Discovery and Targeted Approaches for Comparative Label-Free Proteomic Quantification via Mass Spectrometry

by

Ramanaguru Siva Piragasam

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Biochemistry
University of Alberta

Abstract

Over the last few decades, the field of proteomics have developed rapidly owing to advancement in genomic and transcriptomic technologies contributing to compilation vast quantities of protein sequence libraries. Proteomics, the study of global protein composition in a biological system, have become priority within biological and health sciences for its capability to elucidate subtle complexity of protein biochemistry on a larger system-wide scale with relatively lower time cost. Vast protein sequence libraries coupled with advancement in mass spectrometry and liquid chromatography pushed mass-spectrometry based proteomics to the forefront of protein investigation in terms of protein identification and more recently protein quantification.

Quantification in mass spectrometry is primarily focused on relative approaches where label-free comparative protein quantification has been popular recently compared to label-based methods due to its versatility of applications from simple study of individual proteins to complex samples covering wide range of sample types; purified proteins, cell lysates, biological fluids and tissue and organs. Several label-free quantification techniques have been developed to determine protein abundance. However, label-free quantification has several limitations that require optimization for its implementation. Issues such as data normalization, treatment of missing data and statistical approach and corrections are all active research areas. Thus, the best strategy for the execution of a label-free quantification analysis is yet to be finalized.

Additionally, the focus on quantification by mass spectrometry have allowed for the development of several data acquisition methods each with their own advantages and disadvantages with respect to protein identification, quantification accuracy and sensitivity. Traditionally, proteomics experiments have focused on discovery based techniques, which tries to

maximize protein identification such as the data dependent acquisition (DDA). Later, several quantification-focused data acquisition methods were introduced such as selected reaction monitoring (SRM) and parallel reaction monitoring (PRM) that primarily quantifies protein based on prior knowledge of protein identification. More recently, a more hybrid approach that boasts excellent protein identification rate while retaining reproducible and sensitive quantification have been introduced, the data independent acquisition (DIA).

The practicality and suitability of label-free quantification method to be applied in wide range of biological systems including clinical samples have lead it to become our lab's choice for quantification method. Our group have worked to develop robust, reproducible and reliable approach towards label-free proteomics analyses. We have focused on sample specific normalization utilizing multiple data acquisition strategies for label-free proteomics studies over a wide range of biological systems. This thesis focuses on the application of our mass spectrometry based label-free quantification strategy for comparative proteomics analysis via multiple data acquisition methods in the characterization of:

1. The proteomics changes in HEK293T cell cultures upon over expression of miR23~24 miRNA cluster via data dependent acquisition.
2. The proteomics differences between blood serum of Myasthenia Gravis patients and healthy control to identify potential biomarker candidates. Initial discovery was based on data dependent acquisition and further validation by targeted proteomics approach via parallel reaction monitoring.
3. The differential expression of ribosomal proteins in mouse tissues and cell lines to investigate ribosomal proteins' role in specialization and development of tissues. Ribosomal proteins are quantified primarily by parallel reaction monitoring method.

Preface

This thesis is an original work by Ramanaguru Siva Piragasam and is written as part of requirement of Doctor of Philosophy Degree in Biochemistry at the University of Alberta. The human serum samples used in Chapter 3 of this thesis were de-identified and provided by the Neuromuscular Clinic in Kaye Edmonton Clinic, with approval from the University of Alberta Health Research Ethics Board – Biomedical Panel (Project Title: Metabolomic Profiling of Serum in Myasthenia Gravis: Pilot Study, PRO00030698, 24th May 2012). The study was funded by Stollery Children's Hospital Foundation and supporters of the Lois Hole Hospital for Women through the Women and Children's Health Research Institute (WCHRI) and University Hospital Foundation. This project was additionally supported by a Discovery Grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada to Dr. Richard Fahlman. The work completed in Chapter 4 were funded by NSERC's Discovery Grant while the animals were provided by Janice Cotton-McKay from the Health Sciences Lab Animal Services.

For Chapter 2, cell culture work and RNA analysis via Northern Blotting were completed by Dr. Steven Chaulk and assisted by Braden Millan. Proteomic sample preparation, data acquisition and analysis was done by myself. Protein functional analysis were completed with the help of Dr. Faraz Hussain. Dr.Richard Fahlman and I were responsible for study design.

A version of this chapter has been published as:

Ramanaguru S. Piragasam, S. Faraz Hussain, Steven G. Chaulk, Zaeem A. Siddiqi, and Richard P. Fahlman. Label-free proteomic analysis reveals large dynamic changes to the cellular proteome

upon expression of the miRNA-23a-27a-24-2 microRNA cluster. Biochemistry and Cell Biology. 98(1): 61-69. https://doi.org/10.1139/bcb-2019-0014

For Chapter 3, study design, sample preparation, immunological assays, and data analysis were done by Dr. Faraz Hussain. All mass spectrometry related work including experimental design, sample preparation, data acquisition and analysis were completed by myself. ELISA experiment and the data analysis was performed by Hassan Sarker from Dr. Carlos Fernandez-Patron's team. Dr. Derrick Blackmore obtained patients consent, collected and processed serum samples and provided relevant clinical data. Dr. Elaine Yacyshyn provided clinical data and serum samples of Rheumatoid arthritis patients.

A version of this chapter has been prepared for publication titled "Residual Serum Fibrinogen as a Universal Biomarker for all Serotypes of Myasthenia Gravis" with the following authorship:

Faraz S. Hussain[1,5], PhD, Ramanaguru S. Piragasam[2,5], Hassan Sarker[2], Derrick Blackmore[1], PhD, Elaine Yacyshyn[3], MD, Carlos Fernandez-Patron[2], PhD, Richard P. Fahlman[2,4*], PhD & Zaeem A. Siddiqi[1*], MD, PhD.

[1] Division of Neurology, Department of Medicine, Faculty of Medicine & Dentistry, University of Alberta. [2] Department of Biochemistry, Faculty of Medicine & Dentistry, University of Alberta. [3] Division of Rheumatology, Department of Medicine, Derrick, University of Alberta. [4] Department of Oncology, Faculty of Medicine & Dentistry, University of Alberta. [5] FSH and RSP contributed equally to this work.

For Chapter 4, Dr. Faraz Hussain and I were responsible for tissue harvesting and sample preparation. Proteomics data collection, method optimization and data analysis was performed by

myself while the study design were equal contribution from myself, Dr. Faraz Hussain and Dr. Richard Fahlman.

Supplemental materials utilized for this thesis can be found collectively at : Supplemental Tables

Acknowledgement

The completion of this thesis and the work that went into it has been a journey. A journey that would not have been possible without the contribution of many individuals. Thus I would like to take this opportunity to acknowledge their help in my achievement. First and foremost, I would like to thank my supervisor, Dr. Richard Fahlman, for his mentorship and guidance all this years starting during my undergraduate years where his introductory to mass spectrometry inspired me and steered my path towards research in proteomics. Richard has always been supportive towards me and gave me freedom to explore my research interests and always present to answer any questions I might have. I express my deepest gratitude to you for the friendship, knowledge and mentorship all these years.

The work in this thesis has been made possible with the help of the team in Alberta Proteomics & Mass Spectometry (APM) and Lipdomics Core. I must thank Jack Moore, Paul Semchuk and Audric Moses for their support. They taught me how to operate and troubleshoot chromatography machines and mass spectrometers. I have learned a lot from them and will be forever grateful.

The members of Fahlman lab, present and past, has been instrumental towards the completion of my thesis. Angela Fung, Prarthna Nagar, David Kramer, Luana Leitao, Mohamed Eldeeb have spent their time with me, challenging me and improve my knowledge. Mohamad Eldeeb in particular has been a very dear friend and guided me during my time with him. I would also thank our post-doctoral fellow and my dear friend, Dr. Faraz Hussain for his knowledge, wisdom and friendship. He has shaped me and let me grow in this field under his guidance.

I would also like to thank my supervisory committee Dr. Ing Swie Goping and Dr. Joanne Lemieux for their support and guidance.

Lastly, I would extend my eternal and deepest gratitude to my family, especially my wife, Mira Yong for her patience and love that have kept me going on through tough times providing me unlimited emotional and mental support. My childrens, Amir Danish, Aariz Yunus and Aerina Sofia have always been there for me providing emotional support. Importantly I am grateful to my father, Siva Piragasam Govindasamy, and mother, Amutha Kannan, for believing in me, always rooting for me and care they have given me through my journey in graduate studies as well as their sacrifice to bring me to where I am today. My siblings, Kumaraguru, Parasuramaguru, Vishnooguru and Darsheni Devi have always supported my goals and helped me achieve them. Also I would like to thank my friends in Edmonton for their support, especially, Abul Kalam Azad, Ali Kassim, Marina Omar, Shaedy Saharom, Faizah Razali, Fakrul Luthfi, Farah Shuhaimi, Tarmizi Juhari, Azlyn Mohamad and many more that I have missed.

Table of Contents

1 Chapter 1: Introduction to Proteomics 1

1.1 Overview of Proteomics 2

1.2 Mass Spectrometer 5

1.2.1 Ion Source 6

1.2.1.1 Electrospray Ionization (ESI) 7

1.2.2 Mass Analyzers 9

1.2.2.1 Quadrupole Mass Analyzer 10

1.2.2.2 Orbitrap 12

1.2.3 Mass Spectrum 14

1.3 Proteomics and Mass Spectrometers 17

1.4 Sample Preparation for Mass spectrometry 18

1.5 Data Acquisition by Mass Spectrometry 20

1.5.1 Data Dependant Acquisition (DDA) 22

1.5.1.1 Fragmentation 24

1.5.2 Data Dependent Acquisition (Continued) 25

1.5.3 Data Independent Acquisition (DIA) 27

1.5.4 Selected Reaction Monitoring (SRM) and Parallel Reaction Monitoring (PRM) .. 30

1.6 Protein Quantification in Comparative Proteomics 36

1.6.1 Label-based Quantification 37

1.6.2 Label-Free Quantification 39

1.7 Statistical Approach in Proteomic Data Interpretation 44

1.7.1 T-test 45

1.7.2 Analysis of Variance (ANOVA) and Tukey's Honest Significant Difference (HSD) Test 45

1.7.3 Multiple Testing Problem in Statistics..... 47

2 Chapter 2: Comparative Proteomics of Differential Protein Abundance Upon Cellular Expression of Mir23~24 Cluster 49

2.1 Contribution and Acknowledgement 50

2.2 Abstract 51

2.3 Introduction 52

2.4 Materials and methods..... 55

2.4.1 miR-23~24 expression in cell culture 55

2.4.2 RNA analysis 55

2.4.3 Proteomic analysis 55

2.4.4 Data analysis 57

2.4.5 Parallel reaction monitoring..... 58

2.5 Results 59

2.5.1 MicroRNA cluster expression..... 59

2.5.2 Proteomic analysis 59

2.5.3 Functional analysis of the altered proteome 63

2.5.4 Changes in ribosomal proteins 64
2.5.5 Validation of changes in ribosomal protein abundance 66
2.5.6 Comparison of proteomic data with miRNA predicted targets 67
2.5.7 Top candidate targets of the miR-23~24 cluster 68
2.6 Discussion 70
2.6.1 Proteomic analysis 71
2.6.2 Target identification 73
2.7 Concluding remarks 76
2.8 Future Direction 76
3 Chapter 3: Potential Biomarker Identification Through Serum Proteomics for Myasthenia Gravis 78
3.1 Contribution and Acknowledgement 79
3.2 Abstract 81
3.3 Introduction 82
3.4 Methods 85
3.4.1 Ethics Approval. 85
3.4.2 Serum Samples. 85
3.4.3 SDS-PAGE, Western Blot Analysis & ELISA. 86
3.4.4 Serum Protein Profile by Gel-LC-MS/MS. 86
3.4.5 MS/MS Data Analysis. 87

3.4.6 Quantification and Statistical Analysis. 87

3.4.7 Parallel Reaction Monitoring. 87

3.5 Results 88

3.5.1 Myasthenia Gravis Serum Proteome Profiling. 88

3.5.2 Identification of Proteins of Differential Abundance. 89

3.5.3 High Fibrinogen in MG Sera. 91

3.5.4 Validation of High Fibrinogen in MG Patient Sera. 92

3.5.5 Residual Sera Fibrinogen is Stable at Low Temperature. 95

3.5.6 Blinded Analysis of Residual Serum Fibrinogen Identifies MG Patients. 96

3.5.7 Cohort Analysis by Targeted Mass Spectrometry for Residual Serum Fibrinogen. 97

3.6 Discussion 103

3.7 Future Direction 106

4 Chapter 4: Proteomic Quantification of Ribosomal Proteins in Mice Tissues 107

4.1 Acknowledgement and Contribution 108

4.2 Abstract 109

4.3 Introduction 110

4.4 Methods 113

4.4.1 Tissue harvest, homogenization and ribosome preparation 113

4.4.2 Peptide preparation for LC MS/MS analysis 113

4.4.3 Peptides and ions selection for PRM 114

4.4.4 Proteomics sample preparation and data acquisition. 114

4.4.5 Peptide selection and optimization 115

4.4.6 Data Analysis 115

4.4.7 Statistical Analysis 115

4.5 Results 116

4.5.1 Peptide Selection and Optimization 116

4.5.2 PRM Data Analysis 116

4.5.2.1 Small Ribosomal Proteins 116

4.5.2.2 Large Ribosomal Proteins 119

4.5.3 Tissue Specific Ribosomal Protein Variation 121

4.5.4 Significantly Varying Ribosomal Proteins 121

4.5.5 Tissues With Significant Ribosomal Proteins Variation 123

4.5.6 Tissue Specific Isoform Variation 125

4.6 Discussion 128

4.7 Future Direction 131

5 Chapter 5: Conclusion 140

5.1 Future of Quantitative Proteomics 148

6 References 150

List of Tables

Table 2-1 Pathway enrichment for proteins determined to change in abundance at a statistical cut-off of p < 0.01 64

Table 3-1 Cohort Demographics..... 89

Table 3-2 List of peptides and transition ions used for PRM based cohort study of 31 MG sera, 18 RA sera and 30 control sera...... 99

Table 4-1 List of Peptides of Ribosomal Proteins Quantified..... 139

List of Figures

Figure 1-1 Electrospray Ionization..... 8

Figure 1-2 Quadrupole mass analyzer 11

Figure 1-3 Orbitrap mass analyzer 13

Figure 1-4 Mass resolution and resolving power 16

Figure 1-5 Data Acquisition Methods for Discovery Vs. Targeted Proteomics 23

Figure 1-6 Label Free Quantification: Extracted Ion Chromatogram (EIC) vs. Peptide Spectral Counting 41

Figure 2-1 Expression of the miR-23~24 cluster 60

Figure 2-2 Label free proteomic analysis of cells expression the miR-23∼24 cluster..... 62

Figure 2-3 Global analysis and selected validation of ribosomal proteins. 66

Figure 2-4 Overlap of predicted targets of miR-23∼24 with observed proteomic changes. 68

Figure 2-5 Top candidate target genes for the miR-23∼24 cluster. 69

Figure 3-1 Clinical Proteomic Analysis of Myasthenia Gravis (MG) Patients. 91

Figure 3-2 Immunochemistry Detection of Residual Fibrinogen in Serum. 95

Figure 3-3 Blinded Investigation of MG Serum Samples. 96

Figure 3-4 Targeted Proteomic Analysis of a Clinical Cohort 101

Figure 3-5 Prescription Drug Use of MG Patients..... 102

Figure 4-1 Summary of Protein Quantification of Small Ribosomal Proteins...... 118

Figure 4-2 Summary of Protein Quantification of Large Ribosomal Proteins 120

Figure 4-3 Changes in Ribosomal Proteins 122

Figure 4-4 Specificity curve of ribosomal proteins across tissues and cell lines 125

Figure 4-5 eS10 Peptide Quantification Data 127

Chapter 1:

Introduction to Proteomics

1.1 Overview of Proteomics

Marc Wilkins of Macquarie University introduced the word '*proteome'* in 1994. A *portmanteau* of *protein* and *genome*, proteome refers to the total composition and abundance of proteins expressed in any biological system(Wasinger et al., 1995; Wilkins et al., 1996). Proteins, encoded by the genome, are the major macromolecule in organisms, where all biological processes are carried out by or with the help of proteins (Feijó Delgado et al., 2013; Polakis & Bartley, 1966; Yamada & Sgarbieri, 2005). The fundamental of biology depends on the functions of proteins such as catalyzing metabolic reactions, providing structural integrity to cells, responding to stimuli, transporting molecule. Proteomics is a large-scale analysis of proteins involving but not limited to their structure, spatial distribution, temporal dynamics, response to stimuli in cell, tissue, or organisms under a set condition (Christoforou et al., 2016; Lindqvist et al., 2018; Marx, 2015; Rigbolt et al., 2011).

Improvements in modern nucleotide sequencing techniques for both DNA and RNA as well as genome sequencing projects contributed to the field of genomics and transcriptomics enormous amounts of data with respect to protein-coding sequences (A. Bateman et al., 2015; Craig Venter et al., 2001; Leinonen et al., 2004). Proteomics, an interdisciplinary field, benefited massively from achievements in the field of genomic, which provided the blueprint of possible gene products compiled in expansive databases that are the focus of proteomic studies. While the genetic makeup of an organism determines the system's proteome, it can't predict the dynamics of proteome at any point in time. Even within a cell, as it progresses through its life cycle, distinct set of genes are expressed thus varying its proteome composition depending on spatial and temporal location, cell-cycle progression, epigenetic regulation, cell type, cellular metabolism and

energy demands and the cellular environment (Azimifar et al., 2014; Bartke et al., 2013; Geiger et al., 2012; Grünenfelder et al., 2001; Lindqvist et al., 2018; Ly et al., 2014; Marx, 2015; Minamoto et al., 1999; Rigbolt et al., 2011). Previously, protein quantities were always inferred from quantity of mRNA via RNA analysis. However, this was proven to be inaccurate as the correlation between mRNA amount and protein abundance is very poor (Edfors et al., 2016; Vogel & Marcotte, 2012). mRNAs are not always translated to protein and several factors affect the rate of protein translation from mRNAs such as mRNA half-life, cellular physiology, presence of specialized ribosomes (Wegler et al., 2020). Thus, its incredibly difficult to predict number of proteins translated per molecule of mRNA making mRNA quantification to predict protein abundance almost irrelevant. The complexity of the proteome is further increased by post-translational modification of proteins to further fine-tune their physiological activity and interactions with other biomolecules. Plenty of modifications on proteins has been discovered including but not limited to phosphorylation, ubiquitination, glycosylation, methylation, acetylation, oxidation or combination of modifications that alters the chemical and physical properties as well as the structures of proteins (Hirota et al., 2003; Huddleston et al., 1993; Neubauer & Mann, 1999; Peng et al., 2003). This led to the creation of subfields in proteomics such as phosphoproteomics and glycoproteomics that study post-translational modifications of proteins (Gruhler et al., 2005; Kaji et al., 2003; Qiu & Regnier, 2005; Y. Zhang et al., 2005).

Proteomics, if well uncovered, promises great understanding of cellular and organismal physiology but there are great challenges needed to be overcome by scholars trying to study proteomics. One of the main challenges of proteomics is also the advantage proteomics have on genomics, the complexity of the proteome. Biological processes of a cell require differential gene expression leading to varying abundance of protein – some proteins are expressed in higher

quantities than others for any processes. Thus, in abundance alone, a proteome has a high dynamic range of proteins – in some cases proteome exhibits over 10^6-fold difference in protein quantities(Celis & Gromov, 1999). For example, serum proteome is considered the most complex proteome to study (Anderson et al., 2004). It contains highly abundant proteins such as albumin at a concentration of ~50mg/mL, other cytokines, immunoglobulins, and lipoproteins while other secreted proteins generally of interest often present in very low quantities – in femtomoles range (Thadikkaran et al., 2005). High abundant proteins often interfere during measurement of these proteins of interest. Limited and variable sample availability also limits proteomic studies. Unlike in genomics where samples can be conveniently amplified following the advances in DNA polymerase technologies, proteomics studies are lack sample amplification methodology thus had to depend on reproducible sample collection. Moreover, proteins, made from combinations of 20 amino acids, can have high degree of variation in both chemical and physical properties introducing difficulties in solubilizing as well as managing sample degradation. Solubilizing proteins require careful consideration of pH, hydrophobicity, ionic strength of buffer to ensure maximum protein stability (Bodzon-Kulakowska et al., 2007).

Historically, proteomics experiments had been incredibly tasking. Prior to 1980s, individual proteins were identified by first isolating proteins from complex mixtures through techniques often utilizing electrophoresis, most notably two-dimensional electrophoresis (2DE) (Ames & Nikaido, 1976; Klose, 1975). 2DE separates proteins in mixtures based on their isoelectric point using isoelectric focusing (IEF) followed by their size via sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Although there have been improvements in 2DE technologies, limitation remains. 2DE is very labor-intensive, has poor reproducibility, requires large amounts of sample, has limited dynamic range of protein detection and face difficulties resolving proteins

at extreme ends of size or isoelectric point scale (Alaoui-Jamali & Xu, 2006; Godovac-Zimmermann & Brown, n.d.; Wilkins et al., 1999). Following 2DE, isolated proteins are chemically sequenced via N-terminal sequence degradation (Edman Degradation), developed by Pehr Edman in 1949 (EDMAN, 1949). This technique specifically labels the N-terminal residue of a peptide and cleaves it without disrupting peptide bonds of other residues. Following cleavage, the labelled residue are isolated and can be monitored through chromatography to determine the identity of the amino acid. Repeating this process, a peptide can be sequenced residue by residue and the resulting peptide sequence can be used for protein identification. In 1967, this process was automated, however multiple drawbacks persist; it is very labor and time intensive, requires large quantity of purified sample, and the method fails when peptides are longer than 50 amino acids (Edman & Begg, 1967; Niall, 1973). Additionally, when analysing eukaryotic proteins, this method collapses – most eukaryotic proteins have N-terminal acetylation, which interferes during the chemical labelling step of Edman Degradation. To circumvent this issue, proteins must be digested into peptides and purified before sequencing adding more steps and time needed to identify protein. The knowledge potential of proteomics far outweighs its challenges coupled with development in computer technologies allowed researcher to pursue faster and better methods of separation and identification entailing better signals, sensitivity and dynamic range of detection. For this proteomics community turned to liquid chromatography and mass spectrometry establishing mass spectrometry-based proteomics.

1.2 Mass Spectrometer

Briefly, mass spectrometers are instruments that measures mass-to-charge ratios (m/z) of ions through an electric field. Mass determination of ions begins when vaporized ions are

accelerated to high velocity by electric field and then deflected by magnetic field – heavier ions are deflected less than lighter ions with identical charge in a constant electric and magnetic field. J.J Thomson and his student Francis William Aston developed the first mass spectrometer which led to the discovery of isotopes (Downard, 2007; S. J. J. T. O. M. F.R.S., 2009). This work inspired Arthur Jeffery Dempster to design his mass spectrometer in 1918 and in 1919 (Dempster, 1918; D.Sc., 2009), Aston developed his own design. Modern mass spectrometers still apply fundamental theories and principles of these initial mass spectrometer designs. Now, mass spectrometers are used widely in many different fields – environmental (pollutant determination, air and water quality, food contamination), geological (oil composition for petroleum industry, determination of rare earth metals in soil/ore), sports (dope test for athletes), forensic (drug metabolite determination), pharmaceutical (pharmacokinetics, drug metabolism), clinical (drug testing, hemoglobin analysis), biotechnology (proteins, lipids, polysaccharides analysis) (*Mass Spectrometry Applications Areas | Thermo Fisher Scientific - CA*, n.d.). Mass spectrometers are versatile largely due to its customizability where though the concepts stay the same; the parts can be adapted to handle different kinds of samples (solid, liquid, gas, simple or complex). There are three fundamental parts that builds a mass spectrometer – ion source, mass analyzer and detector.

1.2.1 Ion Source

Mass spectrometers only works with analytes in gaseous state; thus, ion source is responsible to ionizing molecules and introduce the resulting charged molecules to the mass spectrometer in an efficient manner. Many ionization techniques have been developed to handle variety of sample types – which can be separated into two groups, hard ionization and soft ionization (Kai et al., 2012; Kersten & Dorrestein, 2009; Steinmann & Ganzera, 2011; Watrous et al., 2010). Hard ionization techniques are methods that blasts sample analytes with high amount

of residual energy resulting in higher degree of analyte fragmentation (systematic dissociation of ions resulting in bond cleavages and stable fragment ions). In-source fragmentation often results in fragment ions with lower m/z than the parent ion and provides information on structural determination of unknown analytes. The most common hard ionization technique is electron ionization (EI), the earliest ionization method – developed by Arthur J. Dempster in 1918 (Dempster, 1918). The process involves highly energetic electrons bombarding analytes to yield extensive fragmentation. However, this method is not preferred in proteomics as it is difficult to control ionisation efficiency and more importantly the inability to measure the molecular weight of intact analyte (Bhardwaj & Hanley, 2014). Soft ionization techniques developed few decades later helped to address this issue – only low energy imparted to analytes resulting in minimal to no fragmentation of analytes. Some examples of soft ionization techniques include chemical ionization (CI), atmospheric-pressure chemical ionization (APCI), matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) – MALDI and ESI being the most preferred techniques in proteomics (Carroll et al., 1974; Field, 1968; Karas et al., 1985; Yamashita & Fenn, 1984).

1.2.1.1 Electrospray Ionization (ESI)

ESI is a soft ionization technique that utilizes electrospray where high voltage is applied to liquid to produce aerosol. ESI, unlike MALDI, able to produce ions with multiple-charged ions,

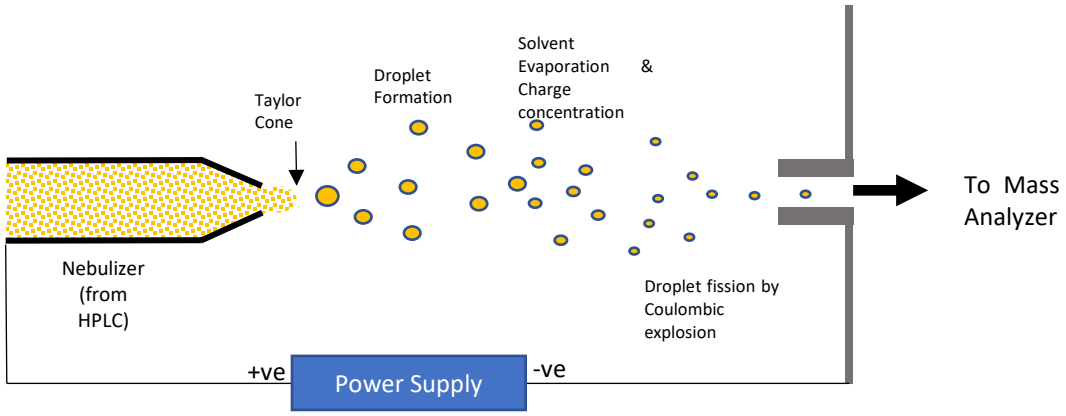


Figure 1-1 Electrospray Ionization

Samples exiting through the nebulizer from the HPLC are subjected to high voltage current between the needle tip and mass spectrometer inlet facilitating droplet formation. Solvents rapidly evaporates from droplets, increasing its surface charge concentration and finally the droplets splits into individual charged ions which are accelerated in the electric field towards the mass analyzer.

thus expanding mass range of analyzer to detect molecules with mass difference of multiple orders of magnitude. First reported by John Fenn and Masamichi Yamashita in 1984, ESI ionizes analytes by first spraying analyte-solvent mixture into fine aerosol across a large potential difference (~3000V) (Yamashita & Fenn, 1984). The tiny droplets will further evaporate leading to decrease in droplet size and increased charge concentration. Ultimately, the droplets will reach the Rayleigh limit – defined as when the electrostatic repulsion of like charges exceeds the liquid surface tension (L. R. F.R.S., 2009). Exceeding the Rayleigh limit, the droplets will undergo Coulombic explosion and form stable tiny droplets. Droplets will undergo Coulomb fission repeatedly until analytes are ionized and ejected into gaseous phase ready to be accelerated into the mass analyzer. In 1989, Fenn and his colleagues were able to ionize biomolecules as large as 76kDa through ESI (Fenn et al., 1989, 1990). Contribution in large biomolecule ionization both via MALDI and ESI for mass spectrometry led to Koichi Tanaka and John Fenn receiving a part in the 2002 Nobel Prize in Chemistry.

1.2.2 Mass Analyzers

Mass analyzers are simply the core of a mass spectrometer, responsible for resolving incoming ions by their m/z thus determining the mass of analyte (Jennings & Dolnikowski, 1990). All mass analyzers work by manipulating the motion charged particles through utilization of electric field, magnetic field, radio frequency (RF) and a combination of these methods. The physics governing the manipulation of the path of charged particles involves the Lorentz force law and Newton's second law of motion;

$$F = Q(E + v \times B)(Lorentz\ Force\ law)$$

$$F = ma \text{ (Newton's } 2^{nd} \text{ Law of Motion)}$$

When equated, the equations can be rearranged to represent force applied to an ion yielding:

$$\left(\frac{m}{Q}\right) a = E + v \times B$$

where m is the mass of ion, Q is ion charge, a is acceleration of ion, E is electric field and v x B is the vector cross product involving velocity of ion and magnetic field (*Introduction to Mass Analyzers : SHIMADZU (Shimadzu Corporation)*, n.d.). This differential equation complemented with ions' initial state, fully explains the motion of ions in time and space in terms of mass over charge (m/Q). However, mass spectrometer data are presented as the dimensionless m/z, where z is the number of elementary charge (e) and can be calculated from:

$$z = Q/e$$

Although m/z is widely referred to as mass-to-charge ratio, it indicates the ratio of mass number (m) with the charge number (z) (Haag, 2016; Jennings & Dolnikowski, 1990). There are many different mass analyzers available with variety of methods to resolve ions based on their

m/z. Additionally, some analyzers are capable of trapping and storing ions thus increasing capability of mass analyzers in manipulating ions. Choice of mass analyzer during modelling an mass spectrometer for commercial productions involves certain factors including but not limited to; (i) analyzer's detection range of m/z, (ii) resolving power of analyzer, (iii) compatibility of mass analyzer with ion source and (iv) limit of detection of the analyzer (Haag, 2016). Some of the commonly used mass analyzers include time-of-flight (TOF), quadrupole, ion trap and Fourier Transform analyzers (Orbitrap) (Haag, 2016). All mass analyzers have strengths and weaknesses and are never suitable for all types of applications. Labs usually have multiple different mass spectrometers that utilize different analyzer to achieve their measurement goals.

1.2.2.1 Quadrupole Mass Analyzer

A quadrupole mass analyzer commonly considered as "mass filter" for its ability to separate and filter ions based on its m/z (Miller & Denton, 1986). It consists of 4 cylindrical rods parallel to each other where opposite poles are electrically connected to each other with radio frequency (RF) potential (***Figure 1-2***). A direct current (DC) is then superimposed over RF potential to manipulate ion trajectories. Operation of quadrupoles involves manipulation of DC and RF potential, which results in stable and unstable trajectories along the applied current (Haag,

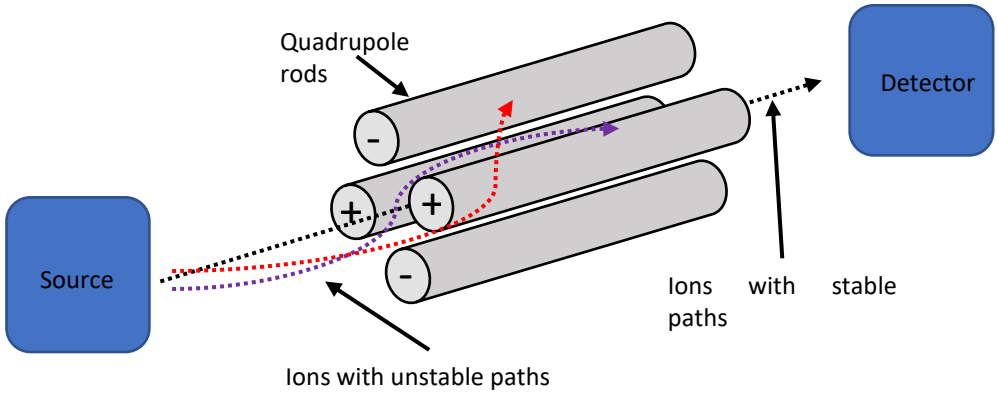


Figure 1-2 Quadrupole mass analyzer

Ionized analytes entering into the analyzer are subjected to varying potential (RF) applied on the two pairs of quadrupoles with opposite polarity. Specific combination of frequency creates resonance and stable path for ions with desired m/z value (black trajectory) while other m/z value carrying ions are non-resonant, experience chaotic trajectory, and will hit the quadrupole thus never reaching the detector.

2016; March, 1997). Ions travelling through z-axis of a quadrupole encounter RF and DC potential causing them to oscillate. Based on the intensity of applied RF and DC potential, only certain ions with specific m/z will have stable trajectory along the quadrupole while other ions with unstable trajectories collide with the rods and filtered out, never reaching the detector. Thus, by varying the RF and DC potential, quadrupoles can be used to scan through or filter ions based on their m/z (Haag, 2016).

Quadrupoles, aside from used as mass filter, do have other functions. They can be operated in RF-only mode, where only RF potential is applied while DC potential is reduced. In this mode, quadrupoles allows all ions to pass through, transforming it into an ion optic component used to transport ions to different part of instrument such as guiding ions from ion source to another analyzer. RF-only quadrupoles also act as collision cell for ion fragmentation via collision-induced dissociation (CID) (Hayes & Gross, 1990). By increasing RF potential and introducing an inert gas into the collision cell, ions within the cell will collide with the applied inert gas and fragment. Controlling the level of RF can vary the level of ion fragmentation in collision cell.

One major advantage of quadrupole mass spectrometer is they are relatively cheaper and small, ideal for many laboratories. They are also very reliable for long period of times and require minimum maintenance making these mass spectrometers popular. They require minimum calibration as the electronics for quadrupole exhibit excellent stability over time. Quadrupoles have

fast duty cycle and need continuous flux of ions making them excellent choice to pair with liquid chromatography (LC) equipments while less suitable with pulsed ionisation sources like MALDI (Arpino & Guiochon, 1979). Some disadvantages of quadrupole include limited mass range and poor resolution. This proves to be a problem when analyzing samples with large molecular weight that do not carry multiple charges and complex mixtures of compounds with similar masses.

1.2.2.2 Orbitrap

The Orbitrap is a modern variant of Kingdon ion trap first described by KH Kingdon in 1923 as a way to generate and studying ion species (Kingdon, 1923). Orbitrap is made up of three main parts, two outer hollow concave-shaped electrodes separated by thin ring of dielectric material covering one inner spindle-like electrode (***Figure 1-3***) (Makarov, 2000). These uniquely shaped electrodes trap ions and leads to electrically induced oscillation of ions to orbit the central electrode that leads to detection of ion masses by Fourier Transformation.

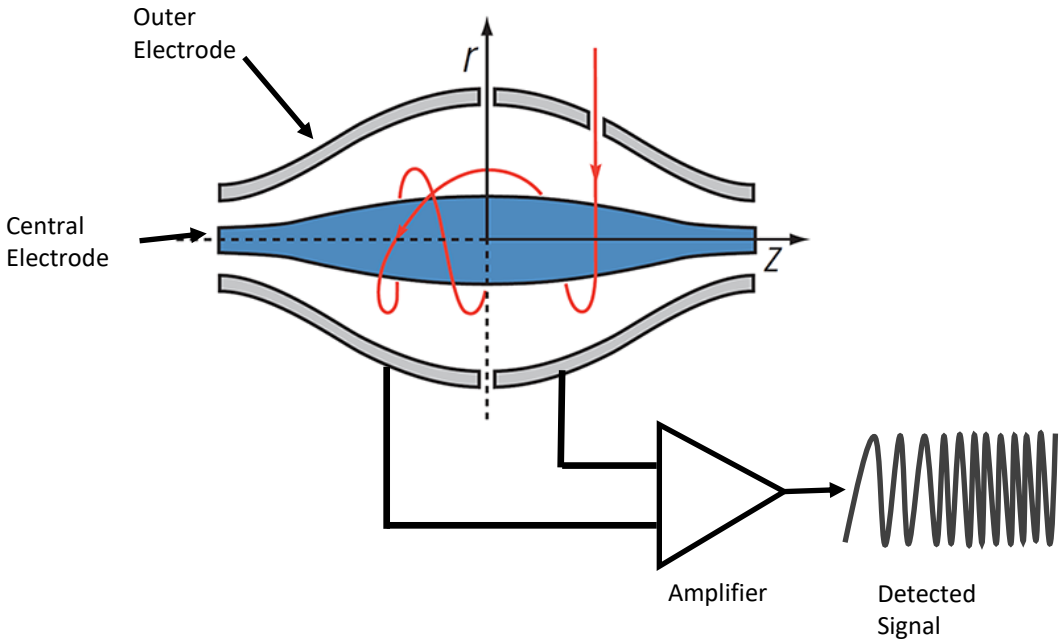


Figure 1-3 Orbitrap mass analyzer

Orbitrap mass analyzer is made up of a spindle-shaped central electrode and a pair of bell-shaped outer electrodes surrounding it. Upon entering into the analyzer, ions spins around the central electrode forming harmonic oscillation along it where the frequency of the harmonic oscillation along the z-axis depends on the m/z of the ion. The outer electrodes are responsible for detecting the image current of the oscillating ions which then transformed by Fourier Transformation. Adapted from www.creative-proteomics.com

A machined hole on one of the outer electrodes is where ion 'packets' are introduced tangentially into the orbitrap, in between the outer and inner electrodes (Haag, 2016). The electric field is increased by ramping up the voltage on the inner electrode as the ions enter the orbitrap causing them to bend and settle into a spiralling oscillation. Additionally, the ions also move back and forth along the axis of inner electrode which following the conical shape of electrodes, induces a harmonic axial oscillation of ions. The axial motion of ions is completely independent of their

orbital motion around the inner electrode as well as any other initial parameters of the ions except their m/z. The angular frequency of the ions, ω is described as:

$$\omega = \sqrt{\frac{k}{m/z}}$$

Where k is a force constant of the potential analogous to spring constant. Thus, ions with specific m/z will spread into rings along the inner electrode based on their angular frequency which can be measured to ultimately detect ion intensities based on m/z.

Currently, orbitrap mass spectrometers are highly preferred for its high-resolution capabilities which is often several times higher than TOFs as well as for its mass accuracy. High resolving power of orbitraps are preserved better in higher m/z range thus making it ideal for analyzing large molecules like proteins. They are also small and low maintenance and since they do not use magnetic field, cryogenic refrigerants are not required, keeping operating cost low. Major cons for orbitraps is the trade-off of scan speed for its high resolution. During high-resolution scans, ions oscillates in the orbitraps for much longer to resolve ions with small m/z difference. However, orbitraps mass spectrometers can be operated at multiple resolution setting thus allowing operators the freedom to prioritize scan speed or resolution.

1.2.3 Mass Spectrum

Mass spectrometers outputs its result as a mass spectrum; an intensity vs. m/z plot representing the distribution of ions in a sample by mass (Murray et al., 2013). As mentioned before, mass spectrometers measures m/z (mass to charge ratio) of ions distinct from individual ions' mass. Ions that exist as two species, singly charged (+1) and doubly charged (+2) have the same molecular mass but the doubly charged species will have m/z that is half of the singly charged

ion. The tallest peak in any mass spectra is referred to as base peak and assigned an intensity of 100. All other peaks are plotted relative to the base peak. Mass spectrometers historically possesses very high mass resolution and accuracy, allowing them to separate isotopic peaks (Downard, 2007). Mass resolution is the ability to distinguish between two closely spaced peaks and can be calculated with the following equation (***Figure 1-4A***):

$$R = \frac{m}{\Delta m_{FWHM}}$$

where resolution (R) of a peak of m/z is the average mass (m) divided by the width of the peak at 50% peak height (full width at half maximum: Δm_{FWHM}) (Murray et al., 2013; "Resolution," 2014; Todd, 1991). Another parameter that often used interchangeably with resolution though carries a different meaning is resolving power. The International Union of Pure and Applied Chemistry's (IUPAC) definition of resolving power is applied for two peaks of same height separated by small increment, where the 'valley' between them is no more than 10% maximum peak height ("Resolving Power in Mass Spectrometry," 2008; Todd, 1991). Resolving power can be calculated by:

$$RP = \frac{m_2}{(m_2 - m_1)}$$

where resolving power (RP) is the heavier peak (m_2) over the difference between heavier peak (m_2) and lighter peak (m_1) (***Figure 1-4B***). Additionally, the rule for valley determination must be stated for resolving power calculation.

In mass spectrometry, mass accuracy (MA), often referred as mass error, is reported as parts per million (ppm) (Brenton & Godfrey, 2010). Mass accuracy is determined as the difference

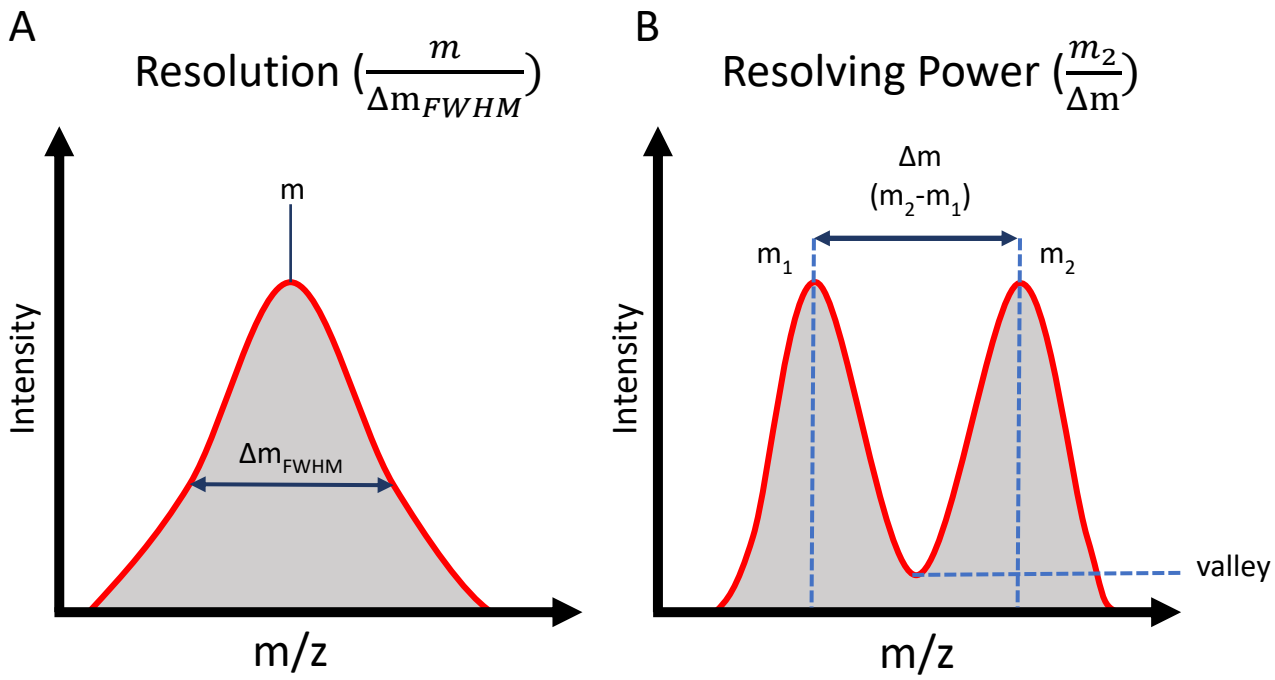


Figure 1-4 Mass resolution and resolving power

A. Mass peak resolution is characterized by the ratio of the measured mass and the peak's full width at half max (50% peak height). **B.** Resolving power is the ability to distinguish overlapping peaks where the peak valley is approximately 10% of the maximum intensity calculated by taking the ratio of the measured mass of the heavier peak and the mass difference between two peaks.

between the measured mass or 'exact mass' (m_E) of a monoisotopic ion and the theoretical 'actual mass' (m_A) of the ion divided by the theoretical mass:

$$MA(ppm) = \frac{(m_E - m_A)}{m_A} \times 10^6$$

Current mass spectrometers can operate in a wide range of resolution depending on length of scan time. For example, orbitrap mass spectrometer with resolving power from 17500 (fast scan) to >100000 (slow scan), can easily separate isotopic peaks of an ion species (Eliuk & Makarov, 2015;

Q. Hu et al., 2005). Together with accepted mass accuracy of ±10ppm and high resolving power determination of mass and charge of an isotopic cluster is quite simple; two peaks in an isotopic cluster have a mass difference of one Dalton (Da) and the charge state will be the reciprocal of the measured difference of m/z of two adjacent peak in an isotopic cluster.

1.3 Proteomics and Mass Spectrometers

A wide variety of mass spectrometers are available due to the potential combinations of available ion sources, mass analyzers and detectors. A select few types of mass spectrometers are however predominant in the field of proteomics (Demartini, 2013). These include mass spectrometers with ion sources: MALDI, ESI and nanospray ionisation (NSI), and mass analyzers: Orbitrap, time-of-flight (TOF), triple quadrupoles (QqQ), and linear ion-traps. Researchers choose the type of mass spectrometers required based on the need of the experiment and two fundamental proteomics approach are considered first: top-down versus bottom-up proteomics.

Top-down proteomic approach involves analysis intact proteins to understand individual protein's structure, post-translational modification and full characterization of proteoforms (all variation of the protein resulting from genetic variation, alternative splicing, and post-translational modifications) (Smith & Kelleher, 2013). This approach requires purified intact proteins directly injected into mass spectrometers for analysis (Kelleher, 2004; Toby et al., 2016). Though this method usually gives 100% sequence coverage, currently it is unable to analyse complex protein mixture thus rendering it unusable to answer many proteomics questions. MALDI-TOF instruments are the predominant choice for this approach.

Bottom-up proteomics (usually known as 'shotgun' approach) refer to a technique where proteins are first proteolytically digested before mass spectrometric analysis (Y. Zhang et al., 2013). Due to the complexity of the resulting proteolytically cleaved sample, this approach needs a better front-end separation: peptides mixture are simplified through chromatography prior to mass spectrometric analysis. Bottom-up proteomics is the most mature and widely used approach because of its versality in handling highly complex sample as well as the availability of bioinformatic tools for protein identification and characterization. Additionally, many quantitative techniques have been developed for bottom-up strategy such as mass tags, stable isotope labels for comparative proteomics to identify up- and down-regulated proteins. Several limitations exists for bottom-up proteomics and most importantly is the low sequence coverage of proteins where only a portion of information is obtained for any protein identifies (Y. Zhang et al., 2013). Therefore, bottom-up strategy sacrifices proteoform information to achieve a better quantification of proteins. Since chromatographic separation is needed for this technique, mass spectrometers with ESI/NSI sources that can work in tandem with liquid chromatography are usually preferred as well as ion-trapping mass-analyzers such as linear ion-traps and orbitraps (Y. Zhang et al., 2013).

My work in this thesis were all done following bottom-up approach utilizing an ultra-performance or high-performance liquid chromatograph (UPLC or HPLC) in tandem with NSI into an orbitrap or quadrupole/orbitrap combination mass spectrometer. Thus, only principles and theories regarding these instruments will be discussed.

1.4 Sample Preparation for Mass spectrometry

All bottom-up proteomic approach requires proteolytic cleavage of protein samples (Y. Zhang et al., 2013). Thus, following protein purification, samples are subjected to proteolytic

digestion prior to mass spectrometric analysis. This is most often performed by proteases with sequence specificity for charged residues such as trypsin, Lys-C, Arg-C, Glu-C, Asp-N, or Lys-N; which results in cleavage of the peptide backbone either preceding or following a charged amino acid (cleavage position is denoted by 'C' for C-terminal or 'N' for N-terminal of specified residue) (Giansanti et al., 2016). Utilization of proteases with specificity for charged residue guarantees the resulting peptide product to have at least a +2-charge following acidification. This allows for easier solvation of the peptides and thus increases efficiency during ionization. Researchers in the proteomics field regard trypsin as the best-suited protease for digestion of protein samples for mass spectrometric analysis (Giansanti et al., 2016; Y. Zhang et al., 2013). It exhibits excellent cleavage specificity to hydrolyze the peptide backbone following basic residues of lysine (K) and arginine (R), though it will not cleave if proline is the next residue following lysine or arginine. Trypsinization of proteins results in relatively short peptides with +2 charge when solubilized in acidic condition. Moreover, trypsin digested peptides are usually in the range of 700 Da to 1500 Da, the ideal mass range for many mass analyzers.

Prior to proteolytic digestion of protein samples for mass spectrometry, few important steps are necessary for preparing the sample. Proteins need to be denatured either by boiling or with the use of denaturing buffer. Disulfide bonds must be reduced with the use of reducing agents such as dithiothreitol (DTT), β-mercaptoethanol (BME) or tris (2-carboxyethyl) phosphine (TCEP) (Rogers & Bomgarden, 2016). The reduced sulfhydryl groups must be alkylated by iodoacetamide to prevent re-forming of disulfide bonds in the later stages of sample preparation (Sechi & Chait, 1998). Following protection of sulfhydryl groups, proteins can be treated with trypsin for proteolysis. The entire process can be done in solution referred to as in-solution digestion. This method is relatively simple but denaturing buffers contains chaotropic agents, salts and detergents

that inactivate trypsin thus, samples must be desalted prior to addition of trypsin (Rogers & Bomgarden, 2016). In-solution digestion is best used when the sample to be analysed is relatively pure, i.e. purified proteins or organelles, proteins isolated from prior proteins separation steps.

The in-gel digestion protocol first explained in 1996 is the most used method for sample preparation for mass spectrometry-based proteomics (Granvogl et al., 2007; Shevchenko et al., 2006). It involves trapping proteins in a polyacrylamide gel and prepare the sample for mass spectrometry by reducing, alkylating, and digesting the proteins while trapped in the gel. Following digestion, the resulting peptides are extracted from gel pieces using organic solvent mixture. By trapping proteins in gels, buffer exchange and desalting become much easier to accomplish. Additionally, researchers usually separate proteins samples in gels by molecular weight via SDS-PAGE thus allowing highly complex samples to be analysed. However, peptide recovery rate from gels are usually not 100% which can affect downstream analysis.

1.5 Data Acquisition by Mass Spectrometry

Proteomics by mass spectrometry involves collection of huge amounts of data achieving near-complete proteome coverage in human and multiple other organisms. The type of data obtained through mass spectrometry heavily depends on the initial research question, often divided into two branches: qualitative or quantitative data.

Qualitative data acquisition in MS, often referred to as survey-based proteomics or discovery proteomics is frequently used to obtain broad overview of proteins in a comparative analysis of various sample complexities ranging from extracted protein samples to tissue lysates (Faria et al., 2017). The strength of discovery proteomics lies in the ability to identify and sort out hundreds or even thousands of proteins in a any given biological states as well as comparing

multiple experimental conditions. Commonly, discovery-based proteomics are initiated without a well-formulated hypothesis and seldomly raises question about the changes of a specific protein of interest within the experimental conditions studied. Traditional qualitative proteomic experiments are designed to produce complex answer to a non-specific question, i.e. what are the proteins that might be up or down regulated with a change in an experimental condition. Historically, discovery-based proteomics data has always been acquired through data dependant acquisition (DDA) as it provides the best results for protein identification in samples from various complexity level (N. W. Bateman et al., 2014; Mann et al., 2001; Ong et al., 2002; Ross et al., 2004). More recently, data independent acquisition (DIA) has been developed to be a more selective discovery-based acquisition method to transition from discovery-based proteomics towards targeted proteomics (Venable et al., 2004).

Quantitative proteomics approaches, often referred to as targeted proteomics, are used for precise quantification of proteins of interest with high sensitivity and reproducibility. Targeted proteomics methods involves analysis of unique peptides of handful of proteins of interest in separate experimental conditions, quantified either through relative or absolute quantification with high sensitivity and specificity in samples with varying level of complexity (Faria et al., 2017). In contrast to qualitative proteomic approaches, which commonly lacks a well-defined hypothesis, targeted proteomics require a clear and specific hypothesis to be tested with its acquisition and analysis method decided before the beginning of experiment. Since targeted proteomics is based on a specific hypothesis, in-depth knowledge of a protein of interest is necessary, from either literature, databases, or previous discovery-based proteomic experiments. Multiple Reaction Monitoring (MRM) was the initial targeted proteomic method developed for quantification of peptides in 1990 by Kusmierz et al (Kusmierz et al., 1990). Later, a more selective targeted method

for peptide quantification was developed, the parallel reaction monitoring (PRM) with better quantification capability than MRM (Gallien, Duriez, et al., 2012; Peterson et al., 2012; Schilling et al., 2015).

1.5.1 Data Dependant Acquisition (DDA)

Commonly referred as shotgun proteomics by the scientific community, DDA refers to a mode of data collection in tandem mass spectrometry. In DDA mode, which refers to a selection criterion of peptides for fragmentation, the mass spectrometer selects most intense peptide ions from the first stage of tandem mass spectrometry, fragments the ions and analyze them sequentially in the second stage of tandem mass spectrometry. Over the last two decades, liquid chromatography-tandem mass spectrometry (LC-MS/MS) running through DDA mode has been the fundamental pillar in the field of proteomics for the broad detection, identification, and relative quantification of thousands of proteins across various biological samples(Wolf-Yadlin et al., 2016).

DDA workflows often includes prior separation of enzymatically digested protein products, peptides, through high-performance liquid chromatography (HPLC) or nano liquid chromatography (nLC)(Wolf-Yadlin et al., 2016; Xu et al., 2013). HPLC and nLC delivers peptides into the mass spectrometer over a solvent gradient (from tens of minutes to a few hours), separating the peptides based on their physicochemical properties, thus reducing sample complexity, and increasing coverage. Data acquisition by DDA can be divided into 2 steps; the first is the MS1 scan and followed by MS2 scans. MS1 scans or full scan are scans that are responsible for identification of m/z of peptide ions injected into the mass spectrometer. During full scan, the mass spectrometer allows all injected ions to pass through the mass filter and reach the mass analyzer for separation by m/z and the resulting data will be transformed into a mass

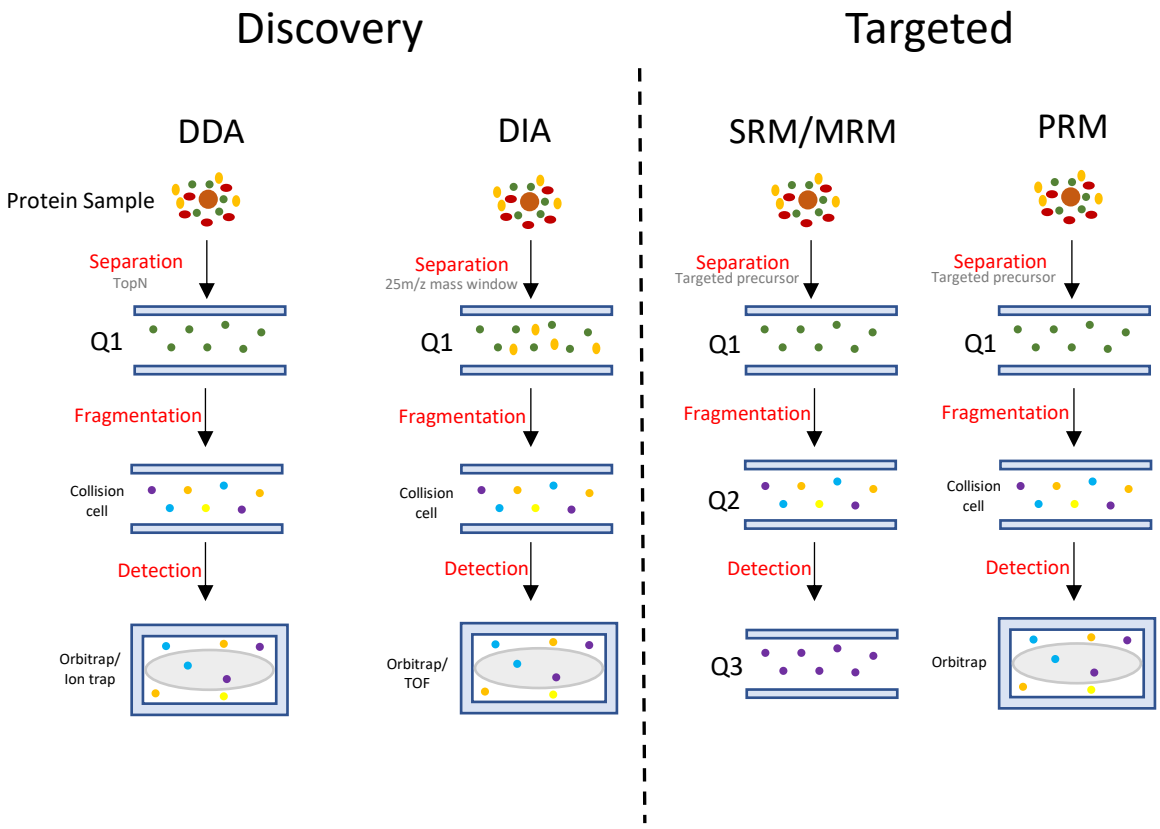


Figure 1-5 Data Acquisition Methods for Discovery Vs. Targeted Proteomics

DDA uses a quadrupole as mass filter to isolate ions passing the TopN criteria for fragmentation in collision cell and followed by detection using a mass analyzer. DIA uses first quadrupole as wide window filter (~25m/z) allowing multiple precursor types through to collision cell for fragmentation and detection by an orbitrap mass analyzer. The instrument repeats the process by scanning the next 25m/z range until entire mass range is covered and peptides are identified during post-acquisition processing of spectra. In targeted proteomics, SRM/MRM uses first quadrupole to isolate user-defined precursor, second quadrupole for fragmentation and third quadrupole to isolate and detect user-defined fragment. PRM works similar to SRM but all the fragment ions from a user-defined precursor are measured in parallel via an orbitrap.

spectrum (***Figure 1-5***). Parameter adjustment needed for a MS1 scan is quite limited, mass range and resolution. Mass range for a mass spectrometer highly depends on the type and capability of mass analyzer equipped in the mass spectrometer with most analyzers capable of detecting m/z

over a wide range while adjusting the resolution of a MS1 scan affects the time spent by the analyzer on a scan. MS2 scans are where peptide ions measured in the MS1 scan are fragmented in the mass spectrometer and the analyzer analyzes the resulting fragments' m/z.

1.5.1.1 Fragmentation

Fragmentation of peptide ions is crucial for the tandem mass spectrometry application. Though many fragmentation methods are available, the most common method is collision-induced dissociation (CID) (Pitt, 2009; Sleno & Volmer, 2004). During CID, peptide ions are isolated and excited to higher kinetic energy before being collided with neutral gases such as helium, nitrogen, or argon, inducing fragmentation of peptide ions.

Orbitrap-based mass spectrometers use a newer form of CID, referred to as 'high energy collisional dissociation' (HCD) where dissociation happens outside of the ion-trap (Perry et al., 2008). Following full-scan, ions with a single m/z are filtered by front optics of the mass spectrometer and allowed to pass through to the C-trap. The ions are then accelerated into the HCD cell at high velocity inducing rapid fragmentation, then shuttled back to C-trap where they are injected into orbitrap for analysis.

Fragmentation of peptides usually happens at characteristic sites; areas corresponding to higher surface area, the peptide backbone, which is made up of three types of atomic bonds, N-C_α, C_α-C_o, and C_o-N (Biemann, 1986; Hunt et al., 1986). The nomenclature of the fragment ions (daughter ions) following peptide fragmentation depends on site of fragmentation, location of ionic charge post-fragmentation and orientation of peptide fragment being read (N-terminus to C-terminus or vice-versa) (Roepstorff & Fohlman, 1984). Depending on whether the charge in the N- or C-terminal ions is retained, A or X ions are produced when the peptide fragments break along the C_α-C_o.

Similarly, dissociation of the peptide bond (C_o-N) produces B and Y ions while fragmentation of N-C_α results in C and Z ion formation. Daughter ions are labelled with numbers indicating the number of residues in the fragment where the numbering originates from their respective terminus. Due to its rigid, massive, planar structure, the peptide bond (C_o-N) is the site of fragmentation that occurs most frequently. Fragment ions produced from peptide bond fragmentation are also the most ideal fragments due to the B and Y ions retaining their relative amino acid compositional mass. B ions' mass can be calculated by the sum of constituent neutral amino acid less the mass of an OH group (17 mass units) where the cleavage at the amide bond leaves the fragment retaining only the carbonyl group. Y ions' mass is the sum of all constituent amino acid's neutral mass plus one mass unit due to internal solvation of proton during fragmentation. Though in theory fragmentation results in two fragments, the analyzer will detect either B- or Y-ion because only one of the fragments will retain the charge while the other will rapidly decompose.

1.5.2 Data Dependent Acquisition (Continued)

"Data dependent" in DDA refers to a predetermined selection criteria of peptide ions for fragmentation during MS2 scans. Analysis of complex samples often result in multiple peptides ionized at the same time (Suna & Mayr, 2018). Attempting to fragment all the peptide ions observed in MS1 scan is not possible as this would overload the mass spectrometer and reduce its scan speed. Thus, peptide ions need to be selected for fragmentation by a predefined selection criterion to limit number of MS2 scans the mass spectrometer runs per MS1 scan. The criteria are abundance, signal intensity and charge state. Selection is made in real time by mass spectrometer software based on the current MS1 mass spectrum where newer mass spectrometers can select the 10 – 40 most abundant peptides observed above a fixed signal intensity threshold (Suna & Mayr, 2018; Xu et al., 2013). Charge state preselection is used to avoid analyzing contaminants. Shotgun

proteomics is mostly performed with peptides digested with trypsin, with preference to cleave the peptide bonds following arginine (R) and lysine (K). Thus, the resulting peptides, tryptic peptides, will always have a basic residue at the C-terminal end and in the acidic conditions the samples are prepared in, the N-terminal is always protonated to yield tryptic peptides of charge +2 at minimum. Charge state preselection can safely exclude singly charged ions and very highly charged ions (+5, +6, +7, +8 …) for shotgun proteomics as those will unlikely be tryptic peptides. Additionally, dynamic exclusion setting prevent repeated selection of the same peptide to increase number of peptide ion analyzed. Dynamic exclusion categorizes an ion previously selected in an exclusion list for a specified time ranging from 10 seconds to 30 seconds depending on the peptide chromatographic elution peak width and the level of identification coverage desired. Once the exclusion time is over, the ion can be selected for fragmentation once again. Excluding highly abundant peptides allows the mass spectrometer to fragment low abundant peptides thus vastly increasing identified protein count from a shotgun proteomics experiment.

DDA is widely used for protein identification in proteomics studies across the world as it is very simple to set up with small number of parameters involved and often once set up in an MS it needs not be changed and thousands of shotgun proteomics experiment can be performed with the same set of parameters. Almost all mass spectrometers are equipped with DDA capability as the hardware requirement for DDA is minimal. DDA results are usually smaller, takes up little space on computer and due to wide availability of database, they are easier to analyze.

Despite the preferential choice of DDA for proteomics analysis, DDA has several inherent flaws that limits its performance, the first being low reproducibility of peptide identification (Domon & Aebersold, 2010). DDA methods can collect thousands of fragment ion spectra in a typical LC-MS/MS analysis. However, tryptic peptides in any proteome digest far outnumbers the

analytical capacity of current DDA systems (that is the separation, detection, and identification of peptides) resulting in inability to perfectly sample a reproducible set of peptides in multiple replicate analysis of the same sample. For every replicate analysis, peptides are chosen by the DDA criteria, which in principle will analyse same peptides, but due to large pool of peptides with matching criteria and limited capacity of instruments, only fixed number of peptides can be analysed thus leading to a stochastic sampling of peptides and affecting reproducibility in DDA analysis. Additionally, DDA methods are set up to maximize identification rate of peptides, which leads to lower accuracy in peptide quantification. DDA features of TopN and dynamic exclusion maximizes protein identification by limiting the frequency of repeated peptide sampling. Though these are efficient for increasing protein identification, it lowers reproducibility of spectral counting and feature based quantification.

1.5.3 Data Independent Acquisition (DIA)

Data Independent Acquisition or DIA, first introduced by Venable et al. in 2004 refer to an MS data acquisition mode that continuously acquires fragment ion spectra (MS2) in an unbiased manner (independent of precursor ion information acquired from MS1 scans) (Venable et al., 2004). DIA, is a hybrid method that was first explored to address the many limitations of DDA and targeted proteomics methods, SRM and PRM, while combining the advantages of discovery and targeted proteomics method. It is mainly implemented to obtain maximum protein identification, >1000s proteins, with high reproducibility and low variation in quantification. In the last decade, DIA has been developing rapidly to challenge the discovery capabilities of DDA and quantification capabilities of SRM and PRM. Many variation of DIA have been introduced over the past decade with differences in type of mass spectrometer used, methods set-up

parameters, and post-acquisition data analysis such as AIF, SWATH, MSX, SONAR, WiSIM, BoxCar DIA and many more (refer to review by Zhang et al. (2020)) (F. Zhang et al., 2020).

One of the most commonly used DIA method is SWATH-MS (Sequential Window Acquisition of all Theoretical Mass Spectra), which was described by Gillet et al (2012) (Gillet et al., 2012). SWATH-MS requires the use of hybrid mass spectrometers capable of fast scanning, high resolution and accurate mass measurements, typically with a quaodrupole as first mass analyzer or selector and a TOF or Orbitrap as second mass analyzer. SWATH-MS and all other DIA methods analyze all peptides in a selected m/z window, typically between 400 m/z to 1200 m/z, where more than 90% of tryptic peptides are found (Koopmans et al., 2018; Pino et al., 2020). Initially, data acquisition begins with MS1 scan to measure precursor population across the whole m/z range. Subsequently, in MS/MS mode, a precursor isolation window of 25m/z is set, beginning with 400 m/z to 425 m/z to isolate and fragment precursors (***Figure 1-5***). All precursors within the selected isolation window will be isolated, fragmented and the resulting fragments analyzed for 50 – 100 ms. Precursor isolation window steps up by 25 m/z to 425 m/z – 450 m/z at the beginning of the next round of isolation for fragmentation. This process repeats several times until the isolation window steps through the whole mass range. A single cycle of MS1 followed by sequential MS2 scans usually takes 3s and cycles through the entire LC gradient. The resulting data, depending on sample complexity, will yield highly multiplexed and complex MS2 spectra with many co-eluting peptides co-fragmented in a single spectrum. Data analysis and subsequent protein identification by traditional genome-wide species-specific databases commonly used to process DDA data were not viable option for these complex MS2 spectra. Many new data-analysis tool emerged specifically for DIA data that are reliable and easy to use such as Spectranaut, OpenSWATH, EncyclopeDIA, Skyline, DIA-NN and more explained in detail by Zhang et al

(2020) (Bruderer et al., 2015; Demichev et al., 2019; Egertson et al., 2015; Röst et al., 2014; Searle et al., 2018).

The major advantage of DIA is its exceptional performance in quantitative analyses of peptides of thousands of proteins with high quantitative reproducibility and accuracy. DIA methods are discovery in nature, similar to DDA, but without the sampling bias of the latter, since DIA samples all precursors in a sample without selection bias. Although, DDA has better protein identification rate than DIA, but that advantage have been reducing over the recent years. There are several reports demonstrating when same samples injected under same condition into same instrument with two data acquisition modes DDA and SWATH-MS/DIA, resulting data suggest SWATH-MS outperforms DDA in peptide identification and associated proteins as well as reproducibility of measurements (Bruderer et al., 2015; Kelstrup et al., 2018). Additionally, DDA methods usually require much higher amount of starting sample, in the range of 100 to 200 micrograms of peptides for off-line pre-fractionation in order to maximize protein coverage. On the other hand, DIA mostly applied in single-shot analysis, requiring about 0.1 to 1 microgram peptides for proteomic analysis with much shorter LC-gradient than DDA (Gillet et al., 2012; Guo et al., 2015; Sun et al., 2020). This means DIA enjoys higher throughput capability and it is ideally suited for projects with large sample number with small sample volume that require accurate and reproducible quantification of major fraction of proteome in each sample.

Despite of the versatility of DIA by excelling in both protein identification and quantification, there exist several limitation to this mode. DIA mode generates very large amount of data for each sample, which places high demand for computational resources in mass spectrometry facilities. Most importantly, these data are composed of highly complex multiplexed MS2 spectra originating form tens to hundreds of co-fragmented peptide precursors. These spectra

requires sophisticated analysis tools many of which are currently in their infancy to isolate fragments of individual precursors for identification and quantification. DDA data analysis pipelines have wealth of tools developed over the past two decades that not available for DIA mode to utilize (Bertsch et al., 2011; Cox & Mann, 2008; Deutsch et al., 2015; Keller et al., 2005). Furthermore, DIA has lower limit of quantification at mid-attomole to low femtomole range, which is better than DDA but still 3 to 10 fold less sensitive than specialized targeted proteomics methods, SRM or PRM measurements (Gillet et al., 2012; Liu et al., 2013; Schmidlin et al., 2016). Due to the complexity of performing DIA experiments, they tend to be more expensive than DDA experiments. This leaves researches to form proper research question and planning before embarking on the DIA route whereas DDA offers cheap and quick data for devising hypothesis thus becoming method of choice for discovery proteomics.

1.5.4 Selected Reaction Monitoring (SRM) and Parallel Reaction Monitoring (PRM)

Desiderio DM et al first introduced concept of targeted proteomics in 1983 by using isotopically labelled peptides as internal standard to measure peptide levels in biological sample (Desiderio & Kai, 1983). In 1990, Kusmierz et al coined the acronym MRM (multiple reaction monitoring) in his report of peptide quantification in human tissue extracts (Kusmierz et al., 1990). Currently, targeted proteomics approach are termed selected reaction monitoring, SRM (used interchangeably with MRM) focus in quantification of selected proteins in varying experimental conditions, either through relative or absolute quantification with high sensitivity and specificity (Catenacci et al., 2014). Parallel reaction monitoring (PRM) is a more recent variation of SRM with similar quantitative capability as SRM but with wider measurement dynamic range and high accuracy despite slower scan speed and sensitivity (Peterson et al., 2012). Data acquisition by

SRM requires specific mass spectrometers such as triple quadrupole (QQQ) or hybrid quadrupole-linear ion trap (QTrap) (Hopfgartner et al., 2004). During SRM analysis on QQQ, the first quadrupole (Q1) acts as a mass filter and selects a predefined list of precursor ions to pass through it while other ions are filtered out (***Figure 1-5***). Selected precursor ions will reach second quadrupole (Q2), which serves as a collision cell to fragment ions. After fragmentation, the products are passed to third quadrupole (Q3) acting as mass filter similar to Q1 to allow only a predetermined list of fragment ions (also called as transitions) reach the detector for analysis. Peak area of each transition are integrated and used for quantification; either by relative or absolute quantification with the prior presence of spiked heavy isotope-labelled internal standard. PRM differ from SRM where PRM analysis are performed on hybrid quadrupole-Orbitrap (q-OT) and quadrupole time-of-flight (q-TOF) mass spectrometers. The orbitrap in q-OT replaces the Q3 of QQQ and this allows all transitions of precursor ions to be measured at the same time in the orbitrap mass analyzer (***Figure 1-5***).

Any targeted proteomics experiment begins with identifying protein candidates for quantification depending on research question to be addressed by researchers. Then quantifiable peptides of proteins of interest will be selected where these peptides should be specific (proteotypic) and stoichiometric (quantotypic) to the protein of interest (Worboys et al., 2014). Proteotypic peptides are peptides that are always observed for a protein, unique to the given protein, regardless of their suitability for quantification. Proteotypic peptides usually fragment well in a mass spectrometer and confidently and consistently identified (Mallick et al., 2006). Quantotypic peptides are proteotypic peptides that are suitable for quantitative analysis. As such, all quantotypic peptides are proteotypic peptides but not all proteotypic peptides are quantotypic peptides. The key characteristic of a quantotypic peptide is its abundance must be stoichiometric

to the abundance of its originating protein. Common cause of concern for a peptide will be modification to its residue. Modified peptide will decrease the abundance of the unmodified peptide thus quantifying unmodified peptide will skew the data for protein quantification.

Below is a list of criteria for selecting a good quantotypic peptide candidate for any protein of interest:

1. Length of peptide: Preferred peptide length are usually between 8-25 amino acids. This ensures the m/z of peptide falls within mass range of mass spectrometer.
2. Uniqueness: Peptide should be unique to the protein of interest. Duplicate peptides can skew the abundance of protein in the sample. Peptide uniqueness can be confirmed by performing a search on Basic Local Alignment Search Tool (BLAST).
3. Miscleavage: Peptides digested using trypsin should be fully tryptic. They should not contain miscleavage sites as miscleavage can affect the abundance of the tryptic peptide with respect to protein abundance. However, presence of proline at the carboxylic side of lysine or arginine shown to produce reproducible miscleavage by trypsin. Thus, peptides with lysine or arginine followed by proline are suitable for quantification (Han & Higgs, 2008). Additionally, fragmentation at the N-terminal of proline in a peptide will result in very intense ion peak that can over lower the peak area distribution of other fragment ions from the same peptide (Breci et al., 2003).
4. Ragged ends: Ragged ends are when tryptic peptides end with a series of ariginine or lysine, for example, KK, KR, RR, RK. These peptides should be avoided as they affect the abundance ratio of peptide to protein.
5. Modification: Modification to peptide residue can alter the peptide cleavage pattern by proteolytic enzymes. Peptides containing amino acids prone to chemical modification

during sample preparation should be avoided if possible. For example, methionine and tryptophan are prone to oxidation, while glutamine and asparagine can undergo deamidation (Fröhlich & Arnold, 2011; Geiger & Clarke, 1987). Additionally, other forms of post translational modification should be considered while choosing the peptide such as phosphorylation (Serine and threonine), or acetylation. However, it is very likely the choice of peptide is very limited, thus modified peptides can be considered if a consistent rate of modification is expected across samples (Fröhlich & Arnold, 2011). One exception is cysteine, where cysteine-containing peptides are consistently reduced and alkylated as part of standard procedure in proteomic sample preparation.

6. Precursor charge: Each peptide can produce multiply charged precursor ion species upon ionization. The precursor ion with a charge state that show better fragmentation and provide measurements that are more sensitive should be selected. Often, doubly (+2) or triply (+3) charged precursors are preferred as they most likely fall into the instrument m/z range. The prevalent charge state of a peptide may differ based on their chemical environment (buffer type, concentration etc.) Histidine-containing peptides should be avoided because it can introduce multiple charges to precursor and product ions. For example, a histidine-containing tryptic peptide can be triply charge instead of doubly charged in acidic condition. Charge state of such peptides must be verified prior to selection for targeted analysis.
7. Chromatographic peak: An ideal peptide's chromatographic peak should have a narrow peak width and symmetrical shape. If multiple peptides are chosen for one protein, ideally they should have different retention times.

8. Signal Intensity: Choose peptide that can ionize efficiently and provide intense and stable signal.

Quantotypic peptide candidates for proteins of interest can be obtained from prior discovery-based experiments performed on similar sample on the same instrument. Additionally, there are many databases available that can aid in planning targeted experiments such as Global Proteome Machine (GPM), PeptideAtlas, or Human Proteinpedia (Beavis, 2006; Deutsch et al., 2008; Prasad et al., 2009). There are also computational algorithms available to predict quantotypic peptides such as PeptideSieve, enhanced signature peptide (ESP) predictor and peptide response predictor (PREGO) (Fusaro et al., 2009; Mallick et al., 2007; Searle et al., 2015). For SRM methods, suitable fragment ions must be selected as well during the method building process.

Once candidate peptides are chosen, they can be tested for practice runs to finalize the correct set of peptides for a protein of interest. This is typically an iterative process to optimize list of precursor before the real data collection begins. For final data collection and quantification, a minimum of two peptides per protein should be measured, though more peptides are recommended in the case where there are discrepancy in quantitative result between the two chosen peptides (Rivers et al., 2007).

For a good quantification, between 10-15 measurement points across the chromatographic elution profile of a precursor are needed. Considering a normal chromatographic peak has an elution width of approximately 30s, the cycle time must be optimized to 3s or less. Cycle time is defined as the time taken to cycle through the entire peptide list. Cycle time is determined by multiplying the number of precursors to the time taken to measure one precursor (transient time). Transient time can vary with the type of instrument and resolution settings for measurement. Lower resolution will give shorter transient time (Gallien, Duriez, et al., 2012). QQQ used in SRM

mode takes around 0.8-10ms per transition, with at least three transitions monitored for each precursor (Holčapek et al., 2012). While, q-OT mass spectrometer such as QExactive used in its lowest resolution mode (17500 at 200m/z) has a transient time of 64ms. Using the parameters of QExactive as an example we can calculate the ideal number of peptides for a PRM method. QExactive has a scan rate of 13 scans per second at its lowest resolution mode. For an ideal cycle time of 3s, QExactive can complete 39 low resolution scans. However, PRM methods often include one high-resolution full-scan at the beginning of the cycle, which will have a transient time of 256ms at resolution of 70000 at 200m/z. Thus, there are ~2.7s of cycle time available for precursor measurements, which gives an ideal number of precursor of 35 percursors for every PRM method. To target large number of peptides in a single PRM method, a scheduled PRM method can be utilized. Scheduled PRM creates a scheduling time window for each precursors in the list thus the instrument will only measure the precursor at the specified time window (Gallien, Peterman, et al., 2012). The time window can be determined with the knowledge of expected peptide retention time using either prior data or retention time predictor. Scheduled PRM approach can measure hundreds of precursors in a single analysis. However, extra attention should be put towards the number of concurrently eluting precursors in the selected time window. The number of concurrent precursors must be below the ideal limit of 35 precursors to ensure proper quantification of the chromatographic peak. Decreasing the size of time window and choosing peptides distributed across the elution gradient can decrease the number of concurrent precursors.

Data analysis in SRM and PRM requires the integration of peak areas of selected fragment ions, which are summed to determine the abundance of peptides and extrapolated to determine protein abundance. Skyline is an open-source tool that is freely available for processing targeted proteomics data including SRM and PRM (MacLean et al., 2010). Skyline automatically processes

data once uploaded into it but the data must still be manually verified to determine any presence of interference in fragment ion peak as well as proper integration of peak area by the software.

Targeted proteomics approaches, SRM and PRM, boasts very good quantification capability with a wide dynamic range (typically 4-5 orders of magnitude), low measurement of coefficients (below 10%) and low limit of detection in the range of femtomoles per milligram of protein (Addona et al., 2009; Gerber et al., 2003; Keshishian et al., 2009; H. Zhang et al., 2011). However, targeted approach are much harder to set up as lots of prior knowledge is required to build the data acquisition method. Additionally, SRM and PRM are time-intensive experiments, as they require extensive method optimization as well as complex post-data acquisition processing for identification and quantification of fragment ion chromatographic peaks (Liebler & Zimmerman, 2013).

1.6 Protein Quantification in Comparative Proteomics

Mass spectrometers have proven to be excellent in qualitative protein analysis through protein identification. The next step in proteomics is moving form 'what' is in a sample to 'what and how much' is it in a sample. That brings about the development of quantitative proteomics in the field of mass spectrometers. However, mass spectrometers are intrinsically limited in terms of quantification due to the difficulties in converting the protein concentration injected into a mass spectrometer into MS-signal. Peptide ionization efficiency can vary between different peptides based on their physicochemical properties by many orders of magnitude (Muntel et al., 2015). Moreover, there is also variability during the generation of peptides from proteins due to different digestion efficiency and peptide solubility. Therefore, converting protein concentration into representative MS-signal are non-uniform and mostly unpredictable. The resulting MS-signal from

any analysis is only and indirect readout of protein abundance in a sample. Nevertheless, protein quantification is still possible from MS-signal is possible following development of quantification techniques to enable both relative and absolute quantification. Before protein abundance can be quantified, peptide signal from every sample must be quantified relatively and integrated to obtain ratio of protein in a sample (Bantscheff et al., 2007, 2012). Quantification of peptides relies on the idea that peptide chromatographic peak size is proportional to number of peptide ions detected in the instrument. Diffrerences in peptide ionization efficiency makes it not possible to compare MS-signal of multiple peptides to determine their abundance. However, MS-signal of a peptide can be compared between multiple samples owing to same ionization efficiency expected for a peptide across different samples, the basis of relative quantification in proteomics (Bondarenko et al., 2002; W. Wang et al., 2003). Absolute quantification in proteomics uses the same principle as relative quantification but peptides' abundance are determined by comparing to the peak of a spiked-in labelled peptide with known concentration (Gerber et al., 2003). There are currently two branches of quantification methods available for researchers for protein quantification through mass spectrometry that are widely used; label-based quantification and label-free quantification.

1.6.1 Label-based Quantification

The major approach to protein quantification by mass spectrometry began with the reliance on stable-isotope labelling introduced by three independent laboratories in 1999 (Gygi et al., 1999; Oda et al., 1999; Paša-Tolic et al., 1999). As explained above, varying peptide ionization efficiency in mass spectrometers lead to difficulty in translating MS-signal to abundance. To address this issue, researchers started introducing labelled peptides into sample for quantification purpose driven by the stable isotope dilution theory. The theory implies that a stable isotope labeled peptide is chemically identical to its non-labeled counterpart, thus should behave in the same manner

during chromatography and ionization and yield distinct peaks due to the isotopic mass difference during mass spectrometry analysis. There are many approaches for labelling proteins/peptides for a label-based quantification, which includes; (i) metabolic labelling techniques such as stable isotope labelling of amino acids in cell culture (SILAC); (ii) chemical labelling techniques such as isotope-coded affinity tags (ICAT), isotope-coded protein labels (ICPL), isobaric tags for relative and absolute quantification (iTRAQ), tandem-mass-tags (TMT), N-terminal labelling, terminal-amine isotopic labelling of substrates (TAILS); (iii) enzymatic labelling techniques such as incorporating heavy ^{18}O during enzymatic digestion of peptides (Desiderio & Kai, 1983; Gygi et al., 1999; Kleifeld et al., 2011; Koehler et al., 2011; Ong et al., 2002; Ross et al., 2004; Schmidt et al., 2005; Thompson et al., 2003).

General label-based experiments are conducted by labelling samples originating from different experimental condition with either heavy (^{2}H, ^{13}C, or ^{15}N) or light (^{1}H, ^{12}C, or ^{14}N) isotopes. Prior to mass spectrometric analysis, the labelled samples are mixed together for simultaneous analysis. Since the samples are chemically identical except for the mass difference due to the introduction of isotope, it is expected all chemical reaction in the sample should occur the same way between the two isotopes tagged groups (Aebersold & Mann, 2003; G. W. Becker, 2008; Chahrour et al., 2015; Z. Zhang et al., 2014). This also applies the following chromatographic separation, with identical peptides from both heavy and light sample co-eluting from column, appearing in the same MS spectra, separated by the mass difference of the heavy and light isotope tags. This allows for direct comparison of peaks from two experimental conditions and can be quantified to reflect changes in protein abundance in samples.

Although isotope-labelling techniques are very powerful and are considered the gold standard for protein quantification in mass spectrometry-based comparative proteomics, there

exists several caveats (G. W. Becker, 2008; Chahrour et al., 2015). The biggest issue with isotope labelling techniques is the increased level of sample handling required. The addition of chemical tags to proteomic samples adds further purification and recovery steps, thus resulting in reduction of starting material for analysis also increasing the time required to perform labelling experiments. Additionally, adding high amount of heavy isotope to a peptide can alter its hydrophobicity thus changing its chromatographic behaviour (Tanaka et al., 1986). Incomplete labelling reaction and ion fragmentation in MS/MS in chemical labelling experiment can result in loss of reporter ion thus introduce an extra level of complexity in data interpretation. Finally, another major hurdle for large scale labelling experiments is the reagents used to perform labelling are incredibly expensive, deterring their use in normal laboratory setting unless necessary.

1.6.2 Label-Free Quantification

Considering the negatives of label-based quantification techniques, a more robust and cheaper alternative is needed for mass-spectrometry based proteomic quantification. With the technological advancement of mass spectrometry and liquid chromatography, label-free quantification have been gaining traction among scientific community as an alternative to labeling techniques (Fabre et al., 2014; Wong & Cagney, 2010). Through careful experimental designs and control, such as automated sample handling, identical amount of starting material or protein concentration, same chromatographic gradient and various methods to 'normalize' peptide ion intensity between runs; label-free quantification can provide information regarding changes in protein abundance with respect to experimental conditions (Karpievitch et al., 2012; Ranjbar et al., 2015; Välikangas et al., 2018). Label free quantification are based on two methods of converting MS-signal to protein abundance; spectral counting or peptide ion extracted chromatogram (al

Shweiki et al., 2017; Arike & Peil, 2014; Cox et al., 2014; Fabre et al., 2014; Nahnsen et al., 2013; Wong & Cagney, 2010; Zhu et al., 2010).

Spectral counting is one of the easiest method of converting MS-signal to abundance, where the number of MS/MS fragment spectra observed for each peptide is tallied up to represent the protein abundance (***Figure 1-6***) (Arike & Peil, 2014; Nahnsen et al., 2013; Zhu et al., 2010). While it is a simple, easy to implement technique, it suffers from issues related to PSM assignment; ion suppression and degenerate peptides, which can lead to either under or over representation of peptide. Moreover, the commonly applied dynamic exclusion setting in DDA mode, used extensively for discovery proteomics, that excludes ions sampled previously from repeated fragmentation negatively affects the quantification accuracy (Old et al., 2005) . Spectral counting also relies on the basis of linearity of response across all proteins when in fact peptide spectral count is affected by its chromatographic behaviour (retention time, peak width, ion suppression by co-eluting peptides) leading to varying responses for different peptides. Nevertheless, the concept of counting spectra as a measure for quantification remain attractive and have led researchers to develop it further to the estimation of absolute protein abundance. Protein Abundance Index (PAI) introduced by Rappsilber et al. is a ratio of the number of observed peptides against all the possible tryptic peptides of a protein that is measurable on the mass range selected for the instrument used in the study (Rappsilber et al., 2002). The method was further developed to include the exponentially modified form (emPAI) to better correlate known protein amounts (Ishihama et al., 2005).

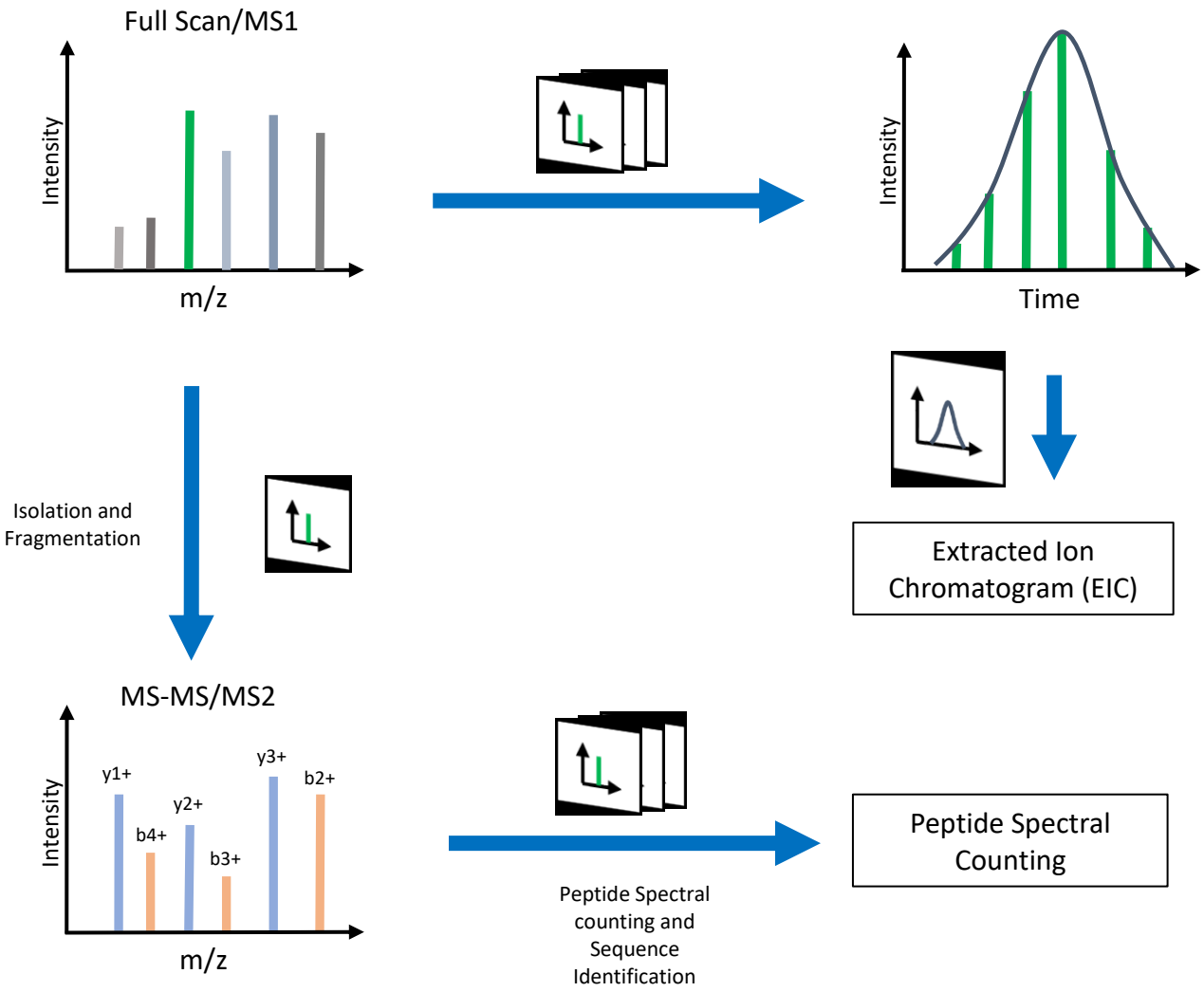


Figure 1-6 Label Free Quantification: Extracted Ion Chromatogram (EIC) vs. Peptide Spectral Counting

Extracted Ion Chromatogram (EIC) is a trace of precursor ion intensity obtained from MS1 scans over time (elution profile). The resulting curve is integrated to calculate the area under curve to represent the peptide's abundance. Spectral counting refers to instances the instrument triggers MS2 scan for a particular precursor. The frequency of MS2 scans for a precursor are compiled and reported as peptide spectral matches (PSM).

Currently, the most widely used technique is quantification based on extracted ion chromatogram (EIC) or MS1 based quantification. As peptides elutes from liquid chromatography column and ionized into the mass spectrometer, their intensity are recorded in the MS1 spectra and their identity can be determined from MS2 spectra upon fragmentation. Once a peptide is identified from the MS2 spectra, the corresponding MS1 signal for a given peptide (based on m/z) can be integrated from all MS1 spectra where the peptide was seen to obtain its chromatographic peak profile. The resulting peak is used to calculate 'area under the curve' as a measure of total number of ion for each peptide (***Figure 1-6***). The calculated area are then used to compare peptide abundance between samples by relative quantification (Bondarenko et al., 2002; Cox et al., 2014; W. Wang et al., 2003). Multiple peptide EICs from a same protein can be combined to reflect the overall protein abundance in a sample. However, as mentioned previously, peptides have different chromatographic behaviour and ionization efficiency, thus can affect the accuracy of protein quantification. One way to account for this issue is the 'Top 3' method employed by the program SEQUEST where protein abundance is determined by averaging the EIC of the three most intense unique peptides (Ahrné et al., 2013; Silva et al., 2006). Intensity based absolute quantification or IBAQ method is another alternative to 'Top 3' method, where total ion intensity for a protein is divided by the theoretical number of tryptic peptide for that protein with the length of 6 – 30 residues (Ahrné et al., 2013; B et al., 2013; Schwanhüusser et al., 2011). In targeted proteomics, quantification are done by taking the area under the curve of fragment ions of a peptide, MS2 signal. Since targeted proteomics does not employ dynamic exclusion like in DDA, peptides of interest are sampled continuously throughout their elution profile, thus summing up fragment ions peak areas to generate peptide abundance increases quantification accuracy. As mentioned earlier, Skyline can help to assign peaks (both MS1 and MS2 signal) for quantification of SRM or PRM

data automatically even though manual checking is required. Nevertheless, the use of MS2 signal for both identification and quantification greatly improves quantification accuracy by avoiding mis-assignment of m/z of peptide of interest that can affect MS1-signal based quantification.

Protein relative quantification is performed by comparing the protein abundance of two different experimental groups, determined through either spectral counting or EIC. Conventionally, an average of protein abundance is used, obtained from averaging protein abundance from multiple replicates of same experimental group. Averaging abundance of biological replicates assumes that all samples were handled identically prior to data collection; however, this is not always the case. Mass spectrometric data collection are affected by drift in elution time of a peptide or the entire gradient and total decrease in ion intensity. The between run variability phenomena does increase the complexity of label-free experiment but researches address the issue by carefully standardizing the front-end procedures, such as (i) reproducibility of liquid chromatography separation, (ii) stability of electrospray ionization, (iii) using computational resources for alignment and integration of multiple LC-MS runs for analysis, as well as normalizing the resulting data. To improve quantification accuracy, normalizing replicates data before protein abundance comparison has become standard procedure in recent times. Some normalization techniques utilized by our lab include spiking identical amount of standard protein in different sample, using a reference protein common in all samples that is known to be in constant abundance, or using a using the total ion current (TIC) which is the sum of EICs of all protein in the sample (Wilm, 2009). Our lab's primary normalization method is to calculate protein abundance as a proportion of the TIC, and less frequently, we use reference protein normalization for studies with good candidate of reference proteins available. Normalization allows researchers

to minimize variation between biological replicates and reducing the impact of EIC of '0' assigned to proteins in some replicates.

1.7 Statistical Approach in Proteomic Data Interpretation

Advances in modern mass spectrometry based proteomics analysis have significantly improved the rate of protein identification and quantification with proteomic analysis capable of identifying hundreds to thousands of proteins in a single sample. Majority of proteomics analysis are comparative in nature (cause and effect) where determination of differential expression of proteins level are crucial to infer important biological mechanisms as well as guide the future direction of a study. When a comparative proteomics study yields thousands of proteins, it becomes a difficult task to interpreting the resulting data and identifying proteins for further validations. Given the nature of validation experiments, which are high cost and labor intensive, a statistical analysis to choose and prioritize proteins of interest becomes an indispensable tool during proteomic data analysis. There are many statistical tests available to test datasets based on data size, conditions, parameters and number of comparisons. In comparative proteomics, the frequently used statistical tests are Student's T test for two population comparisons and analysis of variance (ANOVA) for comparing two or more populations (*ANOVA -- from Wolfram MathWorld*, n.d.; *Student's t-Distribution -- from Wolfram MathWorld*, n.d.). Below are brief introductions of these statistical tests as well as tests applied in this thesis.

1.7.1 T-test

William Sealy Gosset first introduced T-test in 1908 under his pen name 'Student', hence known as Student's T-test. The T-test follows the t-distribution and it is described as the method of testing theory of mean of sample from a normally distributed population where the standard deviation of the population is unknown. In simple terms, t-test is a method to determine the significance of differences between two different populations. There are three types of t-test, each for different types of populations; (i) paired (Student's t-test), (ii) unpaired-homoscedastic, or unpaired-heteroscedastic (Welch's t-test) (WELCH, 1947). Paired t-test assume both populations are equal where they fit normal distribution with equal variance. Meanwhile the second variant of t-test assumes the populations are different but with the measurement of both populations display equal variance. Finally, the unpaired-heteroscedastic t-test assumes unequal population and unequal variance which permits the highest level of 'randomness' of measurement among the three variants of t-test. T-tests calculates t-statistic value based on the means and variance between populations where the higher t-value represent greater difference in the measurement to the sample mean. Once the t-value has been determined, the p-value can be found from the table of values of t-distribution. If the calculated p-value is lower than chosen significance threshold ($p<0.05$), then the null hypothesis will be rejected and the measurement will be considered significant.

1.7.2 Analysis of Variance (ANOVA) and Tukey's Honest Significant Difference (HSD) Test

Analysis of Variance (ANOVA) or in its basic form referred to as One-way ANOVA is the generalized version of two-sample t-test. ANOVA is used to determine the presence of significant difference between two or more groups, where both t-test and ANOVA will give same results for two-sample comparison (*Analysis Of Variance (ANOVA) | Introduction, Types & Techniques*,

n.d.). In ANOVA, measurement groups are compared for their variance between different groups and variance within a group. These two values are computed and the ratio of between group variability and within group variability are reported as the F-statistic value. If the calculated F-value is higher than the F-critical value based of the chosen significance level (α=0.05), then the measurements are significantly different between each other. Although ANOVA can determine the presence of significance difference among groups, it has a limitation in determining the identity of the individual group that is significantly different from the others. Additional statistical tests referred to as post-hoc tests need to be used to further identify groups contributing to the differences among all groups. There are many post-hoc tests available such as Bonferroni correction, Fisher's Least Significant Difference method, Scheffee method and Tukey's Honest Significant Difference Test (HSD) where each of these test have their own strengths and weaknesses and can be applied based on the specific research question explored by a researcher (Bland & Altman, 1995; Haynes, 2013a, 2013b; Tukey, 1977). A summary of comparison between these tests can be found in a review written by Mary L. McHugh (McHugh, 2011).

Tukey's HSD first introduced by John Tukey, is a post-hoc test based on the studentized range distribution that uses pairwise comparisons to determine the presence of difference between the mean of all possible pairs of means. The Tukey test is the most used post hoc test that computes the ratio of the absolute value of difference between two means to the standard error of the mean determined through the ANOVA test. HSD assumes equal sample size however modification can be made to the equation to include unequal sample sizes and this modified method is referred to as the Tukey-Kramer method.

1.7.3 Multiple Testing Problem in Statistics

Statistical tests help determine if a groups of measurements are different from each other due to a specific factor that are being tested. In statistics, a result is considered significant, if the null hypothesis, a statistical jargon that defines the relationship of variables being tested, is rejected. Every time a statistical test is performed, four outcomes may occur, depending on the null hypothesis being found true or rejected by the statistical test. They are (i) a true null hypothesis is rejected (false positive, Type I error), (ii) a true null hypothesis not rejected (true negative), (iii) false null hypothesis is rejected (true positive), and (iv) false null hypothesis not rejected (false negative, Type II error) (Pounds & Morris, 2003). Therefore, there exist a probability a wrong inference drawn from a statistical test result but this can be controlled by carefully planning the experiment and setting a reasonable significance threshold (p-value<0.05). However, proteomics data, similar to other omics field data, require many statistical hypothesis tests to be calculated. In a proteomics dataset to identify differentially expressed proteins, researchers will conduct statistical tests on every identified proteins comparing thein mean expression levels across all experimental groups. Each statistical test have probability to introduce erroneous inference and when computed across thousands of proteins the error rate compounds to a huge number thus decreasing the power of the statistical test. Hence, this forms the crux of the multiple testing problem where the resulting p-values from several consecutive statistical tests increases the false positive rate associated with a p-value cut-off. P-values are often misinterpreted, as the probability the measurement being tested is false. In fact, p-value describes the measure of significance in terms of false positive rate. A false positive rate of 5% (p=0.05) means on average 5% of true negatives will be considered significant by random. In terms of protein comparison, a data set of 1000 identified proteins compared at significance level of α=0.05, then a total of 5% or 50 proteins

will show significant difference by chance even if null hypothesis of no effect is true for all 1000 proteins. Hence, it is much easier to choose a false positive result when choosing a significant result in experiments where multiple testing is involved.

There are several methods that correct for multiple testing when deciding which results are statistically significant, referred to as 'multiple testing procedures' or 'multiple comparison procedures'. Most common among the procedures is the Bonferroni Correction, which control the family-wise error rate (FWER), the probability of making one or more false positive error in a set of tests (Bland & Altman, 1995; Dunn, 1961). Another method which is more popular for biological experiments, is the false discovery rate (FDR) estimation introduced by Benjamini and Hochberg in 1995 (Benjamini & Hochberg, 1995). The FDR method works by determining a threshold of p-value above which only a determined proportion are false positive and the rest are true positive. The FDR defines the proportion of false positive can be expected in a set of significant results. For example, a FDR of 5% means among all results considered significant based on a selected significance threshold, 5% of these results are false positives. Storey and Tibshirani in 2001 redefined FDR as positive FDR (pFDR) and introduced the q-value to provide a measure of significance while automatically taking into account the simultaneous testing of thousands of measurements (Efron et al., 2011; Storey & Tibshirani, 2003). While the p-value cut-off, as a measure of false positive rate, does not provide information on content of set of results considered significant. On the other hands, the q-value of pFDR estimates the probability of the presence of false positives in the set of results considered significant directly.

Chapter 2:

Comparative Proteomics of Differential Protein Abundance Upon Cellular Expression of Mir23~24 Cluster

2.1 Contribution and Acknowledgement

This chapter has been published as:

Ramanaguru S. Piragasam, S. Faraz Hussain, Steven G. Chaulk, Zaeem A. Siddiqi, and Richard P. Fahlman. Label-free proteomic analysis reveals large dynamic changes to the cellular proteome upon expression of the miRNA-23a-27a-24-2 microRNA cluster. Biochemistry and Cell Biology. 98(1): 61-69. https://doi.org/10.1139/bcb-2019-0014

Dr. Steven Chaulk was responsible for cell culture work helped by Braden Millan as part of his summer work and RNA analysis. Dr. Faraz Hussain helped in the analysis of data in regards to functional proteome analysis. I was responsible for mass spectrometric sample preparations, data acquisition for both shotgun and PRM, and data analysis while study design was by Dr. Richard Fahlman and myself.

2.2 Abstract

In deciphering the regulatory networks of gene expression controlled by the small non-coding RNAs known as microRNAs (miRNAs), a major challenge has been with the identification of the true mRNA targets by these RNAs within the context of the enormous numbers of predicted targets for each of these small RNAs. To facilitate the system-wide identification of miRNA targets, a variety of system wide methods, such as proteomics, have been implemented. Here we describe the utilization of quantitative label-free proteomics and bioinformatics to identify the most significant changes to the proteome upon expression of the miR-23a-27a-24-2 miRNA cluster. In light of recent work leading to the hypothesis that only the most pronounced regulatory events by miRNAs may be physiologically relevant, our data reveal that label-free analysis circumvents the limitations of proteomic labeling techniques that limit the maximum differences that can be quantified. The result of our analysis identifies a series of novel candidate targets that are reduced in abundance by more than an order of magnitude upon the expression of the miR-23a-27a-24-2 cluster.

2.3 Introduction

For decades since their discovery (R. C. Lee et al., 1993), gene expression regulation by microRNAs (miRNAs) has been widely investigated with respect to their mechanisms of gene regulation and their networks of mRNA targets. A major mechanism of gene regulation by these small RNAs is their guiding of the RISC complex to target mRNAs by partial base pairing to target sequences within the mRNA sequence (Fabian & Sonenberg, 2012). As a result, to unravel the complexity of miRNA:mRNA regulation networks, numerous investigations have been pursued to identify the targets of individual miRNAs. A summation of many these investigations has been archived in the validated miRNA target database, miRTarBase (Chou et al., 2018). Despite this extensive archive of validated targets, the number of predicted miRNA targets by bioinformatic algorithms, as those utilized by TargetScan (Agarwal et al., 2015), far surpasses the number of validated targets.

While bioinformatic tools have proven invaluable in the identification of potential miRNA targets, these targets still require experimental validation as target prediction algorithms predict many false positive identifications (Pinzón et al., 2017). In addition, we have an incomplete molecular understanding the targeting of mRNAs by miRNAs, for example miR-23a and miR-23b differ by a single nucleotide outside of their common seed region, and thus are predicted to have identical miRNA targets by bioinformatic analysis. On the other hand, laboratory experiments have revealed that in some instances these two miRNAs share targets, such as Pdcd4(X. Hu et al., 2017), but other examples have revealed unique targets for these miRNAs (Li et al., 2016).

Regarding the biological function of miRNAs, an underappreciated aspect is the occurrence of multiple miRNA in the primary miRNA transcript (pri-miRNA). While not all miRNAs

encoded in close proximity in the genome are transcribed as a cluster on a single pri-RNA transcript, many are (Chaulk et al., 2015). The forms of miRNA clusters include those encoded in the introns of coding mRNAs, such as the miR-17∼92 cluster (He et al., 2005) and those expressed from their own promoter, such as the miR-23a-27a-24-2 (miR-23∼24) cluster (Kong et al., 2010). This linked transcription of the miRNAs within a cluster dictates their coordinated expression. Mechanisms have been identified that can lead to differential miRNA maturation of the individual miRNA from the pri-miRNAs, for example differences in the maturation of the miRNA from the miR-17-92 cluster as a result of the folding of the pri-miRNA transcript (Chaulk et al., 2011; Chaulk, Xu, et al., 2014). Despite this, the miRNAs of a pri-miRNA transcript must be considered as a biological unit. For the miR-23∼24, synergistic functions between miR-23a, miR-27a, and miR-24-2 have been identified bioinformatically (Chhabra et al., 2010) and by comprehensive proteomic analysis (Ludwig et al., 2016).

Overshadowing the rapid expansion in the literature regarding various miRNA:mRNA interactions, recent investigations and hypotheses are suggesting that much research on miRNA targeting maybe highly over interpreted (Seitz, 2017). Investigations have demonstrated that the often-reported small changes in gene expression, often under 2-fold, by miRNAs is often less than the variability observed in gene expression between individuals (Pinzón et al., 2017). As opposed to the large-scale changes in gene expression initially observed with the initial identification of miRNAs (R. C. Lee et al., 1993), it is suggested we may need to seriously reconsider the interpretation of small changes in mRNA regulation that is often reported (Seitz, 2017).

In light of this issue regarding the importance of identifying the most significant changes in gene expression by miRNAs to identify the potentially most physiologically relevant miRNA:mRNA associations, we are investigating the utility of label-free proteomics to quantify

the most significant changes to the proteome upon miRNA expression. While proteomic labeling methods, such as iTraq isobaric tags, have been demonstrated to be of high utility to identify changes to the cellular proteome as a result of miRNA expression (Ludwig et al., 2016), the magnitude of changes are typically underestimated as a result of experimental artifacts, that are not widely discussed, that leads to a suppression of the quantified differences observed between samples (Saw et al., 2009; H. Wang et al., 2012). Here we report on the label-free quantification of the proteome changes that result upon the expression of the miR-23∼24 cluster, where previous reports using iTraq almost exclusively only observed changes of under 2-fold.

2.4 Materials and methods

2.4.1 miR-23~24 expression in cell culture

To express the miR-23~24 cluster, the genomic sequence including the 2 kb sequence downstream of the cluster was cloned into a pcDNA 3.1(+) vector. HEK293T cells obtained from ATCC were grown in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum. Transfection of the miR-23~24 plasmid or empty vector control were performed using the calcium-phosphate-based method as described previously (Jordan et al., 1996) to a monolayer of cells in a 10 cm dish. Then, 48 h after transfection, the cells were lysed for either RNA or protein analysis as described below.

2.4.2 RNA analysis

Total RNA was isolated using TRIzol (Invitrogen) for Northern blot analysis as we have previously described (Chaulk et al., 2011; Chaulk, Lattanzi, et al., 2014). For Northern blot analysis, the following DNA probes purchased from IDT were used for each miRNA: miR-23a-3P, 5′-GGAAATCCCTGGCAATGTGAT-3′; miR-27a-3P, 5′-GCGGAACTTAGCCACTGTGAA-3′; miR-24-2-3P, 5′-CTGTTCCTGCTGAACTGAGCCA-3′.

2.4.3 Proteomic analysis

For proteomic analysis, the cells in a 10 cm dish were harvested 48 h after transfection by lysis in 150 μL denaturing lysis buffer [50 mm Tris (pH 6.8), 8% glycerol (*v*/*v*), 1% sodium dodecyl sulfate (SDS, *w*/*v*), 0.125% β-mercaptoethanol (*v*/*v*), 1 mmol/L PMSF, and 1 μg/mL of leupeptin]. To facilitate pipetting, the samples were then sonicated to shear the genomic DNA in the samples. Then, 30 μL of lysate of each sample was resolved by 10% SDS–PAGE and then

visualized with Coomassie Blue staining. Each lane was excised and subsequently cut into 14 equal bands, where each band contains proteins from a different molecular weight range. Each gel fraction was subjected to in-gel tryptic digestion beginning with de-staining twice with 50:50 100mM Ammonium bicarbonate (NH_4HCO_3)/acetonitrile. Once gel fractions are stainless, they are reduced with 10mM β-mercaptoethanol in 100 mM NH_4HCO_3 at 37ºC for 1 hour followed by alkylation with 55mM iodoacetamide in 100 mM NH_4HCO_3 at 37ºC for 1 hour. Gel pieces are dehydrated before the addition of trypsin (6 ng/uL) to cover the gel pieces overnight (~16 hours) at room temperature. Digested peptides are extracted through two extraction stages at 37ºC for 30 minutes; first with extraction buffer (97% water/ 2% acetonitrile/ 1% formic acid) and second extraction with 50:50 extraction buffer/acetonitrile. The resulting peptides were dried and resuspended in 60 µL of 0.2% formic acid in 5% acetonitrile (ACN). Digested peptides were analyzed by LC-MS/MS using a ThermoScientific Easy nLC-1000 in tandem with a Q-Exactive Orbitrap mass spectrometer. Each sample (5 µL) was resolved using a 120 min gradient [0%–45% Buffer B; Buffer A (0.2% formic acid in 5% ACN); Buffer B (0.2% formic acid in 100% ACN)] on a 2 cm Acclaim 100 PepMap Nanoviper C18 trapping column in tandem with a Thermo EASY-Spray column (PepMap® RSLC, C18, 3 µm, 100 Å, 75 µm × 150 mm). For data-dependent analysis, full scans were acquired at 35000 resolution at a range of 400–200 *m*/*z* while 17500 resolution was used for MS/MS scans. Only top 15 ions with +2 and +3 charges were selected for MS/MS with 10 s dynamic exclusion applied to prevent continuous reanalysis of abundant peptides. Following data acquisition, raw data files were compiled for each gel lane and searched with Proteome Discoverer 1.4's SEQUEST search algorithm using the reviewed, non-redundant *homo sapiens* complete proteome retrieved from UniProtKB. The search parameters and quantification were as previously described (Kramer et al., 2017).

2.4.4 Data analysis

The false discovery rates (FDRs) for the identifications of the analyzed samples were as follows: "vector control" samples' actual relaxed FDRs for individual peptides were 0.0438, 0.0416, and 0.0430, while the actual strict FDRs were 0.0068, 0.0067, and 0.0067 for replicates 1–3 respectively; "miR23∼24" samples' actual relaxed FDRs were 0.0427, 0.0358, and 0.0394, while the actual strict FDRs were 0.0070, 0.0078, and 0.0070 for replicates 1–3 respectively. Subsequent analysis was carried out in Microsoft Excel. Extracted ion chromatogram (EIC) was used as a measure of protein abundance with only proteins with EIC > 0 in ≥1 sample were used for comparative data analysis. Relative total ion current (TIC) was calculated by summing the EIC for all the proteins identified in a sample. Then each proteins' EIC were divided by TIC to obtain the "proportion of total" value per sample. To determine changes in abundance in two sample sets (miR-23∼24 vs. vector control), a 2-tailed, heteroscedastic Student's *t* test was applied. Resulting *p* values were sorted and uploaded to the "*q*-value estimation for FDR control" web utility (qvalue.princeton.edu) (Storey & Tibshirani, 2003) to generate estimates of the FDR (*q*-values). The complete set of proteomic data are provided in the Supplementary data, Table S1[2]. Functional analysis of proteins with $p < 0.01$ were analyzed with Database for Annotation, Visualization, and Integrated Discovery (DAVID) version 6.8 (https://david.ncifcrf.gov/) (Huang et al., 2009) and enriched for Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway identifiers. Predicted miR23∼24 targets were obtained from TargetScanHuman version 7.2 (http://www.targetscan.org/vert_72/) (Agarwal et al., 2015), whereas the validated miR23∼24 targets were obtained from miRTarBase version 7.0 (http://mirtarbase.mbc.nctu.edu.tw/php/index.php) (Chou et al., 2018).

2.4.5 Parallel reaction monitoring

Proteins of interest were quantified via parallel reaction monitoring on a Q-Exactive Orbitrap mass spectrometer coupled with ThermoFisher Scientific Easy nLC-1000. Samples were prepared as described above with 40 µL lysate were resolved to approximately 30% of gel length by SDS–PAGE then subjected to in-gel trypsin digestion. The resulting samples were then combined to be analyzed in a single LC-MS/MS run. Multiple peptides per protein (Supplementary data, Table S2[2]) were included for method building with acquisition beginning with a high-resolution full scan (35000) followed by individual precursor isolation by quadrupole (isolation window 1.6 *m*/*z*), HCD fragmentation (NCE 27%) and analyzed by the orbitrap analyzed at 17500 resolution. Raw data were imported and analyzed using Skyline software (MacLean et al., 2010). Peptides' chromatographic peak identity were confirmed by importing Proteome Discoverer result files obtained from the PRM raw data to Skyline and aligning the peptide's retention time to raw chromatogram.

2.5 Results

2.5.1 MicroRNA cluster expression

To investigate the miR-23∼24 cluster, the entire 2.2 kb sequence was cloned into a pcDNA vector. This plasmid or a vector control were transfected into HEK 293T cells and the total RNA from these cells were analyzed by Northern blot analysis for miR23a (***Figure 2-1*A**). In addition, the total RNA from a HeLa cells and a series of breast-cancer derived cell lines was also analyzed for endogenous miR23a levels. As seen in ***Figure 2-1*A**, the expression of the miR-23∼24 cluster from the plasmid in HEK293T cells led to miR23a levels that are comparable to endogenous levels of the miRNA observed in MDA-MB-468 and MDA-MB-231 cells. As with the absence of detectable miR23a in HEK293T cells, miR27a and miR24-2 are only detectable in these cells upon transfection of the vector for miR-23∼24 expression (***Figure 2-1*B**).

2.5.2 Proteomic analysis

To investigate the proteome-wide changes in cells upon expression of the miR-23∼24 cluster, HEK293T cells were transfected with the miR-23∼24 expression vector or an empty vector control. Triplicate samples of the controls and the miR-23∼24 cluster expressing cells were then lysed and analyzed by Gel–LC-MS/MS. The analysis identified a total of 4349 proteins (Supplementary data, Table S1) that were quantified for their relative intensity by quantifying the EICs for the three most intense tryptic peptide observed for each protein and normalizing this data to the total ion current quantified for the sample. The triplicate relative EIC intensities of each protein were compared using individual Student *t* tests. The complete set of statistical comparisons of the normalized data are listed in the Supplementary data, Table S3.

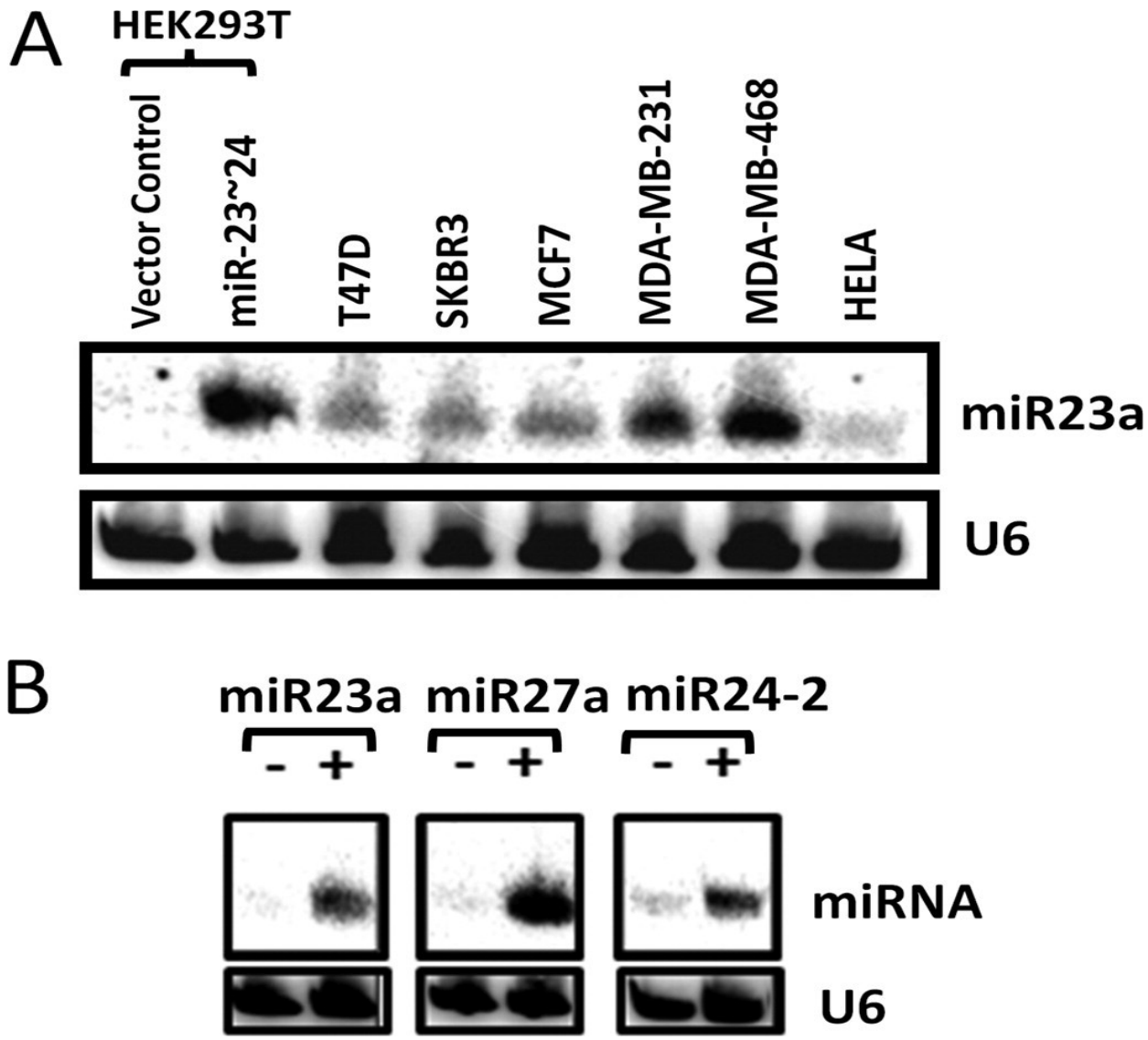


Figure 2-1 Expression of the miR-23~24 cluster

(A) Northern blot analysis for the expression of miR-23a in HEK 293T cells transfected with a plasmid to express the miR-23~24 cluster or the vector control. Total RNA from the indicated cell lines were included for comparison of endogenous miR23a levels in these cell types.(B) Northern blot analysis for expression of miR23a, miR27a and miR24-2 in cells expressing miR23~24 cluster (+) or vector control (-).

The reproducibility of the label-free quantification is relatively high. The comparison of the quantified data from two individual control samples reveals only minimal run-to-run variability for proteins quantified below the relative ion intensity of 1×10^{-3} (***Figure 2-2*A**), whereas the majority of the data exhibit reproducible quantification. In contrast, when the average relative ion intensities for the three replicate vector control samples are plotted against the average relative ion intensities of the miR-23∼24 expressing cells, an overall reduction in the levels of many proteins is observed with the number of proteins observed below the diagonal trendline (***Figure 2-2*B**).

For a more comprehensive comparison of the miR-23∼24 expressing and the vector control cells, including the statistical analysis between the datasets, the fold-change of each protein's relative average ion intensities are depicted as a volcano plot in ***Figure 2-2*C**. Proteins uniquely observed in a single experimental condition were assigned a $\log_2$ value of ±10 to facilitate their visualization on the log scale plot. In this case, proteins only observed once or twice in a particular sample and absent in the other exhibit high *p*-values, whereas proteins repetitively observed in only a single experimental condition result in a low *p*-value; for example, Orc6 was uniquely observed in all three vector control samples but was not detected in any sample from the cells expressing miR-23∼24.

Overall, the analysis again reveals an over-representation of down-regulated proteins upon miR-23∼24 cluster expression, as viewed by the asymmetry of the volcano plot with more proteins being observed to the left of the *Y*-axis. To query the quality of the data, a series of proteins often utilized as loading controls were analyzed for their relative extracted ion intensities. The ion intensities for Gapdh, ActB, GPI, and CFL1 are shown in ***Figure 2-2*D**, and none reveal an observable difference between the control cells or cells expressing the miR-23∼24 cluster. For comparison with these invariant protein controls, we plotted two histone proteins, Hist1H2BK

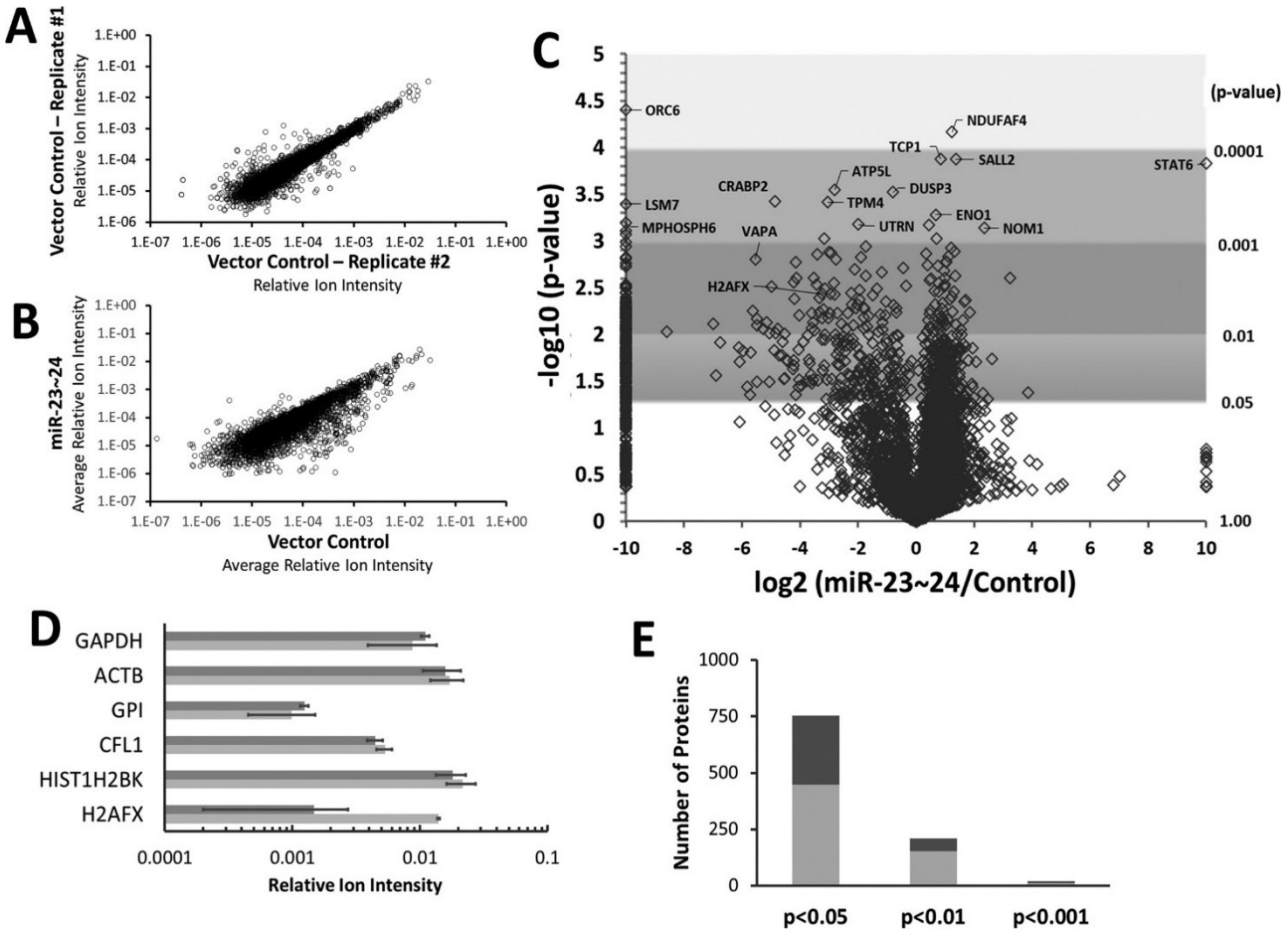


Figure 2-2 Label free proteomic analysis of cells expression the miR-23~24 cluster.

A. The relative ion intensities for identified proteins from two biological replicates of control samples were plotted against each other to demonstrate the reproducibility of quantification and detection. **B.** The average relative ion intensities for proteins identified in the triplicate analysis of cells expressing the miR-23~24 cluster or vector control cells were plotted to demonstrate the global changes observed in the cellular proteomes. **C.** Volcano plot of the data described in (**B.**), which includes the p-values determined for each protein quantified in the comparison of cells expressing the miR-23~24 cluster or vector control cells. Selected proteins from the data set are individually labeled. **D.** Quantified data from selected proteins are plotted to demonstrate the variability between experimental conditions. Major control proteins do not exhibit significant differences between the cells types (dark grey, cells expressing the miR-23~24 cluster; light grey, vector control cells), whereas H2AFX, a known target of miR24a, exhibits a nearly 10-fold reduction upon the expression of the miR-23~24 cluster. **E.** Summation of the number of changes in protein expression observed at the indicated statistical cut-offs.

(Hist H2B type 1-K) and H2AFX (H2A histone family member X), where Hist1H2BK again exhibits no observable variation between samples, but a nearly 10-fold decrease is observed for H2AFX. H2AFX was predicted to be down-regulated upon expression of the miR-23~24 cluster because it is a previously reported target for miR-24-2 (Srivastava et al., 2011).

***Figure 2-2*E** summarizes the observed number of up- and down-regulated proteins at *p*-value cut-offs of 0.05, 0.01, and 0.001. As a result of the multiple testing problem when comparing large datasets (Storey & Tibshirani, 2003), the *q*-values for each of these *p*-value cut-offs were determined as an estimate of the false discovery rate. The corresponding *q*-values for these *p*-values were 0.139, 0.099, and 0.094 respectively.

2.5.3 Functional analysis of the altered proteome

For a global analysis of the proteome-wide changes observed in the cells upon expression of the miR-23~24 cluster, bioinformatic analysis was performed on the function of the proteins that exhibited a change in abundance. For this analysis, proteins that met the criteria of a *p*-value of <0.01 were analyzed. The proteins were analyzed with the DAVID functional annotation tool for KEGG pathway identifiers (Huang et al., 2009) to identify potential enrichments of functions for the proteins observed to change in abundance. As seen in **Table 2-1**, a series of functional groups were identified including ribosomes, RNA processing, and metabolic processes, including glutathione metabolism and oxidative phosphorylation.

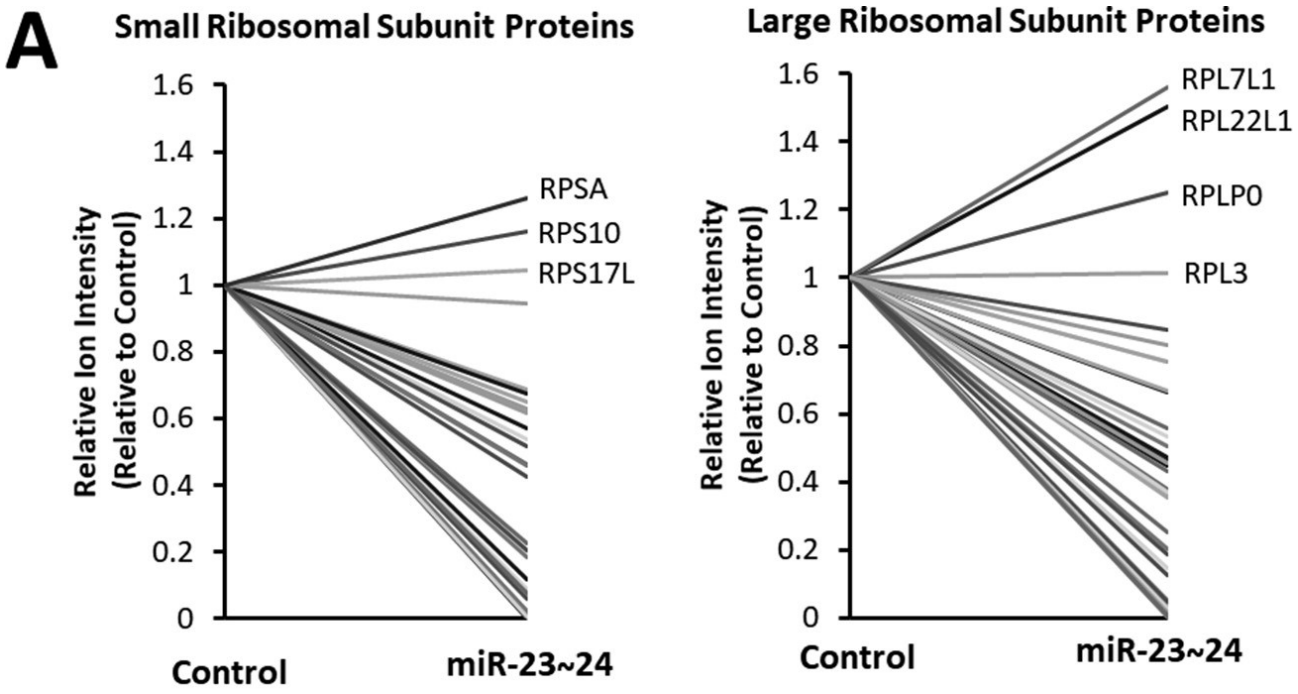
Table 1. Pathway enrichment for proteins determined to change in abundance at a statistical cut-off of $p < 0.01$.

Pathways	Count		p value	KEGG
	Up	Down		
Ribosome	RPLP0	MRPL12, MRPL14, MRPL3, RPL11, RPL19, RPL23, RPL32, RPL38, RPS11, RPS14, RPS6, RPS8	5.2E–07	hsa03010
Spliceosome	PUF60	LSM3, LSM7, RBM8 A, PPIH, SRSF10, SRSF9, SNRPF, ZMAT2	6.6E–04	hsa03040
RNA degradation	ENO1, PNPT1	LSM3, LSM7, MPHOSPH6	0.025	hsa03018
Peroxisome	NUDT19	DECR2, PEX11B, SCP2, SOD1	0.032	hsa04146
Glutathione metabolism	—	GSTM2, GSTM3, GSTM4, MGST1	0.038	hsa00480
Oxidative phosphorylation	—	ATP5F1, ATP5L, NDUFB6, NDUFB8, NDUFS5, UQCRB	0.044	hsa00190

Table 2-1 Pathway enrichment for proteins determined to change in abundance at a statistical cut-off of p < 0.01

2.5.4 Changes in ribosomal proteins

With ribosomal proteins topping the list from the DAVID analysis, we took a closer look at all of the ribosomal proteins observed in the data. **Figure 2-3A** shows the quantification of ribosomal proteins observed in the control cells and cells expressing the miR-23∼24 cluster, where a general global down-regulation is observed, apart from a few ribosomal proteins such as RpsA and Rpl7L1. This observed general overall reduction in ribosomal proteins (which may reflect an overall reduction in ribosomes) upon miR-23∼24 cluster expression is consistent with the global reduction in the large number of proteins observed (***Figure 2-2***).


A
Small Ribosomal Subunit Proteins
Large Ribosomal Subunit Proteins
Relative Ion Intensity
(Relative to Control)
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
RPSA
RPS10
RPS17L
RPL7L1
RPL22L1
RPLP0
RPL3
Control
miR-23~24

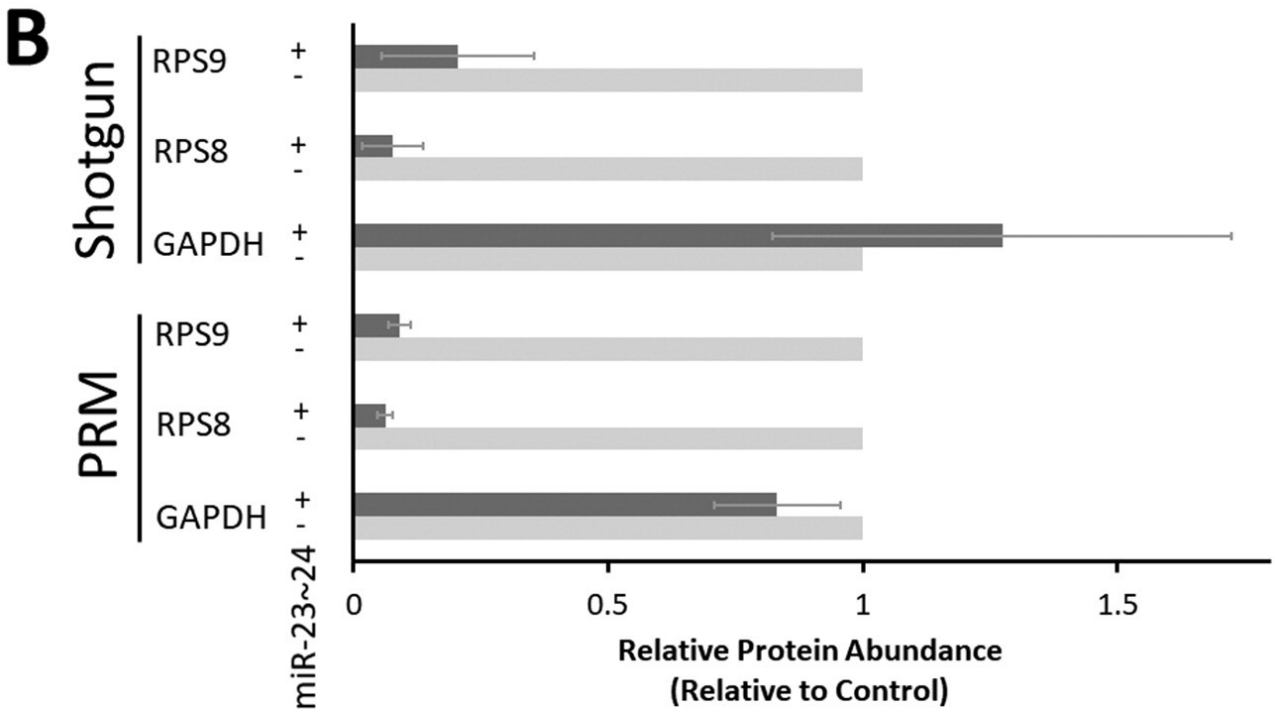

B
Shotgun
PRM
RPS9
RPS8
GAPDH
RPS9
RPS8
GAPDH
miR-23~24
- +
0
0.5
1
1.5
Relative Protein Abundance
(Relative to Control)

Figure 2-3 Global analysis and selected validation of ribosomal proteins.

A. Analysis of all ribosomal proteins detected reveals an overall reduction in both small (left) and large (right) ribosomal subunit proteins upon expression of the miR-23∼24 cluster. In contrast to the global reduction in ribosomal proteins, the indicted ribosomal proteins exhibited increased abundance upon miRNA cluster expression. **B.** Validation of the large fold-changes observed upon expression of the miR-23∼24 cluster with the initial shotgun proteomic analysis by parallel reaction monitoring of the indicated proteins.

2.5.5 Validation of changes in ribosomal protein abundance

With our label-free shotgun proteomic analysis revealing large fold-changes in a number of proteins upon the expression of the miR-23∼24 cluster, some of these changes were validated by an alternative method. For this, we utilized the quantitative approach of parallel reaction monitoring (Gallien, Duriez, et al., 2012) to quantify the changes in abundance of select proteins. The data in **Figure 2-3B**, reveals the quantification of Rps8, Rps9, and GAPDH from both the initial three shotgun analyses and from three PRM analyses of the control cells and cells expressing miR-23∼24. Of methodological note, the PRM analysis was performed on whole cellular lysate and not lysates pre-fractionated by SDS–PAGE. As seen in **Figure 2-3B**, the data between the two analytical methods are in good agreement regarding the magnitudes of change in protein abundance. As predicted, the PRM data resulted in higher precision quantification of these changes.

2.5.6 Comparison of proteomic data with miRNA predicted targets

With the large-scale proteome changes observed upon miR-23~24 cluster expression being a result of both direct targeting by miRNAs and indirect as a result of the miRNA targets regulating the expression of others, a comparison of the proteomic changes observed was made with both the predicted targets and validated targets listed in the Target Scan and miRTarBase bases. For this analysis the proteins that were observed to decrease or increase in abundance in the proteome datasets were determined whether they were targets in either of the databases. The complete comparison is listed in the Supplementary data (Table S4), but the summary of the data in **Figure 2-4** strikingly reveals a minimal overlap of the down-regulated proteins with the miRNA target

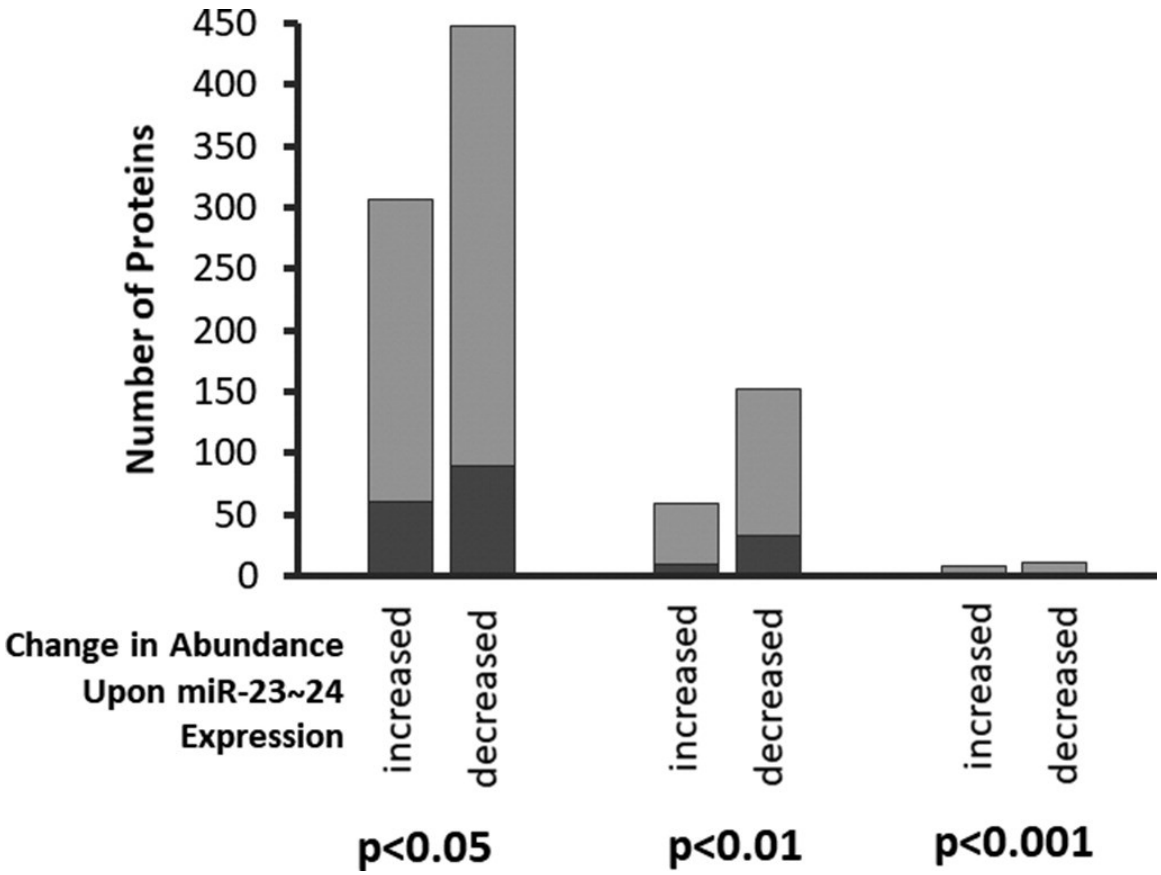


Figure 2-4 Overlap of predicted targets of miR-23~24 with observed proteomic changes.

Dark grey bars indicate the number of proteins identified to be targets of miR23a, miR27a, or miR24-2 in either TargetScan or miRTarBase. Light grey bars indicate the number of proteins observed to change in abundance, but are not predicted targets of any of the three miRNAs. The change in whether protein abundance increased or decreased in abundance upon miR-23~24 expression is indicated. The changes in abundance are grouped according to the statistical confidence of the change in abundance observed. Of note is the lack of a significant correlation of proteins predicted to be targets of the miRNAs and them decreasing in abundance upon miRNA cluster expression.

datasets. In addition, numerous proteins observed to increase in abundance are predicted to be targets of either miR23a, miR27a, or miR24-2.

2.5.7 Top candidate targets of the miR-23~24 cluster

While many of the down-regulated proteins are likely a result of the indirect activity of the miRNAs within the miR-23~24 cluster, the intersection of predicted targets and the down-regulated proteins by proteomic analysis was used to identify candidate targets. While the overlapping still leads to a significant number of candidates, 89 with a p value of <0.05 (**Figure 2-4**), we have also filtered the data for the largest magnitude change in abundance. **Figure 2-5** displays the quantified data for 11 proteins that were either only observed in the vector control cells or exhibited at least a 10-fold reduction in expression, while also exhibiting a p-value < 0.005.

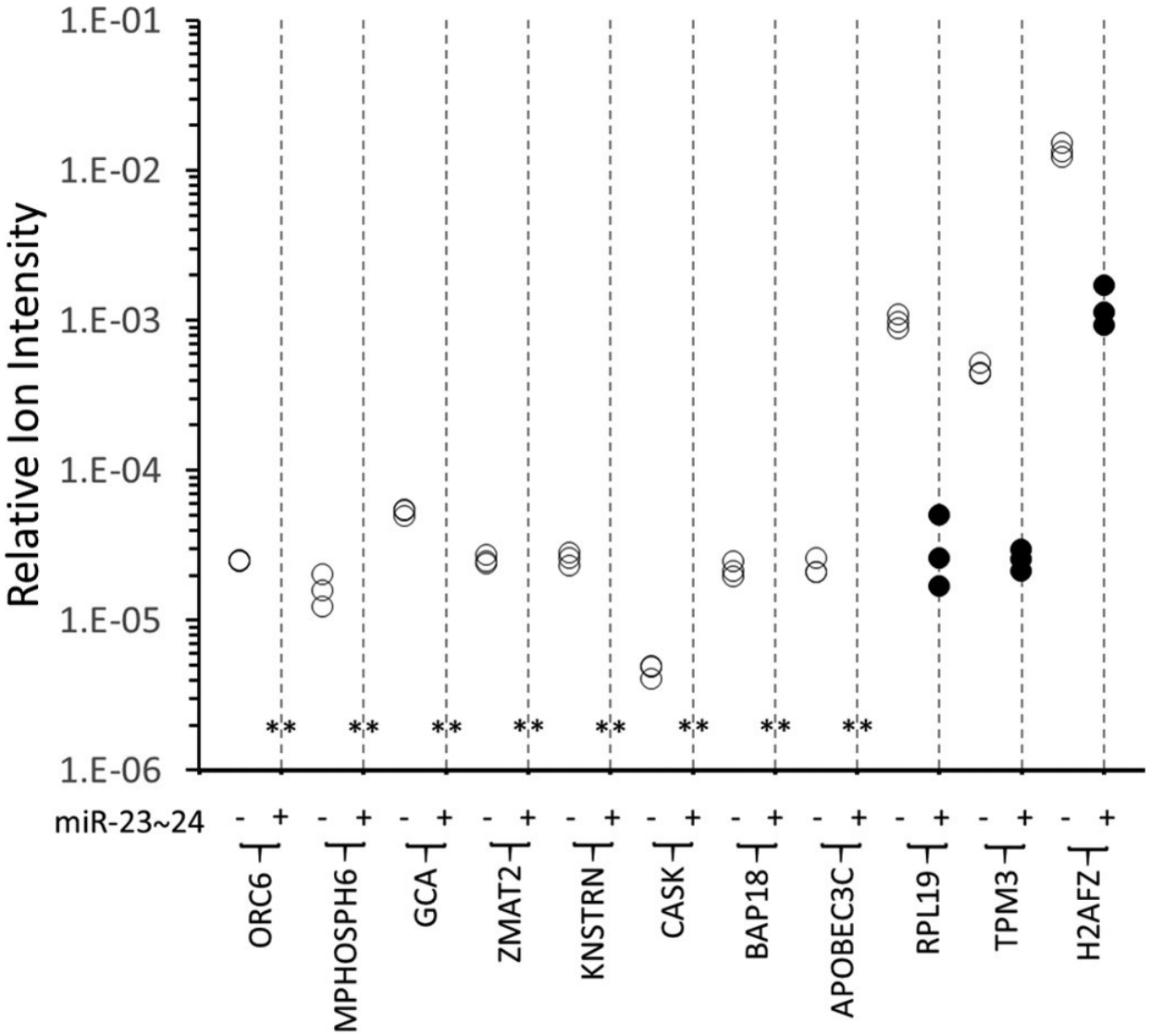


Figure 2-5 Top candidate target genes for the miR-23~24 cluster.

The top candidate potential direct targets of the miRNAs from the miR-23~24 cluster were identified by those were only detected in the control cells or exhibited a greater than 10-fold reduction upon miR-23~24 expression. To minimize this list to the 11 proteins shown, only those with a p value <0.005 are displayed. **, Indicates that the protein was not detected in the cells expressing the miR-23~24 cluster.

2.6 Discussion

Here we have utilized the over-expression of the miR-23∼24 cluster in HEK 293T cells to demonstrate the utility of label-free proteomics to identify large-scale alterations to the cellular proteome. While the expression of genes from a CMV promoter in these cells can lead to artificially high levels of gene expression, in this model it is apparent from the miRNA expression analysis in ***Figure 2-1*** that expression of the miR-23∼24 cluster leads to high levels of mature miRNAs but is comparable to levels observed endogenously in widely utilized breast cancer derived cells lines, such as MDA-MB-231 and MDA-MB-468 cells. These high levels may be relevant in some aspects of breast cancer, because the miR-23∼24 cluster has been linked to the progression of these diseases (Ell et al., 2014; Y. Wang et al., 2017).

In light of a growing concern regarding the over-interpretation of small changes in gene expression often reported regarding mRNA targeting by miRNAs (Seitz, 2017), the identification of the most potent large-scale changes in gene regulation may be the most biologically relevant interactions. The potential questionable significance of modest changes in gene expression as a result of mRNA–miRNA interactions has been brought to question, and as a result, the variability observed in gene expression between individuals can overshadow these small changes (Pinzón et al., 2017). To identify the largest scale changes in protein expression upon the expression of the miR-23∼24 cluster, we have focused on label-free quantitative proteomics using the model system of HEK 293T cells, which do not express the miR-23∼24 cluster at detectable levels (***Figure 2-1***).

While isotopic labeling methods, such as iTraq isobaric tags (Wiese et al., 2007), are widely considered to be superior for quantitative proteomics, technical issues regarding their limitation in detecting large differences in abundance is often overlooked (Saw et al., 2009; H.

Wang et al., 2012). The suppression of the differences quantified by these labeling methods is likely a result of contaminating co-eluting peptides into the collision cell of the mass spectrometer. While modern instruments have sufficient resolution to resolve nearly isobaric masses, they nonetheless suffer from poor resolving power for the isolation of parental ions for MS/MS fragmentation, which is typically conducted by quadrupole. As a result, multiple species may contaminate the collision cell and lead to contaminating reporter ions; an occurrence that will be more significant for low abundance peptides. This results in the compression of the data with lower fold-differences being observed between samples as has been previously reported with head-to-head comparisons of iTraq and label-free proteomic analysis (Latosinska et al., 2015; Trinh et al., 2013). For example, a previous study on the proteome changes upon miR-23∼24 cluster expression reported a maximal ∼2-fold change for the most significant protein alterations (Ludwig et al., 2016). In this previous work, Dhfr and Nolc1 were two of the more significant changes observed upon expression of the miR-23∼24 cluster, with reduced levels of protein of approximately 40% and 60% respectively. While these changes may be significant, our label-free quantification was unable to detect a significant reduction in these proteins. Our data revealed a reduction of ∼30% for both of these proteins through the variability between replicate analysis, which resulted in no statistically significant difference being observed (Supplementary data, Table S1).

2.6.1 Proteomic analysis

While label-free analysis may be insensitive to the quantification of small changes in protein abundance, the analysis of replicate data in ***Figure 2-2*A** reveals sufficient reproducible quantification between analysis. Between these replicates the overall r^2 for all the data was 0.92. Nonetheless, the reproducibility between runs is not exact, with increasing variability for low

abundance proteins, and highlights the need for replicate analysis for reliable label-free proteomic analysis.

The complete comparison of the proteomes of miR-23∼24 cluster expressing, or vector control cells lead to the identification and quantification of 4349 proteins (Supplementary data, Table S1) that range in over five orders of magnitude in abundance, as determined by their relative ion intensity (***Figure 2-2***). The comparison of protein abundance between the samples revealed a large-scale reduction in a large number of proteins. Careful analysis of the volcano plot in ***Figure 2-2*****C** reveals that a large number of proteins are observed to change in abundance by 10- to 100-fold, even if excluding those proteins uniquely observed in one experimental condition. This is in stark contrast to the differences reported by iTraq analysis (Ludwig et al., 2016) and demonstrates the utility of label-free proteomic analysis for the quantification of large changes to the proteome. The statistical comparison of the quantified differences in the proteomes of the miR-23∼24 cluster expressing cells versus the vector control cells identified variations in large numbers of proteins at different levels of statistical confidence (***Figure 2-2*****E**).

With the large numbers of changes in protein abundance determined, ∼750 at a *p* value of <0.05 and ∼200 at a *p* value of <0.01 (***Figure 2-2*****E**), a simple interpretation of the molecular pathways regulated by the miR-23∼24 cluster is not possible. Gene Ontology analysis of the proteins observed to change using a cut off of a *p* value <0.01, revealed an enrichment of the five KEGG groups listed in **Table 2-1**. While this list includes ribosomes, peroxisomes, splicing, and a few metabolic pathways, it is not entirely insightful regarding the specific molecular function of the miR-23∼24 cluster; however, it is in agreement with previous reports linking the miRNAs from this cluster to metabolism (Gao et al., 2009; Kelly et al., 2015). Surprisingly, within the groups of protein populating the functions listed in **Table 2-1**, most are not identified or predicted

targets of miR23a, miR27a, or miR24-2, with the exception of a couple proteins like Rpl19, which is a predicted target of miR27a (Chou et al., 2018). This lack of association suggests that many of the observed changes are a result of indirect effects of the activities of the miRNAs from the miR-23~24 cluster. The occurrence of the regulation of broad pathways by this cluster are not surprising given diverse impact of this miRNA cluster in cell differentiation (Cho et al., 2016; Kurkewich et al., 2017) and disease (Bang et al., 2012; Ell et al., 2014; Y. Wang et al., 2017).

To further query the top group identified in **Table 2-1**, ribosomes, we then went on to investigate the data obtained from all the ribosomal proteins detected and not just those that met the statistical threshold of $p < 0.01$. This secondary analysis was revealing in that it is suggestive of a general global down-regulation of ribosomal proteins, with the exception of a few ribosomal proteins that are observed to increase in abundance (**Figure 2-3**). While the increase in abundance of select ribosomal proteins may at first be counter intuitive with regards to a model for a global down regulation in ribosome biogenesis, the individual increase in select ribosomal proteins may reflect their specific or alternative roles. For example, while RpsA is a known small ribosomal subunit protein (Malygin et al., 2011) it is also known to have extra-ribosomal functions as a laminin receptor (Scheiman et al., 2010) so its observed increase in abundance may be related to this function.

2.6.2 Target identification

While proteomics analysis reveals the global changes to the proteome upon expression of the miR-23~24 cluster, it does not provide insight regarding which changes may be a direct result of miRNA targeting. Direct target identification and validation has demonstrated to be challenging, with current reports having to focus on one or two targets of a miRNA in a single study. For

example, a recent report on miR27a has identified both interleukin-10 and TGF-β-activated protein kinase 1 binding protein as targets of this miRNA (Hussain et al., 2018). A challenge in the identification of miRNA targets is that prediction algorithms, such as those used by TargetScan (Agarwal et al., 2015) predict enormous numbers of candidates, which needs to be validated on a case by case scenario as the success rate for identification can not be reliably estimated as a result of a lack of negative results being published in the literature. Within our own data set, the predictions almost appear random (**Figure 2-4**), which is in agreement with previous reports where minimal overlap of target prediction and proteomic changes are observed (T. Lee et al., 2015). When classifying the up- and down-regulated proteins observed upon miR-23∼24 expression, the percent of predicted targets in both groups are similar. This is not to say that these prediction algorithms are not of high utility, but one must be cautious regarding the interpretation of pure bioinformatic analysis of miRNA targeting (Chhabra et al., 2010). The utility of mRNA target prediction algorithms and bioinformatic analysis are of greatest utility when combined with biological data to identify key target candidates. For example, this cross-platform analysis was demonstrated successfully with the identification of the miR27a-calreticulin signaling axis (Colangelo et al., 2016).

With our focus on identifying the most potent mRNA targets of the miRNAs derived from the miR-23∼24 cluster, our proteomics data was cross-correlated with the miRNA targets from miRTarBase (Chou et al., 2018) and TargetScan (Agarwal et al., 2015). As mentioned, the overlap of predicted miRNA targets and our quantified proteomic changes is marginal (**Figure 2-4**); of the 448 down-regulated proteins, at a p value of <0.05, only 89 are predicted targets of either miR23a, miR27a, or miR24-2. As not all of the alterations to the proteome are likely to be a result of direct regulation by miRNAs, this finding is somewhat expected. In contrast to this, of the 307 proteins

observed to increase in abundance upon the expression of the miR-23∼24 cluster, 60 are predicted to be targets of one of the three miRNAs. At first glance the somewhat random occurrence of miRNA predicted targets and cellular changes to the proteome lead to what may be considered a poor result regarding the predictive power of identifying miRNA targets using these algorithms. Nonetheless, in combination with biological data the results are still very promising, because the intersection of the datasets reveals the most promising candidate genes targeted by the miRNAs from the miR-23∼24 cluster.

In returning to the original question regarding the identification of candidate miRNA targets that lead to large reduction in expression by a miRNA, our data was cross-referenced with the miRTarBase and TargetScan databases for miRNA targets to identify proteins that met the following criteria. Proteins that either were reduced in abundance by at least 10-fold or were only detected in the vector control cells, exhibited a p value of <0.005 for the t test between the control and miR-23∼24 replicates, and were predicted to be targets of one of the three miRNAs. These very restrictive constraints lead to the identification of the 11 proteins listed in **Figure 2-5**. While most of these proteins, such as Orc6 have not been previously reported to be associated with any of the miRNAs from the cluster, grancalcin (GCA) has been reported previously to be regulated by miR27a (Schoolmeesters et al., 2009). For future investigations on this cluster, these proteins are top candidates for regulation. If the criteria are relaxed to a p value <0.01 for example, but keeping the criteria of the 10-fold change, the number of protein candidates jumps to 25 (Supplementary data, Table S3).

2.7 Concluding remarks

Here we report on the utilization of label-free proteomics to characterize the global changes to the cellular proteome upon the expression of the miR-23∼24 cluster and reveal the utility of this approach to identify the most promising potent targets of the miRNAs originating from this cluster. While label-free proteomics is traditionally not viewed to be as quantitative and more problematic than isotopic labeling techniques, such as iTraq, we demonstrate that it is of high utility in identifying the largest magnitude changes, which are desired for the identification of the most potent targets for miRNA regulation. While our analysis for identifying the most significant inhibition by miRNAs from the miR-23∼24 cluster revealed a previously identified target, our data reveals an additional series of previously unreported potential targets that exhibit reduced expression by over an order of magnitude upon expression of the miR-23∼24 cluster. This is in contrast to previous proteomic investigations on this cluster using iTraq labeling, which only reported on changes of less than 2-fold in protein abundance.

2.8 Future Direction

Our findings identified many previously unreported targets that exhibit reduced expression levels, several of them were downregulated by more than an order of magnitude. These targets needs to be investigated to determine if they are direct targets of the miR23~24 cluster or their reduced expression level is due to indirect effect of the overexpression of the cluster. The functional relevance of the targets could also uncover potential pathways or signaling network that could be affected by miR23~24 cluster. Since higher expression of this microRNAs have been shown to

mimic several breast cancer cell lines, these proteins that show great decrease in expression could be tested as potential biomarkers for breast cancer. There are still many unknowns related to microRNA activities and targeting in vivo that still needs to be investigated as our investigations into miR23~24 cluster overexpression have shown. Other microRNA clusters could also have similar behaviour that have been left unreported previously.

Chapter 3:

Potential Biomarker Identification Through Serum Proteomics for Myasthenia Gravis

3.1 Contribution and Acknowledgement

This chapter is a manuscript prepared for publication titled "Residual Serum Fibrinogen as a Universal Biomarker for all Serotypes of Myasthenia Gravis" with the following authorship:

Faraz S. Hussain[1,5], PhD, Ramanaguru S. Piragasam[2,5], Hassan Sarker[2], Derrick Blackmore[1], PhD, Elaine Yacyshyn[3], MD, Carlos Fernandez-Patron[2], PhD, Richard P. Fahlman[2,4*], PhD & Zaeem A. Siddiqi[1*], MD, PhD.

[1] Division of Neurology, Department of Medicine, Faculty of Medicine & Dentistry, University of Alberta. [2] Department of Biochemistry, Faculty of Medicine & Dentistry, University of Alberta. [3] Division of Rheumatology, Department of Medicine, Derrick, University of Alberta. [4] Department of Oncology, Faculty of Medicine & Dentistry, University of Alberta. [5] FSH and RSP contributed equally to this work.

Dr. Faraz Hussain was responsible for study design, sample preparation, immunological assays, and data analysis. I was responsible for mass spectrometric sample preparation, design of mass spectrometric experiments, discovery and targeted proteomics data collection and data analysis. Dr. Richard Fahlman contributed to study design and analysis while Dr. Zaeem Siddiqi contributed to study design and sample collection. Hassan Sarker from Dr. Carlos Fernandez-Patron's team performed the ELISA experiment and the data analysis. Dr. Derrick Blackmore obtained patients consent, collected and processed serum samples and provided relevant clinical data. Dr. Elaine Yacyshyn provided clinical data and serum samples of Rheumatoid arthritis patients.

This research has been funded by the generosity of the Stollery Children's Hospital Foundation and supporters of the Lois Hole Hospital for Women through the Women and Children's Health Research Institute (WCHRI) and University Hospital Foundation. This project was additionally supported by a Discovery Grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada to Dr. Richard Fahlman.

3.2 Abstract

Nonspecific symptoms and a lack of a universal molecular biomarker hinders timely diagnosis and treatment of the autoimmune disease Myasthenia Gravis (MG). Molecular diagnosis of MG by the presence of serum auto-antibodies is only confirmed in ~85% and ~50% of patients with generalized and ocular MG, respectively. Comparing the serum proteomes of MG patients to both healthy controls (HC) and patients with Rheumatoid Arthritis (RA) revealed that residual fibrinogen is universally detected in MG patient sera. Follow-up investigations quantifying the α-, β-, and γ-subunits of fibrinogen in 79 individual sera (31 MG of various serotypes, 18 RA and 30 HC) revealed fibrinogen to be highly specific for MG ($p<0.00001$), with an average higher abundance of >1000-fold over the control groups. Our unanticipated discovery of high levels of residual serum fibrinogen in all MG patients offers an unmatched opportunity in the diagnosis of MG.

3.3 Introduction

Myasthenia gravis (MG) is a humoral autoimmune disorder in which autoantibodies directed against neuromuscular junctions (NMJ) proteins affect the electrical signal transmission across the NMJ resulting in variable weakness of voluntary muscles ranging from mild ocular and/or limb muscle weakness to fulminant life threatening myasthenic crises due to weakness of swallowing and breathing muscles (Gilhus et al., 2019). As a result of the nonspecific symptoms, there still remains reported challenges concerning diagnosis in the emergency room (Abbott, 2010; Spillane et al., 2012) and in elderly populations with comorbid illness (Aarli, 2008; Kukulka et al., 2019).

The prevailing thought is that autoantibodies in MG are directed against the nicotinic acetylcholine receptor (AChR) at NMJ, which is observed in about 85% of patients (Vincent & Newsom-Davis, 1985). A smaller proportion of patients have autoantibodies against other NMJ proteins, including the muscle specific tyrosine kinase (MuSK) or low-density lipoprotein receptor-related protein 4 (LRP4) (Higuchi et al., 2011; Hoch et al., 2001). No antibodies can be detected in about 5-10 % of patients with generalized MG and about 50% with ocular MG, using the current assays (seronegative MG) though such patients manifest clinical features and therapeutic responses similar to those with detectable autoantibodies. Although relatively specific for the diagnosis and subgrouping of MG (Gilhus & Verschuuren, 2015), the serum level of anti-AChR, anti-MuSK or anti-LRP4 antibodies do not correlate with the disease course or treatment outcomes. Limitations of these current molecular diagnostic methods are particularly exemplified with the number of seronegative patients with ocular form (Class I) of this disease (Peeler et al., 2015).

The identification of robust serological biomarkers in MG has previously been attempted. In one study, the blood levels of proliferation-inducing ligand (APRIL) and several cytokines, such as IL-19, IL-20, IL-28A and IL-35, were found to be upregulated in the sera of MG patients as compared to controls (Uzawa et al., 2016). A second study has reported correlations between MG and the serum levels of matrix metalloproteinase 10 (MMP-10), transforming growth factor alpha (TGF-α) and receptor for advanced glycation end-products binding protein (protein S100-A12) (Molin et al., 2017). Most recently, high κ free light chain has been reported for seropositive and seronegative forms of MG (Wilf-Yarkoni et al., 2020).

Although these studies provide promising results, the limited dynamic range (less than 2-fold) of the proteins between controls and individuals with MG leads to significant challenges for their utility in robust diagnostic testing. Additionally, it is not clear whether the elevation of serum inflammatory proteins in MG patients is disease specific or represents a non-specific, general increase in the inflammatory mediators expected in autoimmune disease. Notably, patients with MG have increased risk of having another autoimmune disorder, with about 13-22% having a second autoimmune disorder (Fang et al., 2015; Nacu et al., 2015). In line with these challenges, a recent report has identified a panel of five serum metabolites, which include three lysophospholipids, glyceric acid and 12-ketodeoxycholic acid, that are reported to differentiate between MG and another autoimmune disease Rheumatoid arthritis (RA) (Blackmore et al., 2019). However, again the limited dynamic range of these metabolites, of less than 2-fold, impairs their utilization for diagnostic testing.

To circumvent some of the current challenges in MG diagnosis, the present study aimed to identify serum proteomic biomarker that may be universally sensitive for all classes and serotypes of MG and exhibit specificity with regards to related diseases. To this end our study design

includes sera from a heterogeneous cohort of MG patients that was not only compared to a normal control group but also to a reference autoimmune disease, RA. The choice of RA as the reference autoimmune disease is based on the observation that like MG, RA also has a major humoral component, and the two diseases may co-exist in about 2-4% of MG patients (Mao et al., 2011; Thorlacius et al., 1989). We postulated that the comparison of the serological protein profiles from the two diseases and controls may potentially identify specific proteins in serum that are unique to MG.

3.4 Methods

3.4.1 Ethics Approval.

The sera samples for the current study were obtained from Canadian BioSample Repository (CBSR), collected and stored in a de-identified manner by Dr. Siddiqi's MG study group. Approval to use the sera, to investigate potential biomarker(s), was obtained from the Research Ethics Board, University of Alberta (Project Title: Metabolomic Profiling of Serum in Myasthenia Gravis: Pilot Study, PRO00030698, 24th May 2012).

3.4.2 Serum Samples.

For the study three groups of samples were identified: Myasthenia gravis (MG), Rheumatoid Arthritis (RA) as a reference disease and normal controls. A total of 79 samples were recruited for the study: 31 MG, 18 RA and 30 controls. The demographics of the individuals are summarized in ***Table 3-1***. MG and RA serostatus was confirmed with antibody testing for either anti-AChR/anti-MusSK (MG) or RA. RA patients were diagnosed in accordance with the American Rheumatology Association 1987 criteria (Arnett et al., 1988). To exclude the confounds of race, only Caucasian patients were included in this study. There were no smokers and no statistically significant differences between all groups from time of last meal or BMI. Further, patients had no history of any other autoimmune disease or thymoma. Finally, due to the nature of recruitment, patients were not required to fast. Clinical patients and healthy controls were enrolled in a prospective observational trial to obtain serum. MG and HC were collected within the same clinic. RA samples were collected in multiple clinics.

For collection, blood samples were drawn from the antecubital vein using a 21 G needle and vacutainer red top no additive tubes (Becton Dickenson).

3.4.3 SDS-PAGE, Western Blot Analysis & ELISA.

SDS-PAGE loading buffers contained 100 mM of β-mercaptoethanol and 4% SDS. 5 µl of samples were resolved by 10% SDS-PAGE and either visualized by Coomassie R250 staining or transferred to a nitrocellulose membrane for Western blot analysis as previously described (Eldeeb & Fahlman, 2016). The membranes were then immunoblotted with either a mouse anti-FGA antibody(A-6) (SC-166968; Santa Cruz Biotechnology) derived against residues 750-850 or a rabbit anti-FGA antibody (ab92572, Abcam) derived against residues 21-320. Secondary blotting was achieved with either goat anti-Mouse or donkey anti-Rabbit IRDye 680RD labeled antibodies (LI-COR).

Coomassie stained gels and Western blots were visualized with a LiCOR Odyssey Fc system.

Serum fibrinogen was quantified using an enzyme-linked immunosorbent assay (ELISA) was performed with a kit (ab208036, Abcam) following the manufacturer's recommended procedure.

3.4.4 Serum Protein Profile by Gel-LC-MS/MS.

18 randomly selected samples, 6 from each group, were used for the initial shotgun proteomic analysis. Sample was resolved by 10% SDS-PAGE and visualized by Coomassie staining. Each lane was excised into 26 bands, then each band was subjected to in-gel tryptic digestion as previously described (Khan et al., 2015). The resulting digested peptides were vacuum dried and re-suspended in solvent A (5% Acetonitrile (ACN) in 0.2% formic acid) for LC-MS/MS analysis. LC-MS/MS analysis was performed on an Thermo Scientific EASY-nL 1000 system inline Q-Exactive Hybrid Quadrupole-Orbitrap Mass Spectrometer using identical parameters as previously described (Piragasam et al., 2020) with the alteration of running a 75 min gradient.

3.4.5 MS/MS Data Analysis.

Raw data files of MS/MS spectra for each sample were combined as searched with Proteome Discoverer 1.4.1.14 (Thermo Fisher Scientific, US) against the non-redundant and reviewed proteome of Homo sapiens retrieved from UniProt. The search parameters were as previously described (Kramer et al., 2017). All of the identified proteins from each sample along with their associated extracted ion intensities quantified for abundance are listed in Supplementary Tables 1 and 2. The MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Vizcaíno et al., 2014) repository with the dataset identifier PXD019633.

3.4.6 Quantification and Statistical Analysis.

EICs were used to quantify the protein abundance in individual samples, for which the values were subjected to normalization by the proportion to total ion current (TIC) as previously described (Kramer et al., 2017). One-way ANOVA was employed for comparative analysis of the three sample groups, using a statistical cut-off of a $p<0.05$. All the data used for statistical analysis is provided in Suplementary Table 3.

3.4.7 Parallel Reaction Monitoring.

Serum proteins were resolved by 8% SDS-PAGE then prepared for LC-MS/MS as described above. The resulting fractions from each sample were then pooled prior to LC-MS/MS analysis. Multiple peptides per protein (***Table 3-2***) were included for method building and final quantification using the Skyline software package (MacLean et al., 2010) were implemented as described previously (Piragasam et al., 2020). The raw quantified data from PRM analysis is provided in Supplemental Table 5.

3.5 Results

3.5.1 Myasthenia Gravis Serum Proteome Profiling.

To identify novel proteins in an unbiased manner that may be specifically associated with MG, a pilot study using a gel-based shotgun proteomics method (Gel-LC-MS/MS) was applied to the analysis of unadulterated serum samples collected from six healthy controls, six MG patients and six RA patients, as outlined in ***Figure 3-1A***. As summarized by the Venn diagram in ***Figure 3-1B***, a total of 502 proteins were identified with the criteria of two or more quantifiable peptides per protein to limit the false discovery rate (FDR) of protein identifications. Within the MG group 361 proteins were identified, while for the control and RA groups, 327 and 304 were identified respectively. Of these proteins, 212 were common to all groups whereas the number of unique and additionally shared numbers of proteins for each group are indicated in ***Figure 3-1B***.

Myasthenia Gravis	**n=31**
Average Age (years)	59.0 ± 23.9(range 19 – 93)
Gender (M/F, %)	51.6% / 48.4%
Average On-set Age	54.7(range 15 – 86)
Early/Late On-set (%)	35.5% / 64.5%
Disease Severity (MGFA)	
Class I	25.8%
Class II	32.3%
Class III	41.9%
Control	**n=30**
Average Age (years)	48.2 ± 17.1(range 21-86)
Gender (M/F, %)	40/60
Rheumatoid Arthritis	**n=18**
Average Age (years)	65.1 ± 11.7(range 36-81)
Gender (M/F, %)	33.3/66.7

Table 3-1 Cohort Demographics

Demographic information and clinical profiles of the three groups of individuals from which sera were collected for fibrinogen protein analysis. As per the Myasthenia Gravis Foundation of America (MGFA) classification, Class I represents purely ocular form of MG, Class II represents mild generalized MG and Class III describes moderately severe generalized MG.

3.5.2 Identification of Proteins of Differential Abundance.

The ratios of the average relative extracted ion intensity for each protein compared among all three test groups are shown in the tri-plot in ***Figure 3-1C***. The clustering of the data points in the centre of the graph reveals that the majority of average ion intensities for each protein is similar between all three groups, but that some proteins are higher in some of the groups with the data being closer to the triangle apex from the respective group. As some average ion intensities maybe biased as a result of highly variable measurements from individuals in a group, either from technical variability in quantification or true variability of a protein in an individual, all the data was analyzed using a one-way ANOVA to identify proteins that exhibit statistically significant differences between the MG, RA and control groups. The proteins that met the criteria of a p-value <0.05 are depicted as yellow circles in the tri-plot. While this cut-off may lead to false-positive identifications as a result of the multiple testing problem it does warrant further investigation of these identified proteins.

The individual measurements of the 11 proteins that exhibited statistically significant differences between the MG, RA and control groups are summarized in the heat map shown in ***Figure 3-1D*** along with Apolipoprotein A1 (APOA1) and the constant region of immunoglobulin heavy chain (IGHG2) as controls. In addition, the corresponding ANOVA F-statistic and p-values are also listed for each protein. The data reveals that fibrinogen-α (FGA), fibrinogen-β (FGB), fibrinogen-γ (FGG), keratin, type II cytoskeletal 78 (KRT78), ribonuclease 4 (RNASE4) and

prothrombin (F2) are consistently in higher abundance in MG serum. Serum paraoxonase/arylesterase 1 (PON1), L-lactate dehydrogenase B (LDHB), and sex hormone-binding globulin (SHBG) are consistently high in RA sera, while complement C1q subcomponent subunit B (C1QB) and insulin-like growth factor-binding protein complex acid-labile subunit (IGFALS) are low in both MG and RA sera.

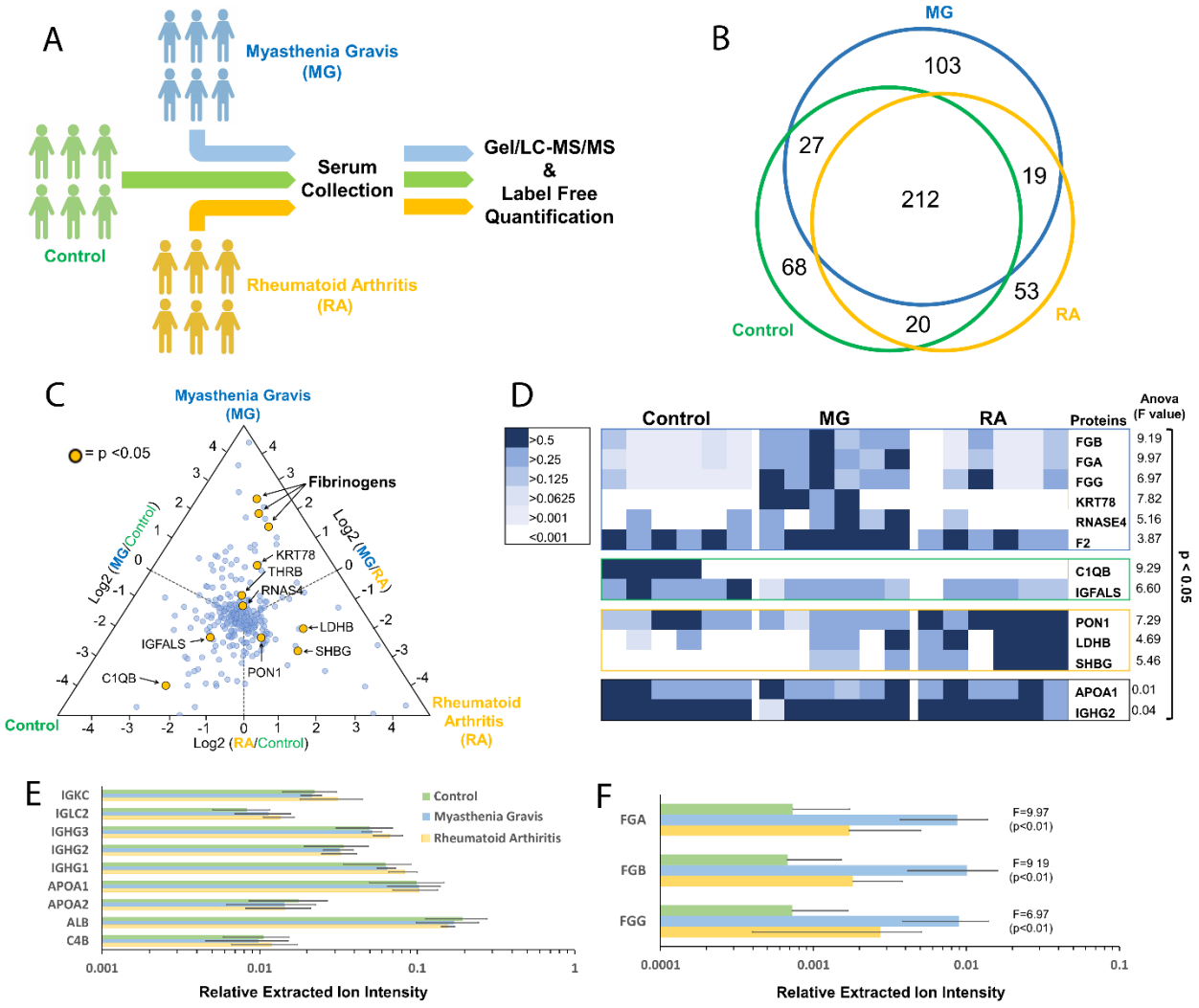


Figure 3-1 Clinical Proteomic Analysis of Myasthenia Gravis (MG) Patients.

A. Schematic workflow for the label free proteomic analysis of serum collected from 6 MG patients, 6 control individuals and 6 with the related disease, Rheumatoid Arthritis (RA). **B.** Venn diagram summarizing the proteins identified and quantified among the three test groups. **C.** A three-way triplot of differential protein abundance observed between MG, control and RA samples. The log2 ratio of the normalized protein abundance was plotted to visualize differential amounts between the three groups. In this plot, proteins observed in higher abundance in one group over others are observed closer to the respective vertices. Proteins that exhibit reproducible differences as revealed by a one-way ANOVA analysis ($p<0.05$) are highlighted in yellow and labelled. **D.** Heatmap representation of the abundance of statistically significant proteins with their respective calculated ANOVA F-statistic shown ($p<0.05$). In addition, two control proteins are shown for comparison. **E.** Validation of label-free quantification by investigating the variability of the measurements between groups for a series of serum proteins, including immunoglobulins (IGKC, IGLC2, IGHG3, IGHG2 & IGHG1); apolipoprotein A (APOA1, APOA2); serum albumin (ALB); and C4b-binding protein alpha chain (C4B). **F.** Relative average abundance level of fibrinogen-α (FGA), fibrinogen-β (FGB) and fibrinogen-γ (FGG) among all sample groups.

3.5.3 High Fibrinogen in MG Sera.

To narrow our focus, proteins consistently exhibiting the largest observed differences were selected for further investigation. As a control for the quantification by extracted ion intensities, the relative extracted ion intensities for a series of control are shown in ***Figure 3-1E***. The quantified data for the control proteins, including several immunoglobulins, apolipoproteins, albumin and C4b-binding protein alpha chain, reveal the spread in the quantified data for all three groups (MG, RA, and controls). For the proteins observed in differential abundance, all three fibrinogen subunits (FGA, FGB and FGG) exhibited the largest differential abundance in MG

patient sera. As seen in ***Figure 3-1E***, the amounts of these fibrinogens were consistently higher in abundance by over an order of magnitude compared to controls ($p<0.01$) and were thus selected for further biochemical validation.

3.5.4 Validation of High Fibrinogen in MG Patient Sera.

To validate our observations of high serum fibrinogen levels in MG patients, serum samples from MG, RA and control individuals, were analyzed for Fibrinogen-α by immunoblotting (***Figure 3-2***). To ensure equal protein loading for analysis, samples were normalized to total protein loading as visualized by Coomassie Blue staining (***Figure 3-2A*** *upper panel*). Immunoblotting for Fibrinogen-α (***Figure 3-2A*** *middle panel*) reveals a very high protein amounts in MG sera, but is essentially undetectable in either the RA or control samples. The epitope for this anti-Fibrinogen-α antibody is mapped to the amino acid residues 750-850, so this large differential detection of Fibrinogen-α was further validated by using an anti-Fibrinogen-α antibody derived against the residues 21-320 (***Figure 3-2A*** *lower panel*), that resulted in essentially the same results in high levels only being observed in sera from MG patients. This second antibody for the N-terminal region of fibrinogen did reveal altered apparent gel mobilities for Fibrinogen-α, suggesting that proteolytic processing is occurring that was not apparent by the analysis with the C-terminal specific antibody. Of note, the C-terminal specific antibody was used in all additional immunoblotting experiments.

This observed high abundance of fibrinogen is peculiar as to our knowledge there have been no previous reports of high fibrinogen in MG patients, nor is there any clotting disorder typically associated with the disease. As a result, we then investigated Fibrinogen-α amounts in the plasma of MG and controls by immunoblotting (***Figure 3-2B***). The analysis revealed no

significant difference in Fibrinogen-α in MG plasma when compared to a control sample. To investigate whether Fibrinogen-α was being removed during the collection and preparation of patient serum, both the plasma and serum of an MG patient and a control sample were analyzed simultaneously (***Figure 3-2C***). What is observed is that when analyzing plasma, the Fibrinogen-α in serum is undetectable as a result of the limited dynamic range of Western blotting. Only when analyzing serum alone, are we able to observe a high level of residual Fibrinogen-α in the serum of MG patients (***Figure 3-2A***).

As a result of the limited dynamic range of detection by immunoblotting, we conducted an ELISA assay for total fibrinogen. Due to the large differences in fibrinogen detected, the MG serum samples required dilution in order for them to be on the same linear range of detection for the assay when comparing to control samples. The resulting data from the ELISA is shown in ***Figure 3-2D*** and like the proteomic analysis, reveals over an order of magnitude difference when comparing total fibrinogens in MG and control sera.

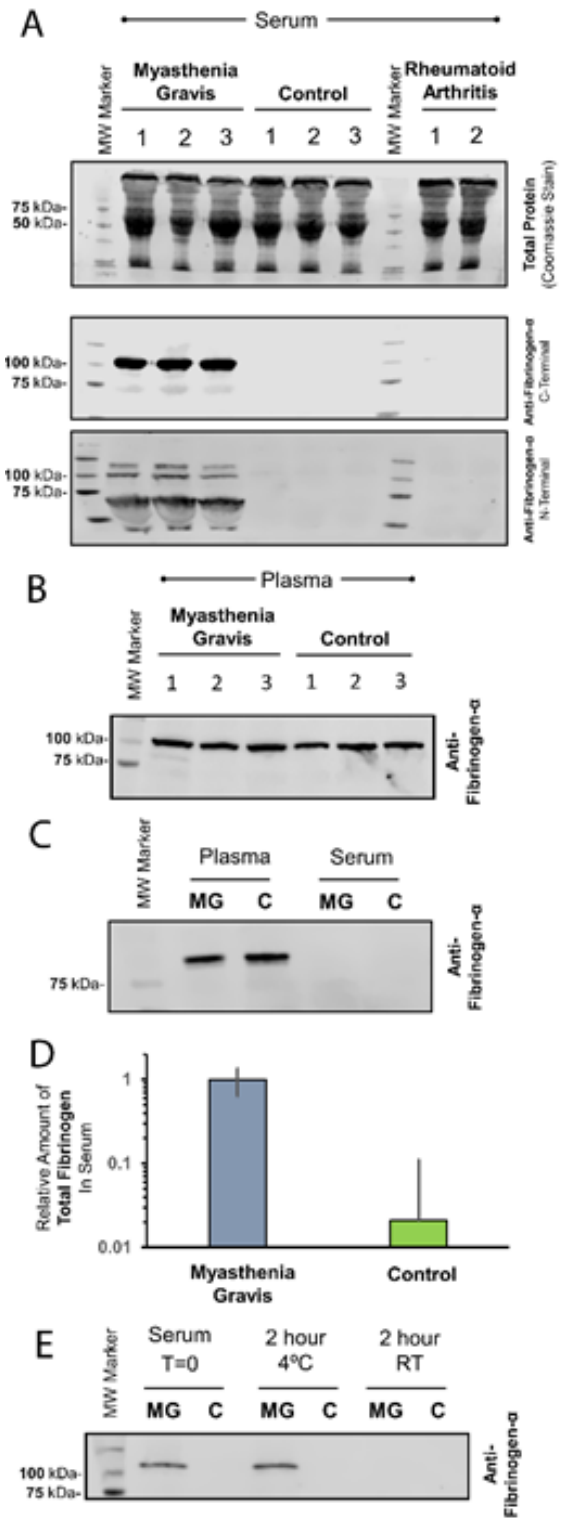

A
Serum
MW Marker
Myasthenia Gravis
Control
MW Marker
Rheumatoid Arthritis
1 2 3 1 2 3 1 2
75 kDa-
50 kDa-
Total Protein (Coomassie Stain)
100 kDa-
75 kDa-
Anti-Fibrinogen-α C-Terminal
100 kDa-
75 kDa-
Anti-Fibrinogen-α N-Terminal
B
Plasma
MW Marker
Myasthenia Gravis
Control
1 2 3 1 2 3
100 kDa-
75 kDa-
Anti-Fibrinogen-α
C
MW Marker
Plasma
Serum
MG C MG C
75 kDa-
Anti-Fibrinogen-α
D
Relative Amount of Total Fibrinogen In Serum
1
0.1
0.01
Myasthenia Gravis
Control
E
MW Marker
Serum T=0
2 hour 4°C
2 hour RT
MG C MG C MG C
100 kDa-
75 kDa-
Anti-Fibrinogen-α

Figure 3-2 Immunochemistry Detection of Residual Fibrinogen in Serum.

A. (*Upper*) Total protein loading of patient serum in a SDS-PAGE gel visualized by Coomassie staining. Analysis of the same protein loading of samples analyzed by Western blotting with anti-Fibrinogen-α antibodies derived against the C-terminal 750-850 residues (*middle)* or the N-terminal 21-320 residues (lower) of Fibrinogen-α. **B.** Western blot analysis for Fibrinogen-α in plasma does not reveal observable differences between MG or control samples. **C.** A comparative analysis of Fibrinogen-α in both plasma and serum samples from MG and control individuals reveals that the residual Fibrinogen-α observed in the sera of MG patients (**A**) is below the dynamic range of detection by Western blot analysis when compared to plasma Fibrinogen-α. **D.** ELISA was used to quantify the difference in total fibrinogen in MG or control serum samples. Of note, in order to compare MG and control samples by ELISA, the MG samples were diluted prior to analysis to achieve a signal in the linear response range for the method. The resulting dilution factor is incorporated in the normalized values presented. **E.** Stability of residual Fibrinogen-α after initial processing (T=0) was investigated by Western blot analysis after incubating the serum sample for an additional two hours at either 4 °C or room temperature.

3.5.5 Residual Sera Fibrinogen is Stable at Low Temperature.

To investigate the stability of the observed residual fibrinogen in the MG patient sera, the presence of fibrinogen was investigated upon prolonged incubations. Freshly thawed MG and control serum samples (T=0) were further incubated for 2 hours at either 4°C or at ambient room temperature (~20°C). Fibrinogen in the samples was then detected by immunoblotting with an anti- Fibrinogen-α antibody. As shown in ***Figure 3-2E*** the residual fibrinogen persisted at low temperature but was no longer detectable upon incubating at room temperature, indicating the transient nature of the residual serum fibrinogen.

3.5.6 Blinded Analysis of Residual Serum Fibrinogen Identifies MG Patients.

Prior to investigating a large cohort of patients, a blinded study was initiated to verify whether the correlation of serum fibrinogen and MG held true with a new set of blinded samples. A total of 10 blinded sera samples (A-J) were obtained from the MG clinic and analyzed for Fibrinogen-α by Western blotting. The resulting analysis revealed protein bands corresponding to Fibrinogen-α for samples A, B, D, H and I (***Figure 3-3***). Unblinding of the clinical data revealed a 100% identification rate for MG patients from normal controls (patients C, E, F, G & J).

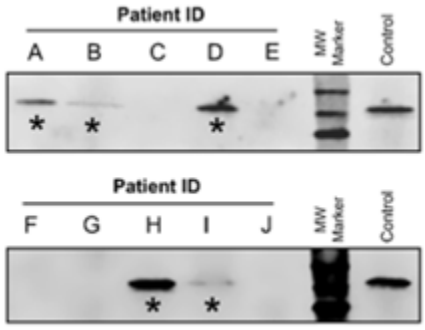


Figure 3-3 Blinded Investigation of MG Serum Samples.

Ten blinded clinical serum samples from either controls or MG patients (A-J) were analyzed by Western blot analysis for residual Fibrinogen-α with serum from a MG patient used as a control. The serum samples A, B, D, H and I all exhibited detectable Fibrinogen-α upon analysis which exactly corresponded to the five samples from patients with MG (*).

3.5.7 Cohort Analysis by Targeted Mass Spectrometry for Residual Serum Fibrinogen.

With the promising findings of the initial proteomic study with regards to the correlation of fibrinogens (***Figure 3-1***) and MG diagnosis along with the result of the blinded study with fibrinogen-α (***Figure 3-3***) we then conducted a validation study on a larger patient cohort consisting of sera from 31 MG patients (27 anti-AChR antibody positive, one anti-MuSK antibody positive, three seronegative), 18 from RA patients and 30 controls. The demographic information and clinical profiles for this cohort are outlined in ***Table 3-1***.

To circumvent the challenges of the limited dynamic range of detection observed for Western blot and ELISA assays, the serum samples were digested with trypsin and simultaneously analyzed for the tryptic peptides derived from fibrinogen-α, fibrinogen-β, fibrinogen-γ and serum albumin (as a control) by parallel reaction monitoring (PRM). For PRM analysis of the three peptides for fibrinogen-α, four peptides fibrinogen-β, and five fibrinogen-γ targeted for quantification are listed in ***Figure 3-4A***. The four peptides for the serum albumin control are listed in the ***Table 3-2***. For the present study, extracted ion chromatograms (EICs) of minimum five co-eluting fragment ions were measured for the quantification for each peptide, while being verified by aligning the retention time with the precursor ion for the given peptide to eliminate any possible mass spectrometry artifacts and quality assurance of the identified fragment ions. Samples raw data for the EIC of the fibrinogen-α derived DSHSLTTNIMEILR peptide is shown in ***Figure 3-4B***. The quantified abundance of each protein was the sum of EICs of corresponding peptide fragment ions normalized to Serum Albumin (***Figure 3-4C***).

The quantification of fibrinogen-α, -β and -γ in the 79 serum samples by PRM analysis, shown in the Beeswarm plots in ***Figure 3-4D***, clearly reveals an essentially equally high abundance of all three fibrinogens in MG patient sera when contrasted to either control or RA

serum samples. Of note, the sera from three seronegative and one anti-MuSK MG patients were indistinguishable from the other MG patients with regards to all three fibrinogens. While the majority of control and RA samples exhibited a significantly lower abundance of the fibrinogens, few individuals did reveal high fibrinogen levels. Within the dynamic range of detection of about five orders of magnitude with the PRM analysis, the median values for the three fibrinogens in MG samples were 28 to 54-fold higher than the RA samples and 250 to 4000-fold higher than the control samples. The MG sera revealed about a 10-fold variation among individuals for the three fibrinogens, while the controls and RA sera exhibited about a 1000-fold variation. Nonetheless, one-way ANOVA analysis revealed high degrees of statistical confidence for fibrinogen-α, fibrinogen-β and fibrinogen-γ with p values of 1.4×10^{-6}, 2.2×10^{-9} and 2.9×10^{-8}, respectively.

With the similar patterns of outliers for all the fibrinogens (***Figure 3-4D***), the correlations of fibrinogen-α, fibrinogen-β and fibrinogen-γ abundance in all patients was further investigated. Plotting fibrinogen-α abundance in an individual against either fibrinogen-β or fibrinogen-γ revealed significant multicollinearity among the proteins (***Figure 3-4E***). This lack of fibrinogens acting as independent variables is not surprising as they are all subunits of the larger fibrinogen complex.

We also considered the possibility that the observed residual fibrinogen could potentially be a treatment effect and not the result of MG. The available clinical records of the MG cohort were investigated with respect to the current and past drugs used by individual patients. The diversity in the drug regiments of these individuals (***Figure 3-5***) and the occurrence of high residual fibrinogen in some controls and individuals with RA led us to conclude that these treatments are not the source of high residual fibrinogen.

Accession	Protein	Description	Peptides	Transition Ion
P02671	FIBRINOGEN-α	Fibrinogen alpha chain OS=Homo sapiens GN=FGA PE=1 SV=2 - [FIBA_HUMAN]	GLIDEVNQDFTNR	y10,y9,y8,y7,y6,y5,y4
			DSHSLTTNIMEILR	y7,y6,y5,y4,y3
			GGSTSYGTGSETESPR	y12,y11,y10,y9,y8,y7,y6,y5,y3
			TVIGPDGHK	y7,y6,y5,y4,y3
P02675	FIBRINOGEN-β	Fibrinogen beta chain OS=Homo sapiens GN=FGB PE=1 SV=2 - [FIBB_HUMAN]	DNENVVNEYSSELEK	y12,y11,y10,y9,y8,y7,y6,y5,y3
			HQLYIDETVNSNIPTNLR	y10,y9,y8,y7,y6,y5,y4,y3
			QGFGNVATNTDGK	y10,y9,y8,y7,y6,y5,y4,y3
			EDGGGWWYNR	y9,y8,y7,y6,y5,y4,y3
			HGTDDGVVWMNWK	y6,y5,y4,y3
P02679	FIBRINOGEN-γ	Fibrinogen gamma chain OS=Homo sapiens GN=FGG PE=1 SV=3 - [FIBG_HUMAN]	DNCCILDER	y7,y6,y5,y4,y3
			YEASILTHDSSIR	y10,y9,y8,y7,y6,y5,y4,y3
			YLQEIYNSNNQK	y10,y9,y8,y7,y6,y5,y4,y3
			QSGLYFIKPLK	y7,y6,y5,y4,y3
			IHLISTQSAIPYALR	y8,y7,y6,y5,y4,y3
P02768	Albumin	Serum albumin OS=Homo sapiens GN=ALB PE=1 SV=2 - [ALBU_HUMAN]	ALVLIAFAQYLQQPFEDHVK	y14,y12,y11,y10,y9,y8,y7,y5,y4,y3
			LVNEVTEFAK	y9,y8,y7,y6,y5,y4,y3
			SLHTLFGDK	y6,y5,y4,y3
			EFNAETFTFHADICTLSEK	y12,y11,y10,y9,y8,y7,y5,y4,y3

Table 3-2 List of peptides and transition ions used for PRM based cohort study of 31 MG sera, 18 RA sera and 30 control sera.

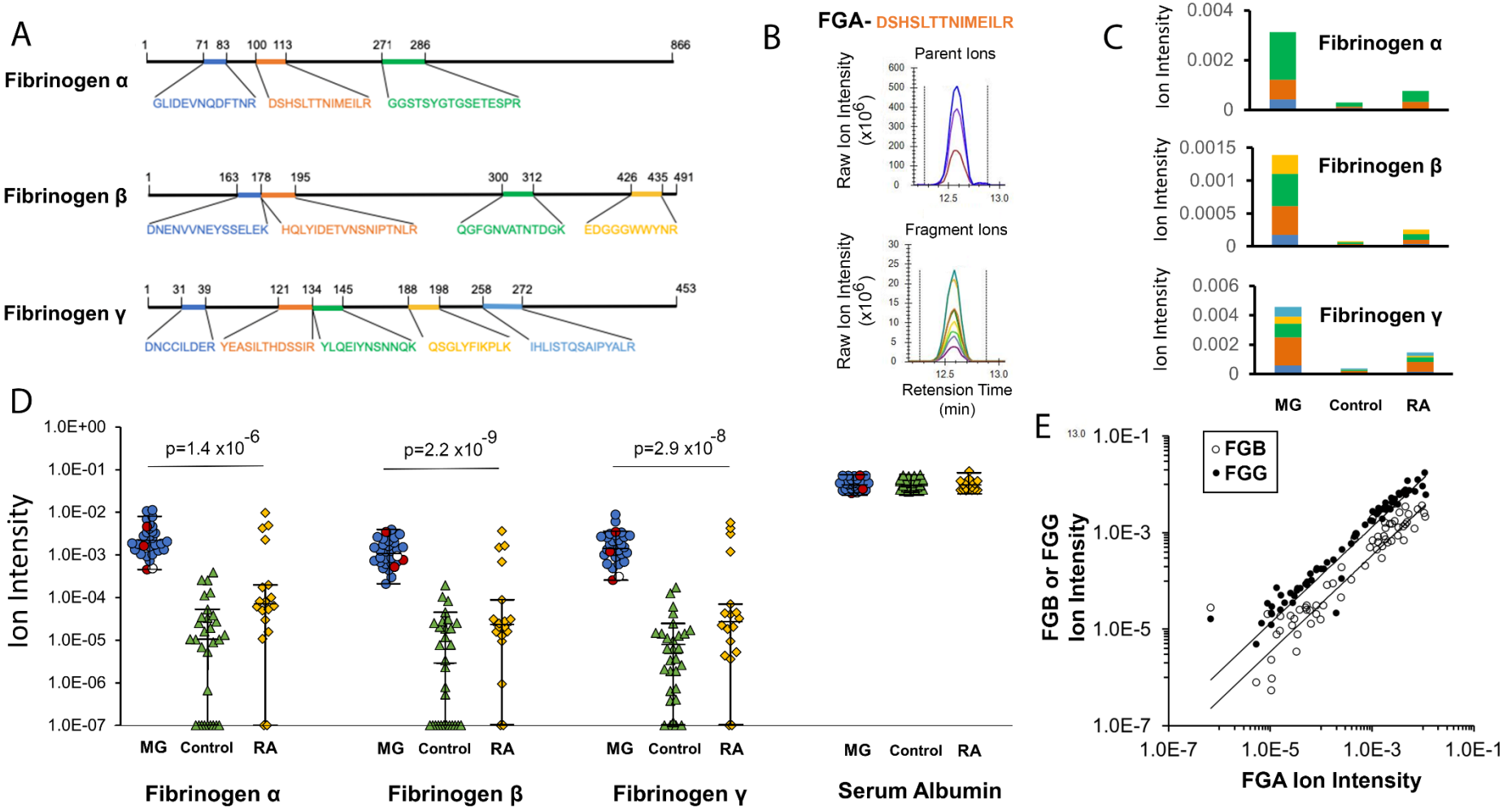

A
Fibrinogen α
1 71 83 100 113 271 286 866
GLIDEVNQDFTNR
DSHSLTTNIMEILR
GGSTSYGTGSETESPR
Fibrinogen β
1 163 178 195 300 312 426 435 491
DNENVVNEYSSELEK
HQLYIDETVNSNIPTNLR
QGFGNVATNTDGK
EDGGGWWYNR
Fibrinogen γ
1 31 39 121 134 145 188 198 258 272 453
DNCCILDER
YEASILTHDSSIR
YLQEIYNSNNQK
QSGLYFIKPLK
IHLISTQSAIPYALR
B
FGA- DSHSLTTNIMEILR
Parent Ions
Raw Ion Intensity (x10^6)
Fragment Ions
Retension Time (min)
C
Ion Intensity
Fibrinogen α
Fibrinogen β
Fibrinogen γ
MG Control RA
D
Ion Intensity
p=1.4 x10^-6
p=2.2 x10^-9
p=2.9 x10^-8
MG Control RA
Fibrinogen α
Fibrinogen β
Fibrinogen γ
Serum Albumin
E
FGB
FGG
FGB or FGG Ion Intensity
FGA Ion Intensity

Figure 3-4 Targeted Proteomic Analysis of a Clinical Cohort

(**A**) For Parallel Reaction Monitoring (PRM), the indicated tryptic peptides from each fibrinogen were used for quantification by targeted LC-MS/MS analysis. (**B**) The extracted ion chromatograms (EICs) of co-eluting fragment ions aligned with the retention time of the precursor ion were quantified. Different coloured traces correspond to different peaks of the isotopic cluster or different fragment ion for the parent or fragment ions respectively. (**C**) Bar diagrams reveal the quantification of the sum of the EICs for the tryptic peptides derived from either fibrinogen-α, fibrinogen-β and fibrinogen-γ for single individuals with Myasthenia gravis (MG), Rheumatoid arthritis (RA) or control. Individual colour represents the EIC for the individual peptide quantified. (**D**) Beeswarm plots of the EICs quantified for the three fibrinogens and albumin as control for 31 MG patients, 18 RA patients and 30 controls. Individuals with undetectable levels of fibrinogen were assigned values at the threshold of detection. Within the MG cohort, of four patients negative for the anti-AChR autoantibody, one was positive for the anti-MuSK autoantibody (white circle) while three were double seronegative (red circles). (**E**) Linear regression analysis of fibrinogen-α, fibrinogen-β and fibrinogen-γ for all individuals reveals a high association for all the fibrinogens.

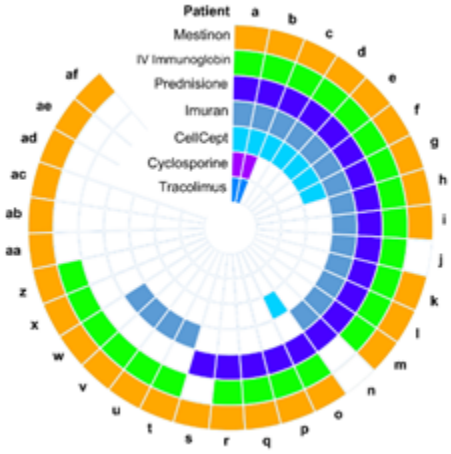


Figure 3-5 Prescription Drug Use of MG Patients

To investigate a possible association of prescribed drugs to MG patients with the observed high serum fibrinogen, the drug regime of the patient cohort was investigated as a potential variable. The sunburst plot reveals the prescription drug use of all MG patients (a-af) in the cohort. Comparison of prescriptions revealed no universal treatment that may have been a potential source of the universally high serum fibrinogen observed in these patients.

3.6 Discussion

Despite recent advances in the understanding and treatment of MG (Dalakas, 2019; Evoli et al., 2015; Gilhus, 2016; Gilhus et al., 2019), challenges remain with the timely diagnosis of this autoimmune disease as a result of its heterogeneous nature and often non-specific symptoms (Abbott, 2010; Spillane et al., 2012). A prior clinical study reported a greater than five-year delay in diagnosis of MG in 13% of patients (Beekman et al., 1997). Accurate diagnosis of MG remains challenging, more so in an emergency room setting, as MG patients may present with atypical symptoms mimicking other neurological conditions, particularly stroke, that like MG, can also cause double vision and/or limb weakness. For molecular testing, the occurrence of the multiple serotypes and seronegative patients leads to limitations in the sensitivity of detection of MG with molecular markers.

To identify a potentially novel general serological marker for MG that may cover all isotypes, we conducted a three-way proteome profiling investigation of the sera of MG patients in comparison with both control individuals and those from a reference autoimmune disease, RA. Our initial pilot study on a limited number of patients revealed potential protein candidates that exhibited statistically significant enrichment for each test group (***Figure 3-1***) for which we believe are well representative of the test groups. For example, our data identified an increased abundance of the sex hormone-binding globulin (SHBG) in the sera of RA patients, a protein previously reported to exhibit a higher risk association for RA in women (Mirone et al., 1996; Qu et al., 2020). From our data, the proteins that exhibited the most striking degree of statistical confidence to be specifically associated with MG, were the three fibrinogen subunits (***Figure 3-1***) that were then chosen for further investigation.

The finding of the α-, β-, and γ-subunits of fibrinogens was unexpected as there are no known associations of MG with any coagulopathy, which one may envision with the detection of unreacted fibrinogen in sera (Kattula et al., 2017). Further investigations revealed that the fibrinogens detected in the sera of MG patients were truly residual, as the comparisons with plasma revealed normal fibrinogen-α levels and the reduction in fibrinogen-α levels during the generation of sera appear normal when contrasted to plasma (***Figure 3-2***) as a result in the limited dynamic range of detection. While the majority of fibrinogen is removed upon the preparation of serum, high residual amounts were constantly observed in MG patients in contrast to normal controls or RA patients. Quantitatively, the magnitude of the residual fibrinogen-α in MG sera was an order of magnitude higher in abundance as determined by mass spectrometry, Western blotting and ELISA (***Figure 3-1*** *and* ***Figure 3-2***). While the observed high residual fibrinogen detected in sera prepared on ice, the fibrinogen was cleared upon incubation at ambient temperature (***Figure 3-2E***), suggesting that this hypocoagulability was not likely to occur *in vivo*.

We investigated whether the high residual serum fibrinogen correlation to MG was a statistical anomaly or whether this association was repeatable in additional samples. Prior to investigating a larger patient cohort, a blinded study was conducted with randomized serum samples, where the detection of fibrinogen-α led to the 100% successful identification of MG patients (***Figure 3-3***).

Based on the positive results of our blinded study, we analyzed a larger patient cohort by mass spectrometry using the quantitative PRM methods to quantify all three fibrinogens in sera (***Figure 3-4***). These analyses confirmed a high specific association of serum fibrinogens with MG when contrasted to controls or RA samples, although the correlation was not absolute as a few samples in the control and RA groups also exhibited high residual fibrinogen levels. The very large

variation observed in controls and RA samples indicates presence of residual serum fibrinogen is common occurrence in population. Nevertheless, residual serum fibrinogen levels observed in MG samples were elevated consistently among all samples measured and were significantly several fold higher than what is observered in control or RA samples. There were four samples of RA exhibiting similar level of residual serum fibrinogens to MG samples, which we speculate to be potentially an undiagnosed secondary autoimmune disease. Further investigation is necessary to understand the cause of elevated fibrinogen in the serum of these individiuals.

The observation of high serum fibrinogen interfering with analysis in cases not attributed to dysglobulinemia, have been reported, but its origin was unknown (Lefèvre & Gillery, 1997). We queried whether the uniform observation of high serum fibrinogen may have been linked to the treatment of MG, as mentioned in the results. This does not appear to be the case due to the considerable diversity of pharmacological treatment of the MG cohort in addition to the detection of high fibrinogen in some non-MG individuals (***Figure 3-5***). Potential cross-reactivity of auto-antibodies to fibrinogen may protect fibrinogen from activation, but this seems unlikely as it appears to occur with all serotypes. Interestingly, a previous proteome-wide screening for cross-reactivity of autoantibodies in MG patients failed to report fibrinogen as a potential target (A. Becker et al., 2013). Nonetheless, other factors may bind fibrinogen and preclude its conversion to fibrin. Further lines of investigation will be required to elucidate the biochemical mechanisms resulting in the elevated residual serum fibrinogen levels.

In summary, we have identified residual fibrinogen in all patients diagnosed with MG. This discovery was enabled by our combination of unbiased shotgun proteomics, Western blotting, ELISA and targeted mass spectrometry of unaltered serum. We found the presence of high serum fibrinogen levels to be highly sensitive to MG with specificities that appear to be superior to current

clinical testing methods. Additionally, high serum fibrinogen is also detected in all serotypes as well in the Class I patients (a purely ocular form of MG), indicating that residual serum testing may be more advantageous than current serological tests. We anticipate our findings to provide the foundation for a routine approach for clinical MG diagnosis.

3.7 Future Direction

A new clinical diagnostic tool for identifying Myasthenia Gravis in patients is necessary and residual serum fibrinogen can be used to develop low-costing diagnostic approach. Translating our finding to clinical application does involve huge amount of work of developing the device, and proving its utility in clinical settings through a clinical trial. Focusing on the basic sciences, the mechanism that leads to the presence of residual serum fibrinogen in MG patients are still needs to be investigated. We speculate the potential presence of auto-antibody in these individuals that prevents the cleavage of fibrinogen to fibrin by thrombin, thus leaving residual serum fibrinogen during blot clotting process. Another possibility is the presence of post-translational modification of fibrinogen in the vicinity of thrombin binding site that prevents proper activity of thrombin during fibrinogen proteolysis. Understanding the mechanism will allows us to further understand Myasthenia Gravis and its implication towards patient's body.

Chapter 4:

Proteomic Quantification of Ribosomal Proteins in Mice Tissues

4.1 Acknowledgement and Contribution

Dr. Faraz Hussain and I were responsible for tissue harvesting and sample preparation. I did the proteomics data collection, method optimization and data analysis. Study design were equal contribution from myself, Dr. Faraz Hussain and Dr. Richard Fahlman. I would like to acknowledge Janice Cotton-McKay from the Health Sciences Lab Animal Services for providing the mice used in this study. The research was conducted with funding from the Discovery Grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada.

4.2 Abstract

Constitutive nature of ribosomes with no regulatory role in protein translation has been a well understood concept until recent findings uncovering heterogeneity in ribosome complexes fuelling interest into their impact towards protein expression. The new ribosomal heterogeneity idea uncovers various questions trying to understand the extent of heterogeneity in many aspects of ribosomes; its structural constituents, post-translational modification, and associated proteins. One factor of heterogeneity could be attributed to differential expression of individual ribosomal proteins that lead to potential specialization of tissues. We addressed this question through a label-free proteomic quantification approach utilizing PRM to measure expression level of ribosomal proteins across multiple mice tissues. We observed seven small and thirteen large ribosomal proteins with significant change in abundance across measured samples of more than two fold difference at $p<0.001$ and $p<0.05$ respectively. Additionally we observed peptide absence for eS10 in kidney and liver tissues that corresponds to potential expression of eS10 isoform.

4.3 Introduction

The ribosome, a “molecular machine par excellence”, plays a substantial role as somewhat of an automaton in expression of genetic information (Frank, 2000; Sonenberg & Hinnebusch, 2009). The orchestrated levels of regulatory specificity of protein expression has always intrigued researchers to better understand the embryo development, disease mechanism and progression, and other life processes. Extensive diversity of molecular instruments involved in the process controls the temporal and spatial precision of protein expression leading to specialized cellular role and behavior. Traditionally, the molecular machinery of ribosome in protein expression is considered to be playing a constitutive role rather than a regulatory one, with no contribution towards cellular diversity.

Despite the constitutive role of ribosomes, recent findings have attributed towards the unique functionality of translational machinery driven from ribosomal heterogeneity. Rather having apparent passive functionality, the ribosome assembly is a complex combination of large and small subunits (such as in eukaryotes, 60S and 40S subunits respectively) while each of the subunit is an intricate architecture of ribosomal RNA (rRNA) and proteins. The eukaryotic large subunit comprises of three rRNAs (28S, 5.8S and 5S) and approximately 45 proteins of which 27 are exclusive to eukaryotes while small subunit has 1 rRNA (i.e. 18S) in combination with ~32 proteins of which 17 are exclusive. This complex assembly primarily attributes towards the heterogeneity of the ribosome exaggerated by the rRNA diversity, ribosomal proteins differential expression and post-translational modifications, and several molecular factors associated with ribosomal activity. The extent of variance in rRNA contributes to ribosomal heterogeneity that could have functional connotation, as rRNAs are transcribed from multiple ribosomal DNA copies.

Upon rRNA sequence data mapping, multiple differentially expressed tissue-specific rRNA alleles have been identified (Parks et al., 2018; Stults et al., 2008). Furthermore this variation is deepened with post-translational modification of more than 200 nucleotides of human rRNA, and among these modifications 2′-O-methyation is the most common (Roundtree et al., 2017). The ribosomal heterogeneity is supplemented by the post-translational modification of ribosomal proteins (RP), as the most commonly reported are phosphorylation and ubiquitination (Simsek & Barna, 2017). The phosphorylation of eS6 has been reported as tissue specific (e.g., smaller pancreatic beta cells) and associated with glucose homeostasis in mice (Meyuhas, 2015). These posttranslational modifications may account for toxic or counter toxic roles to rescue biological functions, as observed in uS19 phosphorylation and Parkinson disease (Martin et al., 2014).

This variation of protein content goes beyond ribosomal constitution and leads towards other ribosome-protein interactions. The association of pyruvate kinase with endoplasmic reticulum (ER) – ribosome in comparison with cytosolic ribosome has been explore in a recent study, sharpen the sub-cellular heterogeneity of ribosomes contributing towards cellular functions (Simsek et al., 2017).

The constitutively changing ribosomal composition and selective interaction may also wield more specialized activities specific to the tissues based on the dedicated needs. This could range from the ribosomal interaction with specific regulatory elements (e.g., internal ribosome entry sites or upstream open reading frames). Different studies have pointed variation in expression profile and tissue specific modification of ribosomal proteins, suggestive of diverse ribosomal protein profile based on tissue specificity (Williams & Sussex, 1995). Studies suggest not all protein components are in equimolar amounts pointing towards regulated translation of ribosomal proteins (Gupta & Warner, 2014). This study aims to explore the potential variation of ribosomal protein

profiles in association with tissue specificity as a suggestive role of ribosomes in tissues specific regulation.

4.4 Methods

4.4.1 Tissue harvest, homogenization and ribosome preparation

Tissue and cell lines homogenization and ribosome were prepared following the protocol available from Cold Spring Harbor Protocols and summarized below (Rivera et al., 2015). The harvested tissues were re-suspend in 2x volumes of Mammalian Ribosome Homogenization Buffer (MRHB) (50 mM Tris-HCL pH7.5, 5mM MgCl2, 25mM KCl, 0.2 M sucrose, prepared in MilliQ water and 0.22-µm filter sterilized) and homogenized. A final concentration of 0.7% NP-40 was added to the homogenate and homogenized further by repeated micropipeting on ice. To pellet out nuclei, suspension was centrifuged at 750*g*, 4C for 10 min. The supernatant centrifuged further at 20,000*g*, 4C for 10 min to obtain a post-mitochondrial fraction (PMT). A 4M KCl solution was added to the PMT fraction at a volume of $V_{KCl}=PMT_{vol}/14$, while total volume should not exceed 600 ul. The KCl- adjusted PMT fraction was layered over a 1 mL sucrose cushion. Tubes were subjected to ultracentrifugation at 250,000*g*, 4C for 2 hours. A compact and translucent pellet of ribosome obtained and rinsed with cold dH_2O. Pellets re-suspended in 100 ul of Tris-Cl buffer and the optical density of the ribosome suspension obtained with a spectrophotometer at 260 nm. An OD of 14 at 260*mm* corresponds to 500ug ribosomal proteins.

4.4.2 Peptide preparation for LC MS/MS analysis

A 100 ul solution of 1 mg (1000 µg) ribosome suspension, 45 µl 6 M Urea and 100 mM Tris was constituted. The reduction of the sample was achieved by adding 3 µl of the 100 mM beta-mercaptoethanol (β-MB) to a final concentration of 4.8 µM. The solution was incubated at 37 °C for 1 hr. Alkylation was completed by 7 µl of 200 mM Iodoacetamide to a final concentration of 20 µM and incubated at 37 °C for 1 hr. Added 12 µl of 100 mM β-MB and incubated for 1 hr at

37 °C. To dilute the urea content, added 465 µl d-H2O. Added 50 µl of trypsin solution (20 µg/100 µl) and incubated at 37 °C overnight. Next day, added 1 µl of Acetic acid. (Note: Expected final peptide conc. of 50 fmol/ µl for each protein.)

4.4.3 Peptides and ions selection for PRM

A skyline document was created and populated with mouse ribosomal proteins acquired from UniProt database (UniProt). Skyline peptide filter was applied on each protein sequence to identify potential quantifiable peptides with 8 - 25 amino acids in length. For each peptide, only precursor ions with +2 and +3 charge were selected for quantification while fragment ions were filtered to only include y-ions with +1 charge excluding y1 and y2 ions. The filtered peptides were screened for prior observations from preliminary shotgun analysis to further narrow down the peptide list to a minimum of two peptides per protein. Potential peptides for proteins with no prior observations were obtained from Peptide Atlas database (www.peptideatlas.org) chosen based on their ranking on peptide suitability score from the "predicted highly observable peptide" section.

4.4.4 Proteomics sample preparation and data acquisition.

Digested peptides were analyzed by LC-MS/MS using a ThermoScientific Easy nLC-1000 in tandem with a Q-Exactive Orbitrap mass spectrometer. Each sample (5µL) was resolved using a 44 minute gradient at 500nL/min (0-22% Buffer B; Buffer A : 0.2 formic acid in 5% ACN, Buffer B: 0.2% formic acid in ACN) on a 2 cm Acclaim 100 PepMap Nanoviper C18 trapping column in tandem with a New Objective PicoChip column (REPROSIL-Pur, C18-AQ, 3µm, 120Å, 105mm). For data acquisition, each duty cycle begins with full scan at 35000 resolution (at 200 m/z) with target automatic gain control (AGC) at 3 x 10^6 and maximum injection time of 200ms followed by PRM scans at 17500 resolution (at 200 m/z), an AGC value of 2 x 10^5 and

maximum fill time of 50ms with isolation window of 1.6 m/z. Fragmentation was performed with normalized collision energy of 27.

4.4.5 Peptide selection and optimization

A total of 210 peptides from 75 proteins were measured through three scheduled PRM injections per sample. Scheduled PRM method was set up with seven minute window with maximum 30 concurrent peptides to ensure 8-10 measurement points per LC peak per transition. A full list of selected ribosomal proteins and peptides, peptide position are presented in ***Table 4-1***.

4.4.6 Data Analysis

Raw data were imported to Skyline software for peptide peak identification and integration (MacLean et al., 2010). 3-5 transition used in tandem with precursor retention time matching as well as mass error for peptide identification and most intense fragment ion used for quantification. Peptide, transition and integrated peak area information were exported to Excel for further analysis. Relative protein abundance was measured by summing peak areas for each measured peptide in a protein. Protein abundances were normalized by first calculating the relative Total Ion Current (TIC) for each sample and dividing each protein abundance to TIC to yield the "proportion of total" value per sample. Normalized protein abundance then averaged across replicates to derive the relative protein abundance.

4.4.7 Statistical Analysis

Statistical tests were manually calculated in Excel. One-way Analysis of Variance (ANOVA) was used to determine significance. $p<0.05$ were considered significant which

corresponds to F-value of 2.46 with degree of freedom (df_{within}= 18 and $df_{between}$=9). As post-hoc analysis, Tukey's Honest Significantly Different (HSD) method was implemented for pair-wise comparison with $p<0.05$ considered significant corresponding to Critical Q value of 5.07. All samples were analyzed in triplicates (n=3) except for testis and EL4 cell line which had n=2.

4.5 Results

4.5.1 Peptide Selection and Optimization

A total of 372 peptides were initially analyzed for 84 ribosomal proteins (***Table 4-1***). For screening of selected peptides, a combination of 519 (+2 / +3) precursor ions were analyzed for quality assurance. To establish a stringent screening, the quantification and quality assurance were based on successful peptide peak identification by skyline's peak picking algorithm, retention time falling in the range of peptide retention time predictor, at least 3 co-eluting fragment ions for identification, as well as retention time matching from peptides identified from proteome discoverer. Based on selection criteria, a total of 210 peptides for 75 ribosomal proteins were shortlisted for further analysis, while uS19, uS14, uL30L, uL16L, eL22L, eL32L, eL37, eL40 and eL41 failed to meet the criteria and/or undetectable during the initial analyses.

4.5.2 PRM Data Analysis

4.5.2.1 Small Ribosomal Proteins

The change significance of small ribosomal proteins among the tissues and cell lines was analyzed by one-way ANOVA to identify the proteins changing among the group. A total of seven ribosomal proteins significant relative abundance with a p-value ≤ 0.001 were eS25, eS19, uS9,

uS7, eS28, eS21, and eS31 with maximum fold change >2. The statistically significant ribosomal proteins are shown with maximum fold change in the **Figure 4-1** with a $p<0.001$ are marked in pink.

To identify the relative abundance based on significantly changing proteins, a Tukey's Honest Significant Difference (HSD) analysis was performed. The ***Figure 4-1*** shows the pairwise comparison of changing abundance among the tissues and the cell lines, as bar diagram demonstrate the abundance and the changing color pattern shows how many tissues are different with the significantly changing abundance.

Among the 31 small ribosomal proteins quantified, three shown more than four fold change with a significance of $p>0.05$. These proteins includes eS21, eS31 and eS30 as shown in ***Figure 4-3A***. The eS31 seems to be a part of the core ribosomal domain while eS21 and eS30 are found associated with surface of the ribosomal structure of the as shown in ***Figure 4-3B***.

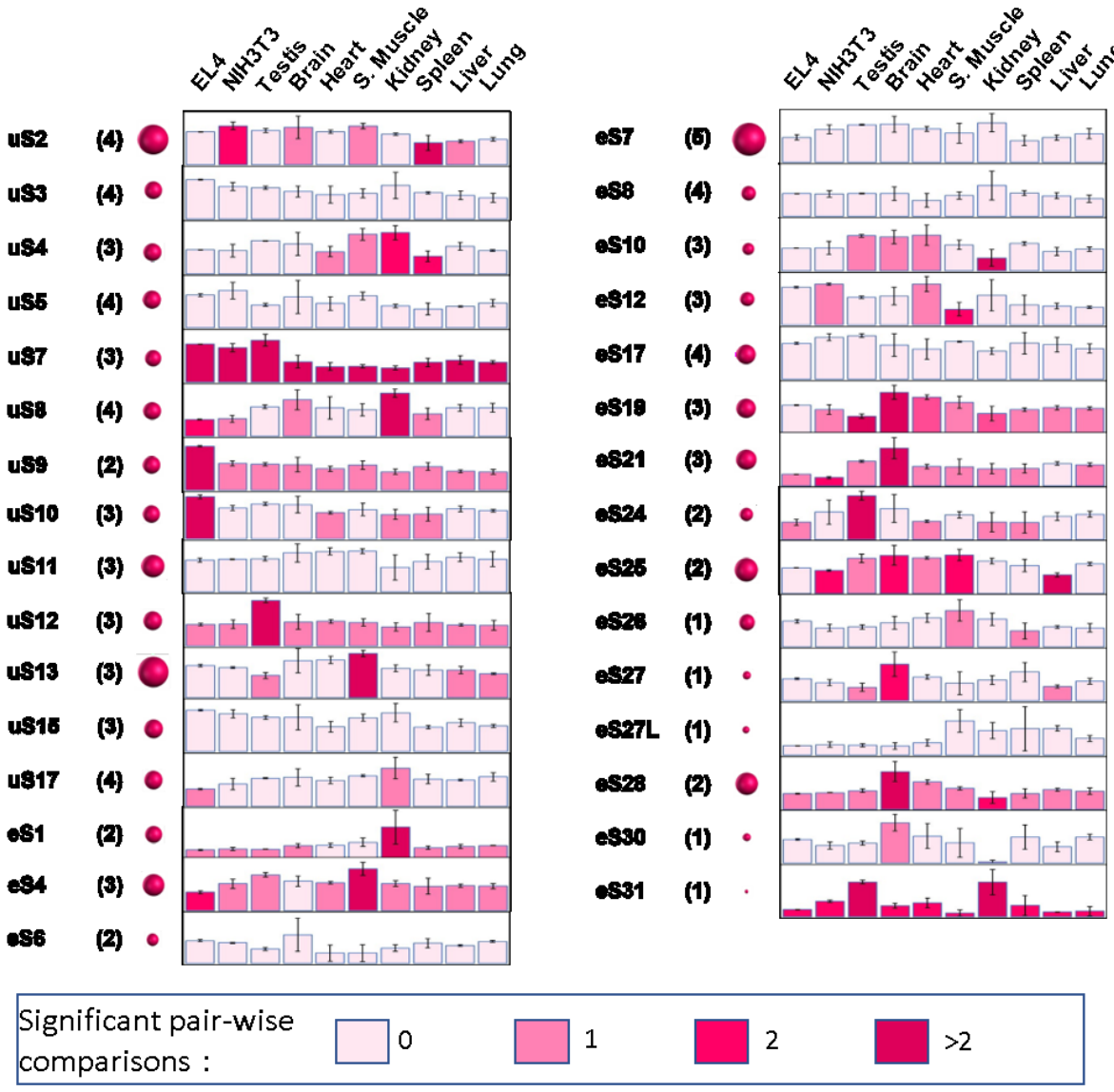


Figure 4-1 Summary of Protein Quantification of Small Ribosomal Proteins.

Relative protein abundance for each protein across multiple tissue and cell lines is represented with bar graph. Number of peptides quantified for each protein is stated within parentheses and volume of spheres for each protein represent their average proportion of total within sample. Statistical results from ANOVA and Tukey's HSD are represented with different shades of maroon with the lightest colored bars indicates no significant pair-wise comparison identified with any tissue or

cell line while the darkest shaded bars indicate presence of more than 2 tissues or cell lines significantly different at $p<0.05$.

4.5.2.2 Large Ribosomal Proteins

The one-way ANOVA was also been recruited for the analysis of large ribosomal proteins to identify the variation among the cell lines and tissue samples. A total of 13 large ribosomal proteins shown more than four-fold change with a significance of $p>0.05$. The relative abundance based on significantly changing proteins is shown in ***Figure 4-3A*** to demonstrate the variability of large ribosomal proteins among the tissues and cell lines. The bar diagram demonstrate the abundance and the changing color pattern shows how many tissues are different with the significantly changing abundance. Among these proteins, eL22 and eL42 shown variability among one tissue, while uL6, uL16, uL14, eL36, eL29, eL30, eL33, eL39 and eL43 were differently expressed between more than two tissues. The large ribosomal proteins eL43, eL38, and uL14 were found to be associated with core domain of the ribosome. The other varying large ribosomal proteins were associated with different ribosomal regions.

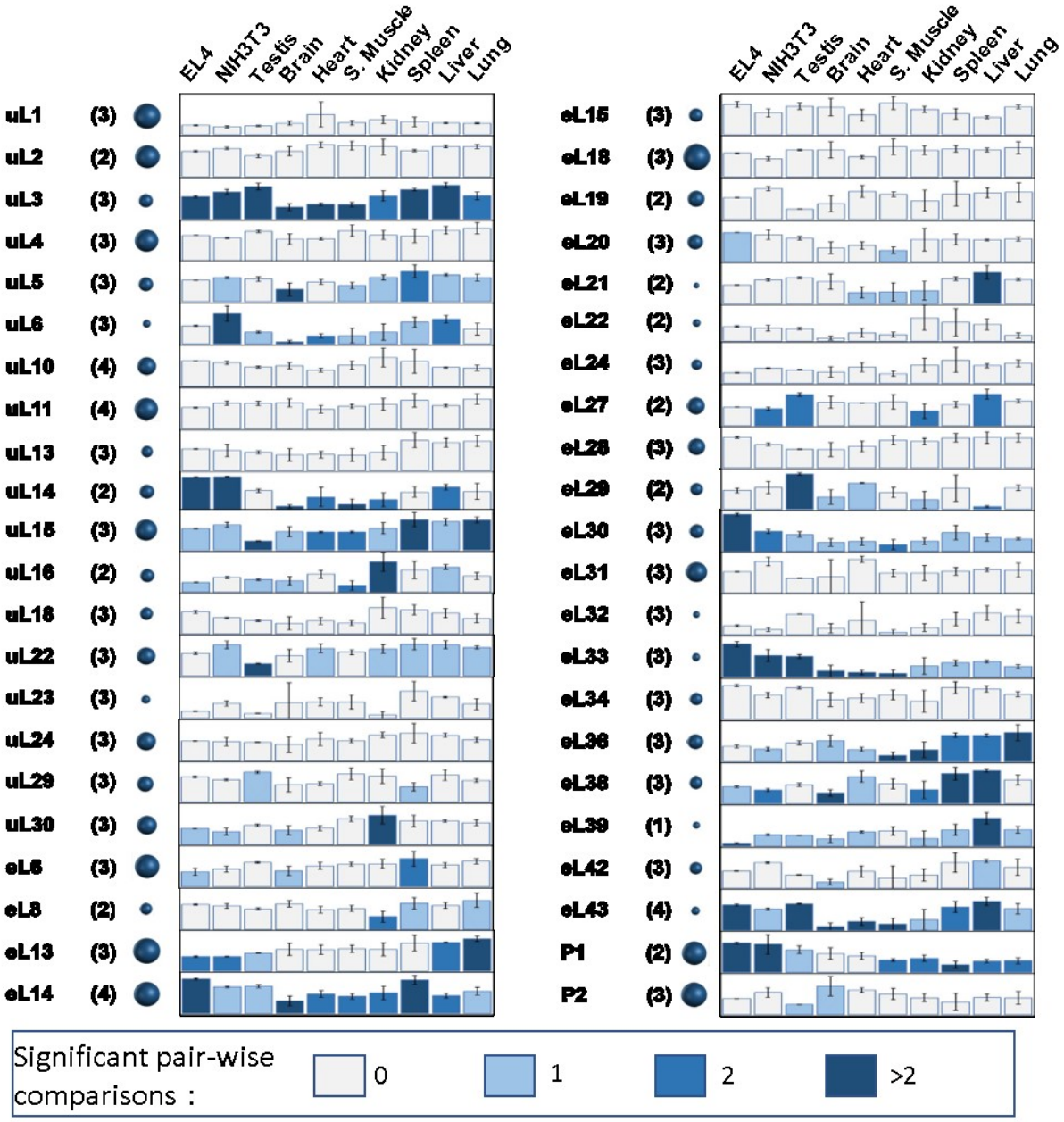


Figure 4-2 Summary of Protein Quantification of Large Ribosomal Proteins

Relative protein abundance for each protein across multiple tissue and cell lines is represented with bar graph. Number of peptides quantified for each protein is stated within parentheses and volume of spheres for each protein represent their average proportion of total within sample. Statistical results from ANOVA and Tukey's HSD are represented with different shades of blue with the lightest colored bars indicates no significant pair-wise comparison identified with any tissue or

cell line while the darkest shaded bars indicate presence of more than 2 tissues or cell lines significantly different at $p<0.05$.

4.5.3 Tissue Specific Ribosomal Protein Variation

These variations among the tissues were different from in terms of up and down regulation of these proteins, as found to be differently associated. As shown in ***Figure 4-4***, eL33 and eL43 are down regulated in heart samples, while in brain tissues, eL14, uL5, uL3, eL38, uL14, eL33, eL43 and uL6 were down regulated. An upregulation of eS28, eS21 and eS19 was observed in brain samples (***Figure 4-4***). While looking more specifically at eL38 in contrast to brain tissues, it was found upregulated in liver cells, along with eL43, eL39 and eL21 (***Figure 4-4***). The downregulated RPs observed in liver tissues were, eS25, eL29 and eS31.

4.5.4 Significantly Varying Ribosomal Proteins

Among the varying ribosomal proteins, concentration among the tissue samples, eL33 and eL43 has been seen most significantly regulated RPs among all as each of these proteins appeared to be down regulated in three tissue samples (Brain, Heart, skeletal Muscles) while up regulated in testis tissue samples and 2 cell lines (EL4 and NIH3T3). The eL33 shown fold change > 8 with a p-value >0.001 while the eL43 has shown a fold change > 4 with a similar amplitude of significance.

With a fold change > 4, uL14 was downregulated in two tissues (Brain and Skeletal muscles) and upregulated two cell lines (EL4 and NIH3T3) with a significance of $p>0.001$. Third most commonly varying ribosomal protein was eL38 as downregulated in NIH3T3 and Brain, and upregulated in two tissues (Spleen and Liver) with a fold change > 2 at a significance of $p>0.001$.

Next in the list, P1 and eL36 were witnessed varying among three samples with a fold change > 4, while eL14 and uL15 with a fold change > 2.

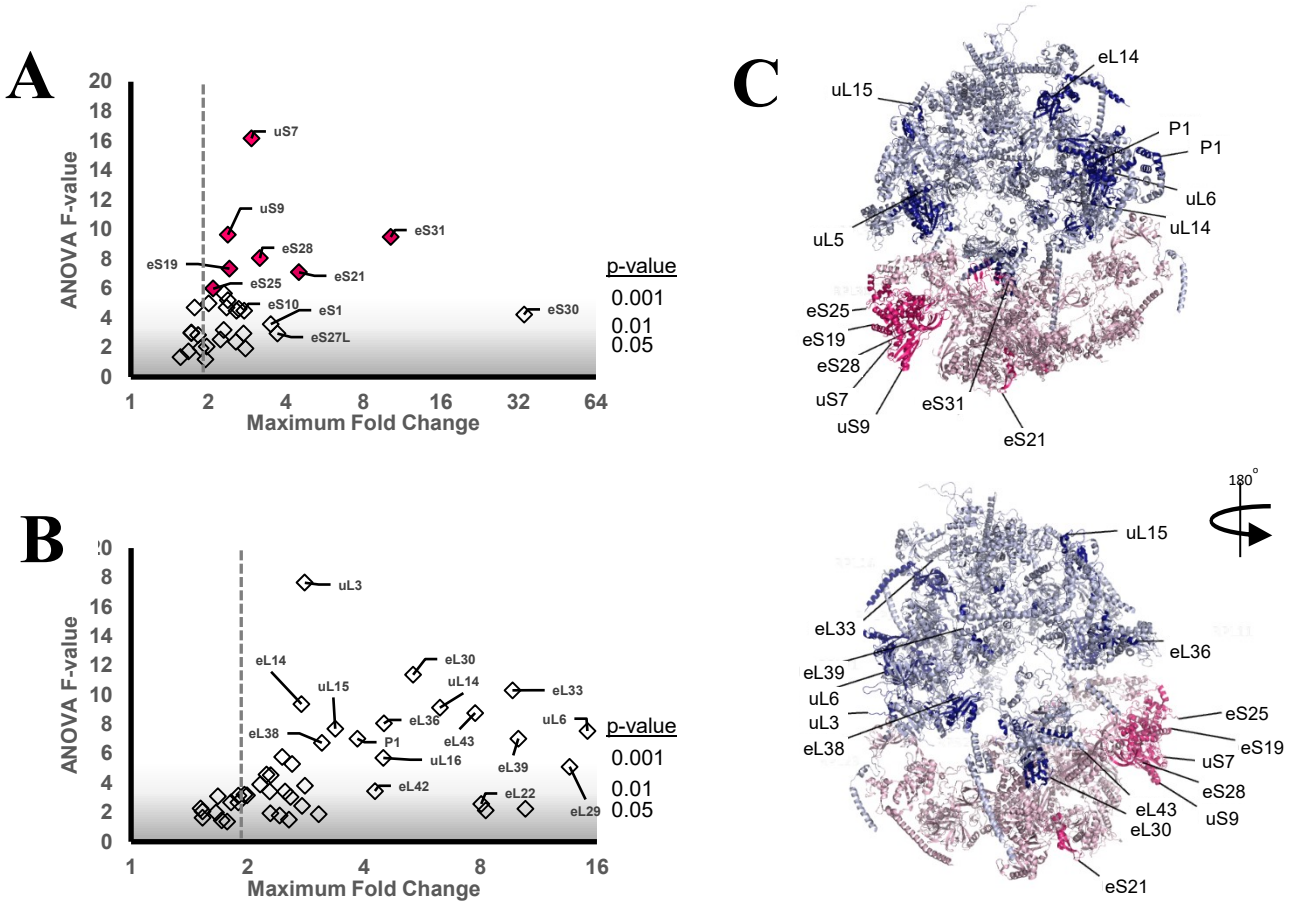


Figure 4-3 Changes in Ribosomal Proteins

A. and **B.** Plot showing maximum fold change between tissue calculated by dividing maximum with minimum relative abundance for each protein in small ribosomal proteins, RPS (**A**) or large ribosomal proteins, RPL (**B**) against the ANOVA F-value. Dark grey shaded area indicates insignificant protein changes at $p<0.05$. p-value cut-offs for 0.05, 0.01, and 0.001 are as indicated. Selected proteins from the data set are individually labelled. **C.** 80S human ribosome complex (PDB: 4V6X) shown without rRNA. RPLs are colored blue while RPSs are colored pink. Proteins with $p<0.001$ are colored in darker shade of blue and pink and individually labelled.

4.5.5 Tissues With Significant Ribosomal Proteins Variation

Among the eight-tissue types, brain tissues showed significant variation in ribosomal protein abundance as eight ribosomal proteins were found downregulated while three were upregulated. The testis tissue shown most number of upregulated ribosomal proteins that is six in total, while three were downregulated attributing to nine ribosomal proteins with varying abundance. The kidney tissues shown eight ribosomal proteins significantly varying among those four were upregulated and four were downregulated. The seven ribosomal proteins were observed significantly varying in liver (four upregulated, three downregulated) and skeletal muscles (two upregulated, five downregulated) tissue samples. (***Figure 4-4***)

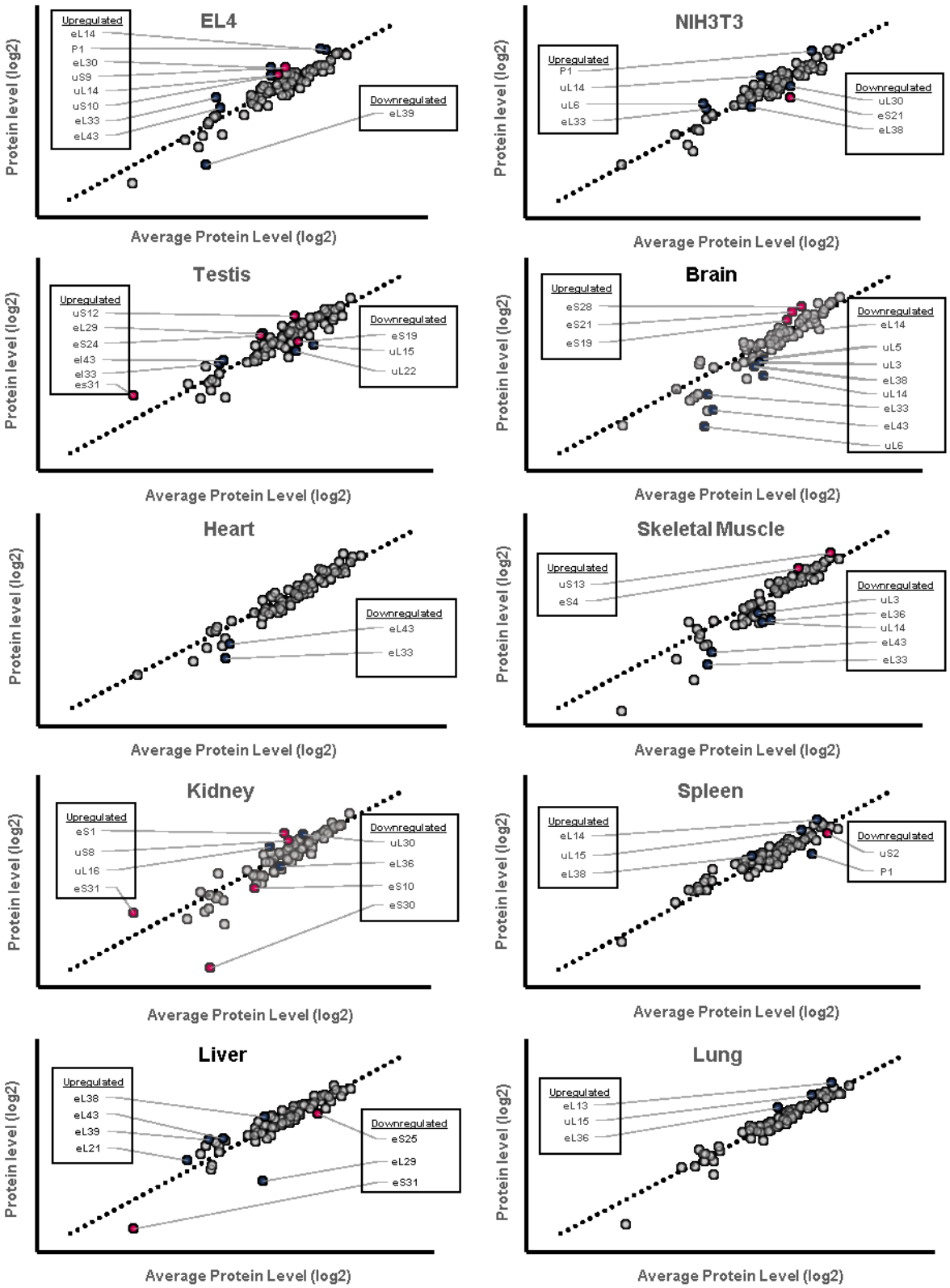
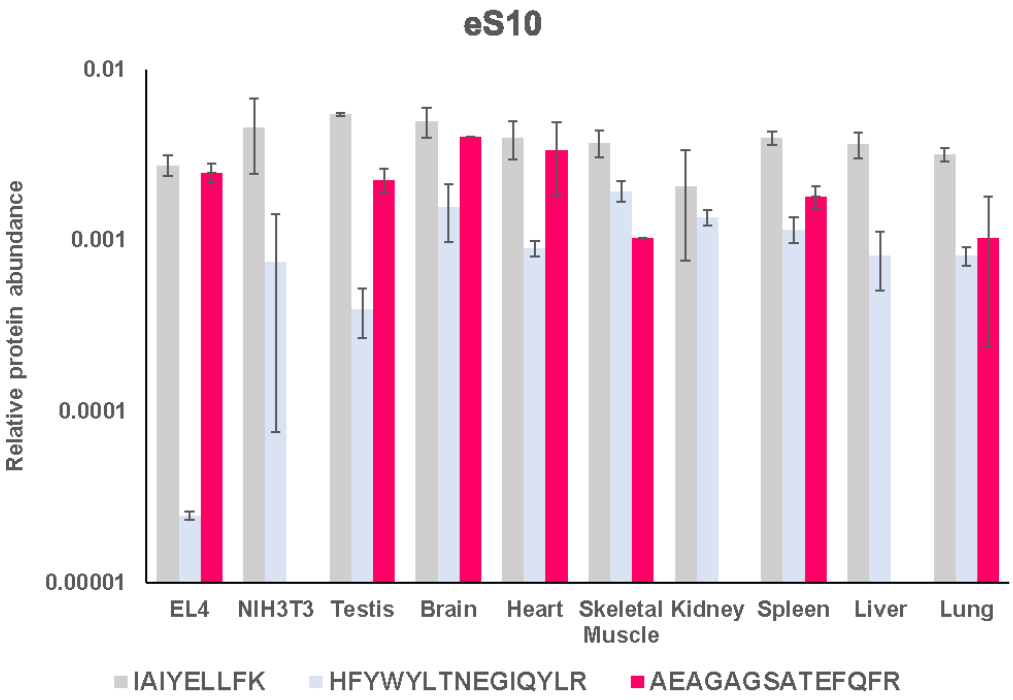

EL4
Upregulated
eL14
P1
eL30
uS9
uL14
uS10
eL33
eL43
Downregulated
eL39
Protein level (log2)
Average Protein Level (log2)
NIH3T3
Upregulated
P1
uL14
uL6
eL33
Downregulated
uL30
eS21
eL38
Protein level (log2)
Average Protein Level (log2)
Testis
Upregulated
uS12
eL29
eS24
eL43
eL33
eS31
Downregulated
eS19
uL15
uL22
Protein level (log2)
Average Protein Level (log2)
Brain
Upregulated
eS28
eS21
eS19
Downregulated
eL14
uL5
uL3
eL38
uL14
eL33
eL43
uL6
Protein level (log2)
Average Protein Level (log2)
Heart
Downregulated
eL43
eL33
Protein level (log2)
Average Protein Level (log2)
Skeletal Muscle
Upregulated
uS13
eS4
Downregulated
uL3
eL36
uL14
eL43
eL33
Protein level (log2)
Average Protein Level (log2)
Kidney
Upregulated
eS1
uS8
uL16
eS31
Downregulated
uL30
eL36
eS10
eS30
Protein level (log2)
Average Protein Level (log2)
Spleen
Upregulated
eL14
uL15
eL38
Downregulated
uS2
P1
Protein level (log2)
Average Protein Level (log2)
Liver
Upregulated
eL38
eL43
eL39
eL21
Downregulated
eS25
eL29
eS31
Protein level (log2)
Average Protein Level (log2)
Lung
Upregulated
eL13
uL15
eL36
Protein level (log2)
Average Protein Level (log2)

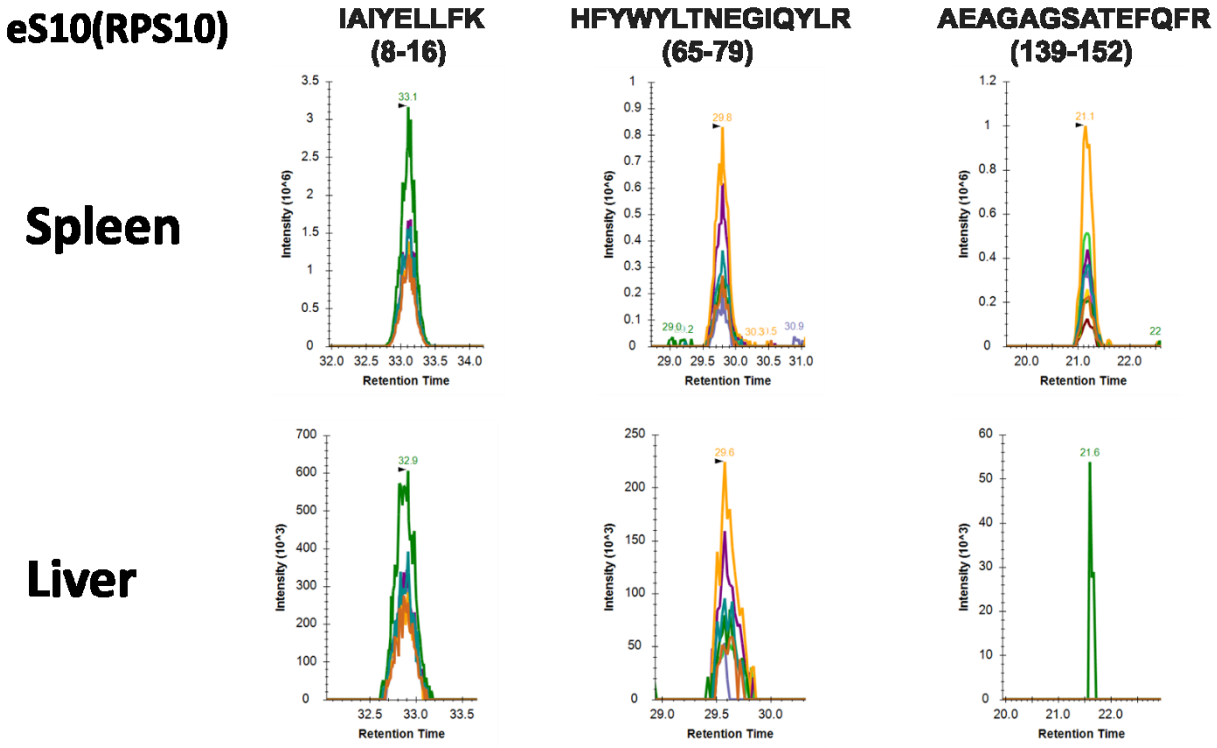
Figure 4-4 Specificity curve of ribosomal proteins across tissues and cell lines

Protein levels for all ribosomal protein for each specified tissue or cell line were plotted against average protein level observed across all sample types studied. Proteins falling above the correlation line (black dotted line) are considered upregulated while downregulated proteins fall below correlation line. Proteins of interest with significant changes for each sample type are labelled and colored (dark blue = RPL, dark pink = RPS).

4.5.6 Tissue Specific Isoform Variation

This ribosomal protein variation was not limited to the concentration, while significant variation was observed in the tissue specific isoform of the ribosome. The eS10 has been observed with a varying isoform between different tissue and cell type. The eS10 isoform (Uniprot ID: Q3UW83) with a missing peptide "AEAGAGSATEFQFR" was specifically present in kidney, liver and NIH3T3, while eS10 present was present in other tissue and cell types (***Figure 4-5***).


eS10
0.01
0.001
0.0001
0.00001
Relative protein abundance
EL4
NIH3T3
Testis
Brain
Heart
Skeletal Muscle
Kidney
Spleen
Liver
Lung
IAIYELLFK
HFYWYLTNEGIQYLR
AEAGAGSATEFQFR


eS10(RPS10)
IAIYELLFK (8-16)
HFYWYLTNEGIQYLR (65-79)
AEAGAGSATEFQFR (139-152)
Spleen
Liver
Intensity (10^6)
Intensity (10^3)
Retention Time

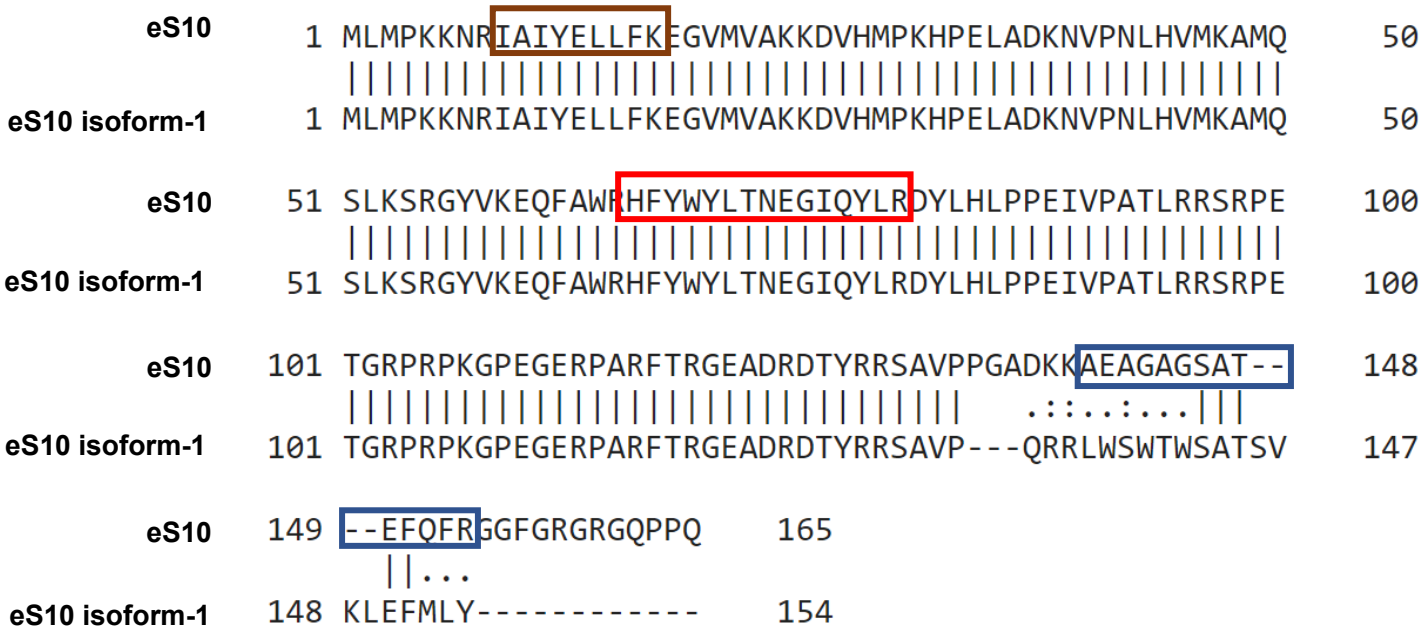


Figure 4-5 eS10 Peptide Quantification Data

Relative protein abundance for 3 quantified peptides of eS10 are shown for tissues and cell lines. "*" indicates undetected peptide. **B.** Overlapping fragment ion abundance curve for each peptide measured for eS10 in both Spleen and Liver sample indicating the absence of peptide "AEAGAGSATEFQFR" in liver. **C**. Sequence alignment of RPS10 and RPS10 isoform 1(Uniprot ID: Q3UW83). Peptide sequences utilized for PRM analysis are indicated in boxes.

4.6 Discussion

The control of translation is increasingly recognized as a major factor in determining protein levels in the cell. The ribosome — the cellular machine that mediates protein synthesis — is characteristically identified as ubiquitous and invariant during the translation process. The ribosomal proteins are known to be highly conserved and vastly expressed; presenting a static outlook that fits the fundamental role in protein synthesis. The recent studied the heterogeneity of ribosomal protein among different tissue types pointing towards differential role and possible regulatory effects. Recent developments have made it clear that heterogeneous ribosome types can exist in different tissues, and more importantly, that these ribosomes can preferentially translate different subsets of mRNAs.

However, ribosomal protein composition in tissues and cells and their role in translational regulations is poorly understood. We attempt to explore this by quantifying all ribosomal proteins (RPs) expressed in tissues to better understand the possibility of heterogeneity within ribosome complex composition that may contribute towards tissue specialization. Previous study attempting to quantify RPs successfully quantified 15 out of 80 (RPs) in mouse embryonic stem cells (mESC) utilizing selected reaction monitoring (SRM) (Shi et al., 2017). They managed to perform absolute quantification of these RPs through the addition of heavy-labelled peptide spiked into the samples; label-based quantification. High cost of labelling as well as difficulty in finding quantotypic peptide for RPs restricted the quantification of all RPs. We circumvent the challenge by utilizing our label-free quantification approach to quantify RPs via PRM.

Initially, we listed 84 RPs obtained from Uniprot database for PRM based quantification for preliminary peptide selection through the measurement of 519 precursor ions. However, for

the final quantification we were able to finalize 210 peptides to quantify 75 RPs. Nine RPs were excluded from quantification due to the lack of suitable tryptic peptides or lack of detection during method optimization. Almost half of these proteins are paralogs of RPs, namely uL30L, uL16L, eL22L, and eL32L where they were undetected during data acquisition. Their lack of detection could be attributed to interference by their paralog counterparts. For example, eL22 has been shown to destabilize its paralog, eL22L1 through binding directly to it mRNA stem loop (O'Leary et al., 2013). Additionally, uL3 mRNA has been reported to have lower abundance compared to its paralog, uL3L selectively in human hearts and skeletal muscle while present in higher abundance in other parts of the body (Chaillou et al., 2014, 2016; Kirby et al., 2015). We have observed similar data in our measurement where uL3 shows lower abundance in heart and skeletal muscle compared to other tissues (***Figure 4-2***).

In ***Figure 4-1*** and ***Figure 4-2***, we have summarized the relative abundance of small and large RPs respectively. Out of 31 small RPs quantified, we observed 19 RPs with significant difference in abundance at $p<0.05$ against at least two tissues. Similarly, 19 RPs from 44 measured large RPs also showed significant difference in abundance against two or more tissue. This indicates that the constitution of ribosome complex do vary between tissues and cells contributed by the variation of abundance of a subset of RPs while majority of RPs do not show significant variation. The work done by Shi and group also indicates variation of RPs stoichiometry in polysome of mESCs where they identified six RPs were significantly substochiometric and four of them only present in 60% – 70% of actively translating ribosome complex (Shi et al., 2017).

We explored the idea of the tissue specific expression of RPs by plotting average abundance of RPs against global average of RPs abundance as summarized in ***Figure 4-4***. Several studies have showed RNA-based evidences of tissue-specific expression of RPs. One example is

testis and prostate where they have been shown to specifically express eS4Y2 than two other paralogs eS4Y1 and eS4X in humans. (Lopes et al., 2010). We have observed several RPs with significantly variable abundance in testis suggesting differential translation of mRNA in testis (***Figure 4-1***,***Figure 4-2*** and ***Figure 4-4***). Additionally, we also noticed a higher number of RPs downregulated in brain with majority if them being large RPs. We speculate that these RPs are required in lower frequency in brains than in other tissues contributing to specialized nature of brain tissue.

We observed during our analysis that several peptides were not detected in all samples that encouraged us to focus on these missing peptides. Although some of these peptides were undetected due to missing data phenomenon (explained further in Chapter 5), we discovered one peptide, AEAGAGSATEFQFR, from eS10 was specifically missing from kidney and liver tissues as well as NIH3T3 cells (***Figure 4-5***). Further investigation revealed the potential expression of eS10 isoform-1, which we determined to be missing the peptide in its amino acid sequence. Nevertheless, further validation in the future is necessary for concluding the presence of eS10 isoform 1 in these tissues though the prospect of protein isoform adding to the level of ribosome heterogeneity is potentially novel.

There are innumerable possibilities that could lead to ribosomal heterogeneity and the finding in this study shows the heterogeneity arises among different tissue types pointing towards the role of heterogeneous ribosomes in molecular functionality. Though the notion of specialized ribosomes is in its infancy, it has conquered the fantasy of researchers, especially the idea of heterogeneity ascending from RP composition. Through this study, we aspire to serve a baseline

reference towards RPs abundance in ribosome complexes across tissues that can help future studies to focus on specific questions to unravel the complexity behind ribosome heterogeneity.

4.7 Future Direction

Our findings have provided foundational evidence of the presence of ribosomal heterogeneity among different tissues. There are numerous avenues yet to be explored to characterize ribosomal heterogeneity. We have to further investigate the possibility ribosome compositional heterogeneity during translation of specific transcript within tissues. Isolating polysomes through sucrose density gradient will allow to study ribosome complexes translating specific transcripts. Our approach focused on global differences in ribosomal proteins in tissues. However, tissues are made up of various cell types that could potentially have difference in their ribosomal composition. Global analysis of ribosomal proteins will overlook these differences that could be identified accurately with specific studies focusing of ribosome compositions in cells. Additionally, single-molecule imaging techniques such as Förster resonance energy transfer (FRET) can be implemented to directly visualize individual ribosomes in real-time, which can uncover information regarding their behaviour and activity. Finally, full characterization of heterogeneity is impossible without the usage of multiple techniques and instruments in combination.

Accension Number	Old Nomenclature	New Nomenclature	Protein Description	Protein Length	Peptides	Sequence Location
Q6ZWV3	RPL10	uL16	60S ribosomal protein L10	214	VHIGQVIMSIR	129 - 139
					LIPDGCGVK	190 - 198
P53026	RPL10A	uL1	60S ribosomal protein L10a	217	DTLYEAVR	8 - 15
					YDAFLASESLIK	107 - 118
					ILGPGLNK	123 - 130
Q9CXW4	RPL11	uL5	60S ribosomal protein L11	178	VLEQLTGQTPVFSK	39 - 52
					IAVHCTVR	68 - 75
					YDGIILPGK	170 - 178
P35979	RPL12	uL11	60S ribosomal protein L12	165	CTGGEVGATSALAPK	17 - 31
					QAQIEVVPSASALIIK	68 - 83
					HSGNITFDEIVNIAR	100 - 114
					EILGTAQSVGCNVDGR	131 - 146
P47963	RPL13	eL13	60S ribosomal protein L13	211	GFSLEELR	75 - 82
					STESLQANVQR	106 - 116
					LATQLTGPVMPIR	146 - 158
P19253	RPL13A	uL13	60S ribosomal protein L13a	203	CEGINISGNFYR	38 - 49
					VLDGIPPPYDK	104 - 114
					LAHEVGWK	141 - 148
Q9CR57	RPL14	eL14	60S ribosomal protein L14	217	VAYISFGPHAGK	12 - 23
					LVAIVDVIDQNR	24 - 35
					ALVDGPCTR	36 - 44
					AAIAAAAAAAAAK	148 - 160
Q9CZM2	RPL15	eL15	60S ribosomal protein L15	204	QLSALHR	32 - 38
					FFEVILIDPFHK	129 - 140
					NTLQLHR	196 - 202
Q9CPR4	RPL17	uL22	60S ribosomal protein L17	184	YSLDPENPTK	4 - 13
					QWGWTQGR	75 - 82
					GLDVDSLVIEHIQVNK	106 - 121

P35980	RPL18	eL18	60S ribosomal protein L18	188	SQDIYLR	20 - 26
					TAVVVGTVTDDVR	79 - 91
					ILTFDQLALESPK	120 - 132
P62717	RPL18A	eL20	60S ribosomal protein L18a	176	NFGIWLR	77 - 83
					DLTTAGAVTQCYR	99 - 111
					VEEIAAGK	129 - 136
P84099	RPL19	eL19	60S ribosomal protein L19	196	VWLDPNETNEIANANSR	22 - 38
					LLADQAEAR	154 - 162
O09167	RPL21	eL21	60S ribosomal protein L21	160	KGDIVDIK	36 - 43
					VYNVTQHAVGIIVNK	64 - 78
P67984	RPL22	eL22	60S ribosomal protein L22	128	AGNLGGGVVTIER	53 - 65
					ITVTSEVPFSK	70 - 80
P62830	RPL23	uL14	60S ribosomal protein L23	140	ISLGLPVGAVINCADNTGAK	16 - 35
					ECADLWPR	124 - 131
P62751	RPL23A	uL23	60S ribosomal protein L23a	156	IEDNNTLVFIVDVK	90 - 103
					VNTLIRPDGEK	124 - 134
					LAPDYDALDVANK	140 - 152
Q8BP67	RPL24	eL24	60S ribosomal protein L24	157	VELCSFSGYK	3 - 12
					VFQFLNAK	28 - 35
					AITGASLADIMAK	81 - 93
P61255	RPL26	uL24	60S ribosomal protein L26	145	FNPFVTSDR	3 - 11
					DDEVQVVR	52 - 59
					YVIYIER	78 - 84
P61358	RPL27	eL27	60S ribosomal protein L27	136	VVLVLAGR	10 - 17
					YSVDIPLDK	85 - 93
P14115	RPL27A	uL15	60S ribosomal protein L27a	148	NQSFCPTVNLDK	66 - 77
					LWTLVSEQTR	78 - 87
					TGVAPIIDVVR	95 - 105
P41105	RPL28	eL28	60S ribosomal protein L28	137	QTYSTEPNNLK	23 - 33
					YNGLIHR	40 - 46

					TVGVEPAADGK	48 - 58
P47915	RPL29	eL29	60S ribosomal protein L29	160	ALVKPQAIKPK	74 - 84
					LAFIAHPK	94 - 101
P27659	RPL3	uL3	60S ribosomal protein L3	403	VACIGAWHPAR	251 - 261
					NNASTDYDLSDK	301 - 312
					FIDTTSK	367 - 373
P62889	RPL30	eL30	60S ribosomal protein L30	115	LVILANNCPALR	45 - 56
					TGVHHYSGNNIELGTACGK	69 - 87
					VCTLAIIDPGDSDIIR	91 - 106
P62900	RPL31	eL31	60S ribosomal protein L31	125	SAINEVVTR	15 - 23
					EYTINIHK	24 - 31
					LYTLVTYVPVTTFK	102 - 115
P62911	RPL32	eL32	60S ribosomal protein L32	135	GQILMPNIGYGSNK	51 - 64
					ELEVLLMCNK	84 - 93
					SYCAEIAHNVSSK	94 - 106
Q9D1R9	RPL34	eL34	60S ribosomal protein L34	117	LSYNTASNK	11 - 19
					IVYLYTK	30 - 36
					AFLIEEQK	94 - 101
Q6ZWV7	RPL35	uL29	60S ribosomal protein L35	123	QLDDLKVELSQLR	20 - 32
					VLTVINQTQK	57 - 66
					YKPLDLRPK	78 - 86
O55142	RPL35A	eL33	60S ribosomal protein L35a	110	AIFAGYK	9 - 15
					DETEFYLGK	37 - 45
					DETEFYLGKR	37 - 46
P47964	RPL36	eL36	60S ribosomal protein L36	105	YPMAVGLNK	5 - 13
					EVCGFAPYER	46 - 55
					VGTHIR	77 - 82
P83882	RPL36A	eL42	60S ribosomal protein L36a	106	DSLYAQGK	31 - 38
					LECVEPNCR	70 - 78
					HFELGGDK	90 - 97

P61514	RPL37A	eL43	60S ribosomal protein L37a	92	KIEISQHAK	28 - 36
					IEISQHAK	29 - 36
					YTCSFCGK	37 - 44
					TVAGGAWTYNTTSAVTVK	63 - 80
Q9JJI8	RPL38	eL38	60S ribosomal protein L38	70	DFLLTAR	10 - 16
					YLYTLVITDK	41 - 50
					QSLPPGLAVK	58 - 67
P62892	RPL39	eL39	60S ribosomal protein L39	51	QNRPIPQWIR	19 - 28
Q9D8E6	RPL4	uL4	60S ribosomal protein L4	419	NVTLPAVFK	21 - 29
					IEEVPELPLVVEDK	144 - 157
					NIPGITLLNVSK	223 - 234
P47962	RPL5	uL18	60S ribosomal protein L5	297	DIICQIAYAR	59 - 68
					FPGYDSESK	180 - 188
					ENPVYEK	249 - 255
P47911	RPL6	eL6	60S ribosomal protein L6	296	SSITPGTVLIILTGR	150 - 164
					AVDLQILPK	260 - 268
					AVPQLQGYLR	271 - 280
P14148	RPL7	uL30	60S ribosomal protein L7	270	AGNFYVPAEPK	100 - 110
					IALTDNSLIAR	189 - 199
					TTHFVEGGDAGNR	246 - 258
P12970	RPL7A	eL8	60S ribosomal protein L7a	266	VPPAINQFTQALDR	76 - 89
					HWGGNVLGPK	236 - 245
P62918	RPL8	uL2	60S ribosomal protein L8	257	AVDFAER	31 - 37
					AVVGVVAGGGR	164 - 174
P51410	RPL9	uL6	60S ribosomal protein L9	192	TILSNQTVDIPENVEITLK	3 - 21
					TGVACSVSQAQK	130 - 141
					DELILEGNDIELVSNSAALIQQATTVK	142 - 168
P14869	RPLA0	uL10		317	IIQLLDDYPK	17 - 26
					GHLENNPALEK	67 - 77

			60S acidic ribosomal protein P0		GNVGFVFTK	84 - 92
					TSFFQALGITTK	135 - 146
P47955	RPLA1	P1	60S acidic ribosomal protein P1	114	AAGVSVEPFWPGLFAK	34 - 49
					ALANVNIGSLICNVGAGGPAPAAGAAPAGGAAPSTAAAPAEEK	50 - 92
P99027	RPLA2	P2	60S acidic ribosomal protein P2	115	YVASYLLAALGGNSSPSAK	3 - 21
					ILDSVGIEADDDR	26 - 38
					NIEDVIAQGVGK	50 - 61
P63325	RPS10	eS10	40S ribosomal protein S10	165	IAIYELLFK	9 - 17
					HFYWYLTNEGIQYLR	66 - 80
					AEAGAGSATEFQFR	140 - 153
P62281	RPS11	uS17	40S ribosomal protein S11	158	VLLGETGK	23 - 30
					EAIEGTYIDK	49 - 58
					EAIEGTYIDKK	49 - 59
					CPFTGNVSIR	60 - 69
P63323	RPS12	eS12	40S ribosomal protein S12	132	TALIHDGLAR	24 - 33
					LGEWVGLCK	85 - 93
					DVIEEYFK	122 - 129
P62301	RPS13	uS15	40S ribosomal protein S13	151	SVPTWLK	21 - 27
					GLTPSQIGVILR	44 - 55
					LILIESR	115 - 121
P62264	RPS14	uS11	40S ribosomal protein S14	151	ELGITALHIK	87 - 96
					TPGPGAQSALR	107 - 117
					IEDVTPIPSDSTR	129 - 141
P62245	RPS15A	uS8	40S ribosomal protein S15a	130	MNVLADALK	4 - 12
					HGYIGEFEIIDDHR	44 - 57
					IVVNLTGR	61 - 68
					FDVQLK	79 - 84
P14131	RPS16	uS9		146	LLEPVLLLGK	51 - 60

			40S ribosomal protein S16		DILIQYDR	110 - 117
P63276	RPS17	eS17	40S ribosomal protein S17	135	VCEEIAIIPSK	34 - 44
					VCEEIAIIPSKK	34 - 45
					RDNYVPEVSALDQEIIEVDPDTK	81 - 103
					DNYVPEVSALDQEIIEVDPDTK	82 - 103
P62270	RPS18	uS13	40S ribosomal protein S18	152	AGELTEDEVER	56 - 66
					IPDWFLNR	79 - 86
					YSQVLANGLDNK	95 - 106
Q9CZX8	RPS19	eS19	40S ribosomal protein S19	145	DVNQQEFVR	8 - 16
					ELAPYDENWFYTR	44 - 56
					VLQALEGLK	103 - 111
P25444	RPS2	uS5	40S ribosomal protein S2	293	SLEEIYLFSLPIK	77 - 89
					GCTATLGNFAK	228 - 238
					ATFDAISK	239 - 246
					SPYQEFTDHLVK	264 - 275
P60867	RPS20	uS10	40S ribosomal protein S20	119	TPVEPEVAIHR	9 - 19
					VCADLIR	35 - 41
					LIDLHSPSEIVK	88 - 99
Q9CQR2	RPS21	eS21	40S ribosomal protein S21	83	DHASIQMNVAEVDR	28 - 41
					TYGICGAIR	52 - 60
					MGESDDSILR	62 - 71
P62267	RPS23	uS12	40S ribosomal protein S23	143	ANPFGGASHAK	38 - 48
					KGHAVGDIPGVR	108 - 119
					VANVSLLALYK	125 - 135
P62849	RPS24	eS24	40S ribosomal protein S24	133	QMVIDVLHPGK	22 - 32
					TTPDVIFVFGFR	50 - 61
P62852	RPS25	eS25	40S ribosomal protein S25	125	LITPAVVSER	67 - 76
					AALQELLSK	86 - 94

P62855	RPS26	eS26	40S ribosomal protein S26	115	NIVEAAAVR	43 - 51
Q6ZWU9	RPS27	eS27	40S ribosomal protein S27	84	DLLHPSPEEEK	6 - 16
P62983	RPS27A	eS31	Ubiquitin-40S ribosomal protein S27a	156	CCLTYCFNKPEDK	144 - 156
Q6ZWY3	RPS27L	eS31L	40S ribosomal protein S27-like	84	DLLHPSLEEEK	6 - 16
P62858	RPS28	eS28	40S ribosomal protein S28	69	VEFMDDTSR	32 - 40
					EGDVLTLLESER	52 - 63
P62908	RPS3	uS3	40S ribosomal protein S3	243	ELAEDGYSGVEVR	28 - 40
					TEIIILATR	46 - 54
					FGFPEGSVELYAEK	77 - 90
					GCEVVVSGK	133 - 141
P62862	RPS30	eS30	40S ribosomal protein S30	59	FVNVVPTFGK	42 - 51
P97351	RPS3A	eS1	40S ribosomal protein S3a	264	LITEDVQGK	86 - 94
					EVQTNDLK	175 - 182
P62702	RPS4X	eS4	40S ribosomal protein S4, X isoform	263	GIPHLVTHDAR	135 - 145
					LSNIFVIGK	222 - 230
					LTIAEER	246 - 252
P97461	RPS5	uS7	40S ribosomal protein S5	204	YLPHSAGR	48 - 55
					AQCPIVER	64 - 71
					VNQAIWLLCTGAR	147 - 159
P62754	RPS6	eS6	40S ribosomal protein S6	249	LNISFPATGCQK	3 - 14
					EEAAEYAK	204 - 211
P62082	RPS7	eS7	40S ribosomal protein S7	194	ELNITAAK	42 - 49
					KAIIIFVPVPQLK	58 - 70
					AIIIFVPVPQLK	59 - 70
					HVVFIAQR	91 - 98

					TLTAVHDAILEDLVFPSEIVGK	121 - 142
P62242	RPS8	eS8	40S ribosomal protein S8	208	IIDVVYNASNNELVR	78 - 92
					NCIVLIDSTPYR	99 - 110
					ISSLLEEQFQQGK	158 - 170
					ADGYVLEGK	185 - 193
Q6ZWN5	RPS9	uS4	40S ribosomal protein S9	194	LIGEYGLR	31 - 38
					IGVLDEGK	84 - 91
					LDYILGLK	94 - 101
P14206	RPSSA	uS2	40S ribosomal protein SA	295	AIVAIENPADVSVISSR	64 - 80
					FAAATGATPIAGR	90 - 102
					FTPGTFTNQIQAAFR	103 - 117
					ADHQPLTEASYVNLPTIALCNTDSPLR	129 - 155

Table 4-1 List of Peptides of Ribosomal Proteins Quantified

Chapter 5:

Conclusion

This thesis tries to integrate two aspects of our lab's focus on mass spectrometry based proteomics: (1) label-free protein quantification strategies that are reliable, reproducible, and robust; and (2) data acquisition methods that are either discovery or targeted in nature to analyze and quantify proteins from samples of various complexity, from purified cellular component extract to highly complex cell lysates.

In Chapter 2, the usage of label-free protein quantification using primarily the DDA method to explore the proteomic changes in cells with overexpression of miR-23∼24 is described. A complete comparison of the proteomes of vector control cells or cells expressing the miR-23–24 cluster was performed using DDA acquisition mode and the resulting chromatographic data were quantified by integrating the area under the curve of precursor ions (MS1). Protein abundances were normalized between replicates and treatment groups and compared for differential expression in a label-free manner. The analysis resulted in the identification and quantification of 4349 proteins, the abundance of which varied by over five orders of magnitude. Our data analysis to understand the most significant inhibition of miRNAs from the miR-2324 cluster did identify previously known targets of the cluster as well as more potential targets that exhibit reduced expression by over an order of magnitude upon miR-2324 cluster expression. While label-free proteomics is traditionally not viewed to be as quantitative as and more problematic than isotopic labeling techniques, such as iTraq, we demonstrate that it is of high utility in identifying the largest magnitude changes, which are desired for the identification of the most potent targets for miRNA regulation. Our approach of utilizing label-free quantification coupled with normalization by representing protein abundances as a proportion of TIC allows us to increase data reproducibility and circumvent the disadvantages of using isotopic labeling

techniques that can suppress observed expression differences as revealed in previous proteomic investigations on this cluster using iTraq labeling, which only reported on changes of less than 2-fold in protein abundance.

Chapter 3 discussed the proteomic comparison of human serum samples to identify novel proteins in an unbiased manner associated with Myasthenia gravis (MG). The first part of the study involved proteomic data collection through the DDA method using unadulterated human serum samples. We conducted a three way comparison between healthy controls, MG patients, and rheumatoid arthritis (RA) patients, used as a reference autoimmune disease, which identified a total of 502 proteins among all groups sampled. Among those proteins, 11 proteins showed a statistically significant difference between the control, MG, and RA groups. To further validate our finding, we utilized a targeted proteomics acquisition method, PRM, to reliably quantify the proteins exhibiting the largest differential abundance in our DDA analysis, FGA, FGB, and FGG. We analyzed a larger cohort of MG patient serum samples consisting of various subtypes (27 anti-AChR antibody positive, one anti-MuSK antibody positive, three seronegatives, for a total of 31 MG sera), compared against 18 from RA patients and 30 sera from the control group. Our PRM analysis showed fibrinogens were present in consistently higher abundance in MG patient sera compared to RA and control groups; 28 to 54-fold higher than the RA samples and 250 to 4000-fold higher than the control samples. Meanwhile, no difference were observed in the abundance of fibrinogens was observed between the different subtypes of MG. Through utilizing two different acquisition method, DDA and PRM, we were able to initially identify potential novel proteins associated with MG as well as quantitatively determine the abundance differential of fibrinogen in MG patients' sera. We found the presence of high serum fibrinogen levels to be highly sensitive to MG with specificities that appear to be superior to current clinical testing methods. Furthermore,

high serum fibrinogen levels were found in all serotypes and in Class I patients (a form of MG that only affects the eyes), suggesting that residual serum testing might be more useful than the current serological tests. We believe that our findings will serve as the cornerstone of a consistent method for diagnosing clinical MG.

In Chapter 4, we explored the proteomic quantification of ribosomal proteins expressed in varying cells and tissue types to study the differences in ribosomal protein abundance in tissues. We performed a targeted proteomic analysis by PRM on purified ribosomal fractions from tissue lysates to quantify 75 ribosomal proteins from 210 peptides. The list of peptides were finalized from 379 peptides and 519 precursors from 84 proteins following repeated method optimization. Our resulting analysis were able to show that the ribosome complexes thought to be constitutive were indeed made up of several ribosomal proteins that are not expressed equally in all tissue. This further fuels the emerging notion of ribosomal heterogeneity that plays a role towards translational regulation providing an additional point of specification and specialization of tissues. We additionally discovered the small ribosomal protein, eS10 showed peptides absence in certain tissues. Further investigation uncovered the potential of an isoform of eS10 expressed in liver and kidney adding another level of heterogeneity to ribosome complexes.

The comparison and relative quantification of proteins originating from various biological samples using our label-free proteomic technique is the main focus of this thesis. While the basic idea behind label-free quantification is by no means new (for reviews, see (Bantscheff et al., 2007, 2012; Domon & Aebersold, 2006; Listgarten & Emili, 2005; Nahnsen et al., 2013), our method combines widely used individual protein quantification methods (Kramer et al., 2017, 2018; Piragasam et al., 2020) with a sample-specific normalization procedure for all observed proteins.

It has been suggested in the past that this technique may have issues when it comes to the prefractionation of proteins found in biological samples (Cox et al., 2014). Theoretically, different protein/peptide combinations in each sub-fraction may behave differently during chromatographic separation, producing results that are difficult to reproduce. To lessen this sub-fraction variation, however, a number of measures can be taken. Though quantification through label-free methods is not direct and more challenging compared to the direct comparison advantage of isotope labeling methods, label-free quantification has become more popular and adopted by the proteomics community in the last few years which can be attributed to the explosion of technological advances in mass spectrometry (increase in instrument sensitivity capable of accurate ion quantification) and in computational performance in the past decade. The trend can be easily noticed by performing keyword searches in the PubMed database for both "label-free proteomics" and "isotope proteomics", where at the time of writing this thesis, publications with the "label-free proteomics" keyword have almost tripled from 2015, 1726 in 2015 to 5160 currently. During the same period of time, the "isotope proteomics" keyword only showed an increase of approximately 1.5 times, from 3676 in 2015 to 5739 currently. Our lab has repeatedly shown that reliable semi-quantitative comparisons can be made that are verifiable when tested visually using immunoblotting techniques or through targeted proteomic quantification (see Chapters 2 and 3 and (McRae et al., 2020; Piragasam et al., 2020)). This has been accomplished by carefully controlling sample pre-fractionation, including how samples are fractionated, and subsequent front-end preparation for replicate samples, as well as careful normalization following data acquisition.

Our aim of focusing on label-free quantification does come with challenges inherent to the approach that become apparent in the MS data analysis stage. The variations in the acquired data caused by systematic biases introduced during various stages of sample processing and data

generation are one of the challenges. The results of quantitative analysis might be skewed if these biases are not taken into consideration. Therefore, careful processing of data by accounting for common problems such as missing data and correcting for instrument variability are necessary.

The method to properly handle missing value/data (MVs) in the field of proteomics have been explored for an extended period of time by many groups. Protein abundances below the instrument detection limit, the absence of proteins, sample loss during preparation, incorrect peptide cleavage during digestion, poor ionization efficiency, and poor peptide-spectrum matches are just a few of the biological and analytical factors that can cause MVs in LC-MS-based proteomics. Large datasets with MVs present a rather special problem because it is impossible for the user to know with absolute certainty whether a value is missing due to total absence or presence below the limit of detection (missing not at random; MNAR), or whether it is present but missing-ness due to the accumulation of stochastic variability/error (missing at random; MAR, or missing completely at random; MCAR) (Karpievitch et al., 2012; Lazar et al., 2016; Rubin, 1976; Webb-Robertson et al., 2015). In proteomics, MNAR may result from experimental effects such as (1) true presence/absence, (2) enzyme miscleavages, and (3) effects from instruments (dynamic range or limit of detection occurring when peptide measurements are low in abundance compared to background noise or constitute low ionization efficiency). MCAR can happen anywhere along the data distribution and is unaffected by the data or observed values. This kind of missingness results from inaccurate instrumentation or mistakes made during the preparation of experimental samples. Compared to MCAR, MAR covers a wider range of missing-ness and is typically conditionally dependent on the observed values. It usually occurs when a peptide sequence is incorrectly mapped or software errors resulting in misidentification of peptide in some samples creating MVs in other samples.

Real label-free proteomic datasets frequently include MVs in the 10–50% range (Albrecht et al., 2010; Lazar et al., 2016), typically existing as a mix of MCAR and MNAR. In order to deal with missing values, there are two options: either removing peptides or proteins from the analysis if there are not enough samples for them to be analyzed, or impute values in their place. For imputing values replacing MVs, several solutions have been developed. The most common solutions are to replace MVs with zeroes, the local minimum (LM) observed in a single sample, or the global minimum (GM) observed across all samples in the comparison (Lazar et al., 2016; Webb-Robertson et al., 2015). However, there is no single method that is fit-for-all or functions the best thus requiring case-by-case determination of suitable solution to handle MVs. Maximum differentiation of protein abundances is made possible by the use of zeroes and GMs in imputation, but in datasets with high MV content, these techniques can introduce significant measurement errors and lower statistical power. Another shortcoming of inputting zeroes is that the following calculations and statistical analysis will be restricted by the division by zero problem. Inputting LMs to replace MVs can overcome problems encountered through the usage of zeroes and GMs. However, there are possibilities of introducing artificial differences in datasets as there may be significant differences in LMs between samples. This can lead to skewed statistical findings and a biased conclusion.

Our approach for MV imputation in this thesis involves considering the dynamics of the data and analysis goal to determine a suitable solution. In Chapter 2, we used LMs imputation due to large number of proteins in the datasets as well as relatively small difference in LMs between the replicate samples. Zeroes were imputed for analysis in Chapter 3 where the data identified proteins with large dynamic range with several proteins of interest were mostly highly abundant. In Chapter 4, our targeted proteomic approach utilizing PRM is inherently better than DDA in

terms of the frequency of MVs observed(Gallien et al., 2015). Therefore, we focused on a combination of methods of addressing MVs, which are inputting zeros and removing peptides and fragments from analysis following identification of MVs.

My research has concentrated on creating effective, dependable, and repeatable methods for using mass spectrometry to carry out bottom-up label-free quantitative proteomics over the past few years. This thesis explores the utility of label-free proteomics quantification over a wide range biological functions that are applicable towards improvement of knowledge in laboratory settings and clinical diagnostic application by addressing various questions such as proteomic changes resulting from miRNAs, identifying biomarkers for neuromuscular disease, and identifying heterogeneity in ribosomal proteins. We also demonstrated the use of the technique on various samples with different characteristics and complexities, including whole cell lysates, fractionated tissue lysates, and human serum. Proteomic studies can now be performed at greater depth thanks to newer techniques and technologies, but their reliance on cutting-edge equipment can sometimes be a drawback. Nevertheless, the methods utilized in this thesis are practical can be applied to answer wide range of biological questions further assisted by the development in the field of proteomics.

Finally, the fundamental purpose of research is to uncover knowledge that can benefit humans. Though not all research impact the progression of human civilization in similar level, the pursue of knowledge through research should always be preserved. Only then new discoveries leading to accumulation of knowledge can occur and further application of knowledge to enhance humanity can be pursued. One such opportunity emerged within our team. The results from our search for biomarker in characterizing Myasthenia Gravis uncovered several candidate proteins that are potentially useful as for identification of all types of the disease. Therefore, our team have

decided to develop a diagnostic device based on our findings in the hope of future commercialization. Our spin-off company, named Decipher MedTech Inc., has filed a patent based on our findings, and we are currently working on the production of the device. I am proud that the work described in this thesis will potentially benefit human life in the near future with the successful development of the diagnostic device for Myasthenia Gravis.

5.1 Future of Quantitative Proteomics

Quantitative proteomics is a rapidly evolving field that can potentially transform our grasp of disease progress and biological systems. It can be further enhanced with development in several key areas. One such is advancement in mass spectrometry technology. New mass spectrometers are constantly being developed and instruments with better sensitivity and resolution as well as throughput allows for enhancement in quantification accuracy and quality. Additionally, quantitative proteomics data can be integrated with other omics data, such as genomics, metabolomics, transcriptomics to provide better view and enhance understanding of biological systems. Muiltiple omics data integration can provide a more detailed insights into interactions of various molecules in a biological system, hence providing solutions to many unanswered question in the field of biology and medicine, such as identification of new disease biomarkers and therapeutic targets. Currently, quantitative proteomics require substantial sample preparation before data collection and data analysis. However, advancement in this area is occurring rapidly and a simplified workflow for protein quantification by mass spectrometry can materialize in the very near future. This will enable quantitative proteomics to be readily applied in clinical research

as well as medical setting where development of personalized medicine will be the focus in the next decade.

References

Aarli, J. A. (2008). Myasthenia gravis in the elderly: Is it different? *Annals of the New York Academy of Sciences*, *1132*, 238–243. https://doi.org/10.1196/ANNALS.1405.040

Abbott, S. A. (2010). Diagnostic challenge: Myasthenia gravis in the emergency department. *Journal of the American Academy of Nurse Practitioners*, *22*(9), 468–473. https://doi.org/10.1111/J.1745-7599.2010.00541.X

Addona, T. A., Abbatiello, S. E., Schilling, B., Skates, S. J., Mani, D. R., Bunk, D. M., Spiegelman, C. H., Zimmerman, L. J., Ham, A. J. L., Keshishian, H., Hall, S. C., Allen, S., Blackman, R. K., Borchers, C. H., Buck, C., Cardasis, H. L., Cusack, M. P., Dodder, N. G., Gibson, B. W., … Carr, S. A. (2009). Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nature Biotechnology*, *27*(7), 633–641. https://doi.org/10.1038/NBT.1546

Aebersold, R., & Mann, M. (2003). Mass spectrometry-based proteomics. *Nature 2003 422:6928*, *422*(6928), 198–207. https://doi.org/10.1038/nature01511

Agarwal, V., Bell, G. W., Nam, J. W., & Bartel, D. P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *ELife*, *4*(AUGUST2015). https://doi.org/10.7554/ELIFE.05005

Ahrné, E., Molzahn, L., Glatter, T., & Schmidt, A. (2013). Critical assessment of proteome-wide label-free absolute abundance estimation strategies. *Proteomics*, *13*(17), 2567–2578. https://doi.org/10.1002/PMIC.201300135

Alaoui-Jamali, M. A., & Xu, Y. jie. (2006). Proteomic technology for biomarker profiling in cancer: an update. *Journal of Zhejiang University. Science. B*, *7*(6), 411–420. https://doi.org/10.1631/JZUS.2006.B0411

Albrecht, D., Kniemeyer, O., Brakhage, A. A., & Guthke, R. (2010). Missing values in gel-based proteomics. *Proteomics*, *10*(6), 1202–1211. https://doi.org/10.1002/PMIC.200800576

al Shweiki, M. H. D. R., Mönchgesang, S., Majovsky, P., Thieme, D., Trutschel, D., & Hoehenwarter, W. (2017). Assessment of Label-Free Quantification in Discovery Proteomics and Impact of Technological Factors and Natural Variability of Protein Abundance. *Journal of Proteome Research*, *16*(4), 1410–1424. https://doi.org/10.1021/ACS.JPROTEOME.6B00645/SUPPL_FILE/PR6B00645_SI_015.XLSX

Ames, G. F. L., & Nikaido, K. (1976). Two-Dimensional Gel Electrophoresis of Membrane Proteins. *Biochemistry*, *15*(3), 616–623. https://doi.org/10.1021/bi00648a026

Analysis Of Variance (ANOVA) | Introduction, Types & Techniques. (n.d.). Retrieved May 7, 2022, from https://www.analyticsvidhya.com/blog/2018/01/anova-analysis-of-variance/

Anderson, N. L., Polanski, M., Pieper, R., Gatlin, T., Tirumalai, R. S., Conrads, T. P., Veenstra, T. D., Adkins, J. N., Pounds, J. G., Fagan, R., & Lobley, A. (2004). The human plasma proteome: a nonredundant list developed by combination of four separate sources. *Molecular & Cellular Proteomics : MCP*, *3*(4), 311–326. https://doi.org/10.1074/MCP.M300127-MCP200

ANOVA -- from Wolfram MathWorld. (n.d.). Retrieved May 7, 2022, from https://mathworld.wolfram.com/ANOVA.html

Arike, L., & Peil, L. (2014). Spectral counting label-free proteomics. *Methods in Molecular Biology (Clifton, N.J.)*, *1156*, 213–222. https://doi.org/10.1007/978-1-4939-0685-7_14

Arnett, F. C., Edworthy, S. M., Bloch, D. A., Mcshane, D. J., Fries, J. F., Cooper, N. S., Healey, L. A., Kaplan, S. R., Liang, M. H., Luthra, H. S., Medsger, T. A., Mitchell, D. M., Neustadt, D. H., Pinals, R. S., Schaller, J. G., Sharp, J. T., Wilder, R. L., & Hunder, G. G. (1988). The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis and Rheumatism*, *31*(3), 315–324. https://doi.org/10.1002/ART.1780310302

Arpino, P. J., & Guiochon, G. (1979). LC/MS coupling. *Analytical Chemistry*, *51*(7), 682–701. https://doi.org/10.1021/AC50043A001

Azimifar, S. B., Nagaraj, N., Cox, J., & Mann, M. (2014). Cell-type-resolved quantitative proteomics of murine liver. *Cell Metabolism*, *20*(6), 1076–1087. https://doi.org/10.1016/J.CMET.2014.11.002

Bang, C., Fiedler, J., & Thum, T. (2012). Cardiovascular Importance of the MicroRNA-23/27/24 Family. *Microcirculation*, *19*(3), 208–214. https://doi.org/10.1111/J.1549-8719.2011.00153.X

Bantscheff, M., Lemeer, S., Savitski, M. M., & Kuster, B. (2012). Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Analytical and Bioanalytical Chemistry*, *404*(4), 939–965. https://doi.org/10.1007/S00216-012-6203-4

Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., & Kuster, B. (2007). Quantitative mass spectrometry in proteomics: a critical review. *Analytical and Bioanalytical Chemistry*, *389*(4), 1017–1031. https://doi.org/10.1007/S00216-007-1486-6

Bartke, T., Borgel, J., & DiMaggio, P. A. (2013). Proteomics in epigenetics: new perspectives for cancer research. *Briefings in Functional Genomics*, *12*(3), 205–218. https://doi.org/10.1093/BFGP/ELT002

Bateman, A., Martin, M. J., O'Donovan, C., Magrane, M., Apweiler, R., Alpi, E., Antunes, R., Arganiska, J., Bely, B., Bingley, M., Bonilla, C., Britto, R., Bursteinas, B., Chavali, G., Cibrian-Uhalte, E., da Silva, A., de Giorgi, M., Dogan, T., Fazzini, F., … Zhang, J. (2015). UniProt: a hub for protein information. *Nucleic Acids Research*, *43*(Database issue), D204–D212. https://doi.org/10.1093/NAR/GKU989

Bateman, N. W., Goulding, S. P., Shulman, N. J., Gadok, A. K., Szumlinski, K. K., MacCoss, M. J., & Wu, C. C. (2014). Maximizing peptide identification events in proteomic workflows using data-dependent acquisition (DDA). *Molecular & Cellular Proteomics : MCP*, *13*(1), 329–338. https://doi.org/10.1074/MCP.M112.026500

Beavis, R. C. (2006). Using the global proteome machine for protein identification. *Methods in Molecular Biology (Clifton, N.J.)*, *328*, 217–228. https://doi.org/10.1385/1-59745-026-X:217

Becker, A., Ludwig, N., Keller, A., Tackenberg, B., Eienbröker, C., Oertel, W. H., Fassbender, K., Meese, E., & Ruprecht, K. (2013). Myasthenia Gravis: Analysis of Serum Autoantibody Reactivities to 1827 Potential Human Autoantigens by Protein Macroarrays. *PLOS ONE*, *8*(3), e58095. https://doi.org/10.1371/JOURNAL.PONE.0058095

Becker, G. W. (2008). Stable isotopic labeling of proteins for quantitative proteomic applications. *Briefings in Functional Genomics*, *7*(5), 371–382. https://doi.org/10.1093/BFGP/ELN047

Beekman, R., Kuks, J. B. M., & Oosterhuis, H. J. G. H. (1997). Myasthenia gravis: diagnosis and follow-up of 100 consecutive patients. *Journal of Neurology*, *244*(2), 112–118. https://doi.org/10.1007/S004150050059

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/J.2517-6161.1995.TB02031.X

Bertsch, A., Gröpl, C., Reinert, K., & Kohlbacher, O. (2011). OpenMS and TOPP: open source software for LC-MS data analysis. *Methods in Molecular Biology (Clifton, N.J.)*, *696*, 353–367. https://doi.org/10.1007/978-1-60761-987-1_23

Bhardwaj, C., & Hanley, L. (2014). Ion sources for mass spectrometric identification and imaging of molecular species. *Natural Product Reports*, *31*(6), 756–767. https://doi.org/10.1039/C3NP70094A

Biemann, K. (1986). Mass Spectrometric Methods for Protein Sequencing. *Analytical Chemistry*, *58*(13), 1288 A-1300 A. https://doi.org/10.1021/AC00126A001/ASSET/AC00126A001.FP.PNG_V03

Blackmore, D., Siddiqi, Z., Li, L., Wang, N., & Maksymowych, W. (2019). Beyond the antibodies: serum metabolomic profiling of myasthenia gravis. *Metabolomics*, *15*(8), 1–12. https://doi.org/10.1007/S11306-019-1571-9/FIGURES/5

Bland, j. M., & Altman, D. G. (1995). Multiple significance tests: the Bonferroni method. *BMJ*, *310*(6973), 170. https://doi.org/10.1136/BMJ.310.6973.170

Bodzon-Kulakowska, A., Bierczynska-Krzysik, A., Dylag, T., Drabik, A., Suder, P., Noga, M., Jarzebinska, J., & Silberring, J. (2007). Methods for samples preparation in proteomic research. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences*, *849*(1–2), 1–31. https://doi.org/10.1016/J.JCHROMB.2006.10.040

Bondarenko, P. v., Chelius, D., & Shaler, T. A. (2002). Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry. *Analytical Chemistry*, *74*(18), 4741–4749. https://doi.org/10.1021/AC0256991

Breci, L. A., Tabb, D. L., Yates, J. R., & Wysocki, V. H. (2003). Cleavage N-Terminal to Proline: Analysis of a Database of Peptide Tandem Mass Spectra. *Analytical Chemistry*, *75*(9), 1963–1971. https://doi.org/10.1021/AC026359I

Brenton, A. G., & Godfrey, A. R. (2010). Accurate Mass Measurement: Terminology and Treatment of Data. *Journal of the American Society for Mass Spectrometry*, *21*(11), 1821–1835. https://doi.org/10.1016/J.JASMS.2010.06.006

Bruderer, R., Bernhardt, O. M., Gandhi, T., Miladinović, S. M., Cheng, L. Y., Messner, S., Ehrenberger, T., Zanotelli, V., Butscheid, Y., Escher, C., Vitek, O., Rinner, O., & Reiter, L. (2015). Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Molecular & Cellular Proteomics : MCP*, *14*(5), 1400–1410. https://doi.org/10.1074/MCP.M114.044305

B, S., D, B., N, L., G, D., J, S., J, W., W, C., & M, S. (2013). Corrigendum: Global quantification of mammalian gene expression control. *Nature*, *495*(7439), 126–127. https://doi.org/10.1038/NATURE11848

Carroll, D. I., Dzidic, I., Stillwell, R. N., Horning, M. G., & Horning, E. C. (1974). Subpicogram Detection System for Gas Phase Analysis Based upon Atmospheric Pressure Ionization (API) Mass Spectrometry. *Analytical Chemistry*, *46*(6), 706–710. https://doi.org/10.1021/AC60342A009/ASSET/AC60342A009.FP.PNG_V03

Catenacci, D. V. T., Liao, W. L., Thyparambil, S., Henderson, L., Xu, P., Zhao, L., Rambo, B., Hart, J., Xiao, S. Y., Bengali, K., Uzzell, J., Darfler, M., Krizman, D. B., Cecchi, F., Bottaro, D. P., Karrison, T., Veenstra, T. D., Hembrough, T., & Burrows, J. (2014). Absolute Quantitation of Met Using Mass Spectrometry for Clinical Application: Assay Precision, Stability, and Correlation with MET Gene Amplification in FFPE Tumor Tissue. *PLOS ONE*, *9*(7), e100586. https://doi.org/10.1371/JOURNAL.PONE.0100586

Celis, J. E., & Gromov, P. (1999). 2D protein electrophoresis: can it be perfected? *Current Opinion in Biotechnology*, *10*(1), 16–21. https://doi.org/10.1016/S0958-1669(99)80004-4

Chahrour, O., Cobice, D., & Malone, J. (2015). Stable isotope labelling methods in mass spectrometry-based quantitative proteomics. *Journal of Pharmaceutical and Biomedical Analysis*, *113*, 2–20. https://doi.org/10.1016/J.JPBA.2015.04.013

Chaillou, T., Kirby, T. J., & Mccarthy, J. J. (2014). Ribosome biogenesis: emerging evidence for a central role in the regulation of skeletal muscle mass. *Journal of Cellular Physiology*, *229*(11), 1584. https://doi.org/10.1002/JCP.24604

Chaillou, T., Zhang, X., & Mccarthy, J. J. (2016). Expression of muscle-specific ribosomal protein L3-like impairs myotube growth. *Journal of Cellular Physiology*, *231*(9), 1894. https://doi.org/10.1002/JCP.25294

Chaulk, S. G., Ebhardt, H. A., & Fahlman, R. P. (2015). Correlations of microRNA:microRNA expression patterns reveal insights into microRNA clusters and global microRNA expression patterns. *Molecular BioSystems*, *12*(1), 110–119. https://doi.org/10.1039/C5MB00415B

Chaulk, S. G., Lattanzi, V. J., Hiemer, S. E., Fahlman, R. P., & Varelas, X. (2014). The Hippo Pathway Effectors TAZ/YAP Regulate Dicer Expression and MicroRNA Biogenesis through Let-7. *The Journal of Biological Chemistry*, *289*(4), 1886. https://doi.org/10.1074/JBC.C113.529362

Chaulk, S. G., Thede, G. L., Kent, O. A., Xu, Z., Gesner, E. M., Veldhoen, R. A., Khanna, S. K., Goping, I. S., MacMillan, A. M., Mendell, J. T., Young, H. S., Fahlman, R. P., & Glover, J. N. M. (2011). Role of pri-miRNA tertiary structure in miR-17~92 miRNA biogenesis. *Http://Dx.Doi.Org/10.4161/Rna.8.6.17410*, *8*(6), 1105–1114. https://doi.org/10.4161/RNA.8.6.17410

Chaulk, S. G., Xu, Z., Glover, M. J. N., & Fahlman, R. P. (2014). MicroRNA miR-92a-1 biogenesis and mRNA targeting is modulated by a tertiary contact within the miR-17~92 microRNA cluster. *Nucleic Acids Research*, *42*(8), 5234. https://doi.org/10.1093/NAR/GKU133

Chhabra, R., Dubey, R., & Saini, N. (2010). Cooperative and individualistic functions of the microRNAs in the miR-23a~27a~24-2 cluster and its implication in human diseases. *Molecular Cancer*, *9*, 232. https://doi.org/10.1186/1476-4598-9-232

Cho, S., Wu, C. J., Yasuda, T., Cruz, L. O., Khan, A. A., Lin, L. L., Nguyen, D. T., Miller, M., Lee, H. M., Kuo, M. L., Broide, D. H., Rajewsky, K., Rudensky, A. Y., & Lu, L. F. (2016). miR-23~27~24 clusters control effector T cell differentiation and function. *The Journal of Experimental Medicine*, *213*(2), 235–249. https://doi.org/10.1084/JEM.20150990

Chou, C. H., Shrestha, S., Yang, C. D., Chang, N. W., Lin, Y. L., Liao, K. W., Huang, W. C., Sun, T. H., Tu, S. J., Lee, W. H., Chiew, M. Y., Tai, C. S., Wei, T. Y., Tsai, T. R., Huang, H. T., Wang, C. Y., Wu, H. Y., Ho, S. Y., Chen, P. R., … Huang, H. da. (2018). miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Research*, *46*(Database issue), D296. https://doi.org/10.1093/NAR/GKX1067

Christoforou, A., Mulvey, C. M., Breckels, L. M., Geladaki, A., Hurrell, T., Hayward, P. C., Naake, T., Gatto, L., Viner, R., Arias, A. M., & Lilley, K. S. (2016). A draft map of the mouse pluripotent stem cell spatial proteome. *Nature Communications 2016 7:1*, *7*(1), 1–12. https://doi.org/10.1038/ncomms9992

Colangelo, T., Polcaro, G., Ziccardi, P., Pucci, B., Muccillo, L., Galgani, M., Fucci, A., Milone, M. R., Budillon, A., Santopaolo, M., Votino, C., Pancione, M., Piepoli, A., Mazzoccoli, G., Binaschi, M., Bigioni, M., Maggi, C. A., Fassan, M., Laudanna, C., … Colantuoni, V. (2016). Proteomic screening identifies calreticulin as a miR-27a direct target repressing MHC class I cell surface exposure in colorectal cancer. *Cell Death & Disease*, *7*(2). https://doi.org/10.1038/CDDIS.2016.28

Cox, J., Hein, M. Y., Luber, C. A., Paron, I., Nagaraj, N., & Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction,

termed MaxLFQ. *Molecular & Cellular Proteomics : MCP*, *13*(9), 2513–2526. https://doi.org/10.1074/MCP.M113.031591

Cox, J., & Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology 2008 26:12*, *26*(12), 1367–1372. https://doi.org/10.1038/nbt.1511

Craig Venter, J., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., … Zhu, X. (2001). The sequence of the human genome. *Science (New York, N.Y.)*, *291*(5507), 1304–1351. https://doi.org/10.1126/SCIENCE.1058040

Dalakas, M. C. (2019). Immunotherapy in myasthenia gravis in the era of biologics. *Nature Reviews. Neurology*, *15*(2), 113–124. https://doi.org/10.1038/S41582-018-0110-Z

Demartini, D. R. (2013). A Short Overview of the Components in Mass Spectrometry Instrumentation for Proteomics Analyses. *Tandem Mass Spectrometry - Molecular Characterization*. https://doi.org/10.5772/54484

Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S., & Ralser, M. (2019). DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nature Methods 2019 17:1*, *17*(1), 41–44. https://doi.org/10.1038/s41592-019-0638-x

Dempster, A. J. (1918). A new Method of Positive Ray Analysis. *Physical Review*, *11*(4), 316. https://doi.org/10.1103/PhysRev.11.316

Desiderio, D. M., & Kai, M. (1983). Preparation of stable isotope-incorporated peptide internal standards for field desorption mass spectrometry quantification of peptides in biologic tissue. *Biomedical Mass Spectrometry*, *10*(8), 471–479. https://doi.org/10.1002/BMS.1200100806

Deutsch, E. W., Lam, H., & Aebersold, R. (2008). PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. *EMBO Reports*, *9*(5), 429–434. https://doi.org/10.1038/EMBOR.2008.56

Deutsch, E. W., Mendoza, L., Shteynberg, D., Slagel, J., Sun, Z., & Moritz, R. L. (2015). Trans-Proteomic Pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics. *Proteomics. Clinical Applications*, *9*(7–8), 745–754. https://doi.org/10.1002/PRCA.201400164

Domon, B., & Aebersold, R. (2006). Mass spectrometry and protein analysis. *Science (New York, N.Y.)*, *312*(5771), 212–217. https://doi.org/10.1126/SCIENCE.1124619

Domon, B., & Aebersold, R. (2010). Options and considerations when selecting a quantitative proteomics strategy. *Nature Biotechnology 2010 28:7*, *28*(7), 710–721. https://doi.org/10.1038/nbt.1661

Downard, K. M. (2007). Historical account: Francis William Aston: the man behind the mass spectrograph. *European Journal of Mass Spectrometry (Chichester, England)*, *13*(3), 177–190. https://doi.org/10.1255/EJMS.878

D.Sc., F. W. A. M. A. (2009). LXXIV. A positive ray spectrograph. *Https://Doi.Org/10.1080/14786441208636004*, *38*(228), 707–714. https://doi.org/10.1080/14786441208636004

Dunn, O. J. (1961). Multiple Comparisons among Means. *Journal of the American Statistical Association*, *56*(293), 52–64. https://doi.org/10.1080/01621459.1961.10482090

Edfors, F., Danielsson, F., Hallström, B. M., Käll, L., Lundberg, E., Pontén, F., Forsström, B., & Uhlén, M. (2016). Gene-specific correlation of RNA and protein levels in human cells and tissues. *Molecular Systems Biology*, *12*(10), 883. https://doi.org/10.15252/MSB.20167144

EDMAN, P. (1949). A method for the determination of amino acid sequence in peptides. *Archives of Biochemistry*, *22*(3), 475. https://pubmed.ncbi.nlm.nih.gov/18134557/

Edman, P., & Begg, G. (1967). A Protein Sequenator. *European Journal of Biochemistry*, *1*(1), 80–91. https://doi.org/10.1111/J.1432-1033.1967.TB00047.X

Efron, B., Tibshirani, R., Storey, J. D., & Tusher, V. (2011). Empirical Bayes Analysis of a Microarray Experiment. *Https://Doi.Org/10.1198/016214501753382129*, *96*(456), 1151–1160. https://doi.org/10.1198/016214501753382129

Egertson, J. D., MacLean, B., Johnson, R., Xuan, Y., & MacCoss, M. J. (2015). Multiplexed peptide analysis using data-independent acquisition and Skyline. *Nature Protocols 2015 10:6*, *10*(6), 887–903. https://doi.org/10.1038/nprot.2015.055

Eldeeb, M. A., & Fahlman, R. P. (2016). Phosphorylation Impacts N-end Rule Degradation of the Proteolytically Activated Form of BMX Kinase. *The Journal of Biological Chemistry*, *291*(43), 22757. https://doi.org/10.1074/JBC.M116.737387

Eliuk, S., & Makarov, A. (2015). Evolution of Orbitrap Mass Spectrometry Instrumentation. *Annual Review of Analytical Chemistry (Palo Alto, Calif.)*, *8*, 61–80. https://doi.org/10.1146/ANNUREV-ANCHEM-071114-040325

Ell, B., Qiu, Q., Wei, Y., Mercatali, L., Ibrahim, T., Amadori, D., & Kang, Y. (2014). The microRNA-23b/27b/24 cluster promotes breast cancer lung metastasis by targeting metastasis-suppressive gene prosaposin. *The Journal of Biological Chemistry*, *289*(32), 21888–21895. https://doi.org/10.1074/JBC.M114.582866

Evoli, A., Iorio, R., & Bartoccioni, E. (2015). Overcoming challenges in the diagnosis and treatment of myasthenia gravis. *Http://Dx.Doi.Org/10.1586/1744666X.2016.1110487*, *12*(2), 157–168. https://doi.org/10.1586/1744666X.2016.1110487

Fabian, M. R., & Sonenberg, N. (2012). The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. *Nature Structural & Molecular Biology 2012 19:6*, *19*(6), 586–593. https://doi.org/10.1038/nsmb.2296

Fabre, B., Lambour, T., Bouyssié, D., Menneteau, T., Monsarrat, B., Burlet-Schiltz, O., & Bousquet-Dubouch, M. P. (2014). Comparison of label-free quantification methods for the determination of protein complexes subunits stoichiometry. *EuPA Open Proteomics*, *4*, 82–86. https://doi.org/10.1016/J.EUPROT.2014.06.001

Fang, F., Sveinsson, O., Thormar, G., Granqvist, M., Askling, J., Lundberg, I. E., Ye, W., Hammarström, L., Pirskanen, R., & Piehl, F. (2015). The autoimmune spectrum of myasthenia gravis: a Swedish population-based study. *Journal of Internal Medicine*, *277*(5), 594–604. https://doi.org/10.1111/JOIM.12310

Faria, S. S., Morris, C. F. M., Silva, A. R., Fonseca, M. P., Forget, P., Castro, M. S., & Fontes, W. (2017). A Timely shift from shotgun to targeted proteomics and how it can be groundbreaking for cancer research. *Frontiers in Oncology*, *7*(FEB), 13. https://doi.org/10.3389/FONC.2017.00013/BIBTEX

Feijó Delgado, F., Cermak, N., Hecht, V. C., Son, S., Li, Y., Knudsen, S. M., Olcum, S., Higgins, J. M., Chen, J., Grover, W. H., & Manalis, S. R. (2013). Intracellular Water Exchange for Measuring the Dry Mass, Water Mass and Changes in Chemical Composition of Living Cells. *PLOS ONE*, *8*(7), e67590. https://doi.org/10.1371/JOURNAL.PONE.0067590

Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., & Whitehouse, C. M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. *Science (New York, N.Y.)*, *246*(4926), 64–71. https://doi.org/10.1126/SCIENCE.2675315

Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., & Whitehouse, C. M. (1990). Electrospray ionization–principles and practice. *Mass Spectrometry Reviews*, *9*(1), 37–70. https://doi.org/10.1002/MAS.1280090103

Field, F. H. (1968). Chemical Ionization Mass Spectrometry. *Accounts of Chemical Research*, *1*(2), 42–49. https://doi.org/10.1021/AR50002A002/ASSET/AR50002A002.FP.PNG_V03

Frank, J. (2000). The ribosome--a macromolecular machine par excellence. *Chemistry & Biology*, *7*(6). https://doi.org/10.1016/S1074-5521(00)00127-7

Fröhlich, T., & Arnold, G. J. (2011). Quantifying attomole amounts of proteins from complex samples by nano-LC and selected reaction monitoring. *Methods in Molecular Biology (Clifton, N.J.)*, *790*, 141–164. https://doi.org/10.1007/978-1-61779-319-6_11

F.R.S., L. R. (2009). XX. On the equilibrium of liquid conducting masses charged with electricity. *Https://Doi.Org/10.1080/14786448208628425*, *14*(87), 184–186. https://doi.org/10.1080/14786448208628425

F.R.S., S. J. J. T. O. M. (2009). XIX. Further experiments on positive rays. *Https://Doi.Org/10.1080/14786440808637325*, *24*(140), 209–253. https://doi.org/10.1080/14786440808637325

Fusaro, V. A., Mani, D. R., Mesirov, J. P., & Carr, S. A. (2009). Prediction of high-responding peptides for targeted protein assays by mass spectrometry. *Nature Biotechnology 2009 27:2*, *27*(2), 190–198. https://doi.org/10.1038/nbt.1524

Gallien, S., Duriez, E., Crone, C., Kellmann, M., Moehring, T., & Domon, B. (2012). Targeted proteomic quantification on quadrupole-orbitrap mass spectrometer. *Molecular & Cellular Proteomics : MCP*, *11*(12), 1709–1723. https://doi.org/10.1074/MCP.O112.019802

Gallien, S., Kim, S. Y., & Domon, B. (2015). Large-Scale Targeted Proteomics Using Internal Standard Triggered-Parallel Reaction Monitoring (IS-PRM)*. *Molecular & Cellular Proteomics*, *14*(6), 1630–1644. https://doi.org/10.1074/MCP.O114.043968

Gallien, S., Peterman, S., Kiyonami, R., Souady, J., Duriez, E., Schoen, A., & Domon, B. (2012). Highly multiplexed targeted proteomics using precise control of peptide retention time. *Proteomics*, *12*(8), 1122–1133. https://doi.org/10.1002/PMIC.201100533

Gao, P., Tchernyshyov, I., Chang, T. C., Lee, Y. S., Kita, K., Ochi, T., Zeller, K. I., de Marzo, A. M., van Eyk, J. E., Mendell, J. T., & Dang, C. v. (2009). c-Myc suppression of miR-23a/b enhances mitochondrial glutaminase expression and glutamine metabolism. *Nature*, *458*(7239), 762–765. https://doi.org/10.1038/NATURE07823

Geiger, T., & Clarke, S. (1987). Deamidation, isomerization, and racemization at asparaginyl and aspartyl residues in peptides. Succinimide-linked reactions that contribute to protein

degradation. *Journal of Biological Chemistry*, *262*(2), 785–794. https://doi.org/10.1016/S0021-9258(19)75855-4

Geiger, T., Wehner, A., Schaab, C., Cox, J., & Mann, M. (2012). Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Molecular & Cellular Proteomics : MCP*, *11*(3). https://doi.org/10.1074/MCP.M111.014050

Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W., & Gygi, S. P. (2003). Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(12), 6940–6945. https://doi.org/10.1073/PNAS.0832254100

Giansanti, P., Tsiatsiani, L., Low, T. Y., & Heck, A. J. R. (2016). Six alternative proteases for mass spectrometry–based proteomics beyond trypsin. *Nature Protocols 2016 11:5*, *11*(5), 993–1006. https://doi.org/10.1038/nprot.2016.057

Gilhus, N. E. (2016). Myasthenia Gravis. *Https://Doi.Org/10.1056/NEJMra1602678*, *375*(26), 2570–2581. https://doi.org/10.1056/NEJMRA1602678

Gilhus, N. E., Tzartos, S., Evoli, A., Palace, J., Burns, T. M., & Verschuuren, J. J. G. M. (2019). Myasthenia gravis. *Nature Reviews Disease Primers 2019 5:1*, *5*(1), 1–19. https://doi.org/10.1038/s41572-019-0079-y

Gilhus, N. E., & Verschuuren, J. J. (2015). Myasthenia gravis: subgroup classification and therapeutic strategies. *The Lancet. Neurology*, *14*(10), 1023–1036. https://doi.org/10.1016/S1474-4422(15)00145-3

Gillet, L. C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., Bonner, R., & Aebersold, R. (2012). Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Molecular & Cellular Proteomics : MCP*, *11*(6). https://doi.org/10.1074/MCP.O111.016717

Godovac-Zimmermann, J., & Brown, L. R. (n.d.). *PERSPECTIVES FOR MASS SPECTROMETRY AND FUNCTIONAL PROTEOMICS*. https://doi.org/10.1002/1098-2787

Granvogl, B., Plöscher, M., & Eichacker, L. A. (2007). Sample preparation by in-gel digestion for mass spectrometry-based proteomics. *Analytical and Bioanalytical Chemistry*, *389*(4), 991–1002. https://doi.org/10.1007/S00216-007-1451-4

Gruhler, A., Olsen, J. v., Mohammed, S., Mortensen, P., Færgeman, N. J., Mann, M., & Jensen, O. N. (2005). Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Molecular & Cellular Proteomics : MCP*, *4*(3), 310–327. https://doi.org/10.1074/MCP.M400219-MCP200

Grünenfelder, B., Rummel, G., Vohradsky, J., Röder, D., Langen, H., & Jenal, U. (2001). Proteomic analysis of the bacterial cell cycle. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(8), 4681–4686. https://doi.org/10.1073/PNAS.071538098/SUPPL_FILE/TABLE2_2.XLS

Guo, T., Kouvonen, P., Koh, C. C., Gillet, L. C., Wolski, W. E., Röst, H. L., Rosenberger, G., Collins, B. C., Blum, L. C., Gillessen, S., Joerger, M., Jochum, W., & Aebersold, R. (2015). Rapid mass spectrometric conversion of tissue biopsy samples into permanent quantitative digital proteome maps. *Nature Medicine 2015 21:4*, *21*(4), 407–413. https://doi.org/10.1038/nm.3807

Gupta, V., & Warner, J. R. (2014). Ribosome-omics of the human ribosome. *RNA*, *20*(7), 1004–1013. https://doi.org/10.1261/RNA.043653.113

Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., & Aebersold, R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology*, *17*(10), 994–999. https://doi.org/10.1038/13690

Haag, A. M. (2016). Mass analyzers and mass spectrometers. *Advances in Experimental Medicine and Biology*, *919*, 157–169. https://doi.org/10.1007/978-3-319-41448-5_7/FIGURES/8

Han, B., & Higgs, R. E. (2008). Proteomics: from hypothesis to quantitative assay on a single platform. Guidelines for developing MRM assays using ion trap mass spectrometers. *Briefings in Functional Genomics & Proteomics*, *7*(5), 340–354. https://doi.org/10.1093/BFGP/ELN032

Hayes, R. N., & Gross, M. L. (1990). Collision-induced dissociation. *Methods in Enzymology*, *193*(C), 237–263. https://doi.org/10.1016/0076-6879(90)93418-K

Haynes, W. (2013a). Bonferroni Correction. In O. and C. K.-H. and Y. H. Dubitzky Werner and Wolkenhauer (Ed.), *Encyclopedia of Systems Biology* (p. 154). Springer New York. https://doi.org/10.1007/978-1-4419-9863-7_1213

Haynes, W. (2013b). Tukey's Test. In O. and C. K.-H. and Y. H. Dubitzky Werner and Wolkenhauer (Ed.), *Encyclopedia of Systems Biology* (pp. 2303–2304). Springer New York. https://doi.org/10.1007/978-1-4419-9863-7_1212

He, L., Thomson, J. M., Hemann, M. T., Hernando-Monge, E., Mu, D., Goodson, S., Powers, S., Cordon-Cardo, C., Lowe, S. W., Hannon, G. J., & Hammond, S. M. (2005). A microRNA

polycistron as a potential human oncogene. *Nature*, *435*(7043), 828. https://doi.org/10.1038/NATURE03552

Higuchi, O., Hamuro, J., Motomura, M., & Yamanashi, Y. (2011). Autoantibodies to low-density lipoprotein receptor-related protein 4 in myasthenia gravis. *Annals of Neurology*, *69*(2), 418–422. https://doi.org/10.1002/ANA.22312

Hirota, J., Satomi, Y., Yoshikawa, K., & Takao, T. (2003). Epsilon -N,N,N-trimethyllysine-specific ions in matrix-assisted laser desorption/ionization-tandem mass spectrometry. *Rapid Communications in Mass Spectrometry : RCM*, *17*(5), 371–376. https://doi.org/10.1002/RCM.924

Hoch, W., Mcconville, J., Helms, S., Newsom-Davis, J., Melms, A., & Vincent, A. (2001). Auto-antibodies to the receptor tyrosine kinase MuSK in patients with myasthenia gravis without acetylcholine receptor antibodies. *Nature Medicine 2001 7:3*, *7*(3), 365–368. https://doi.org/10.1038/85520

Holčapek, M., Jirásko, R., & Lísa, M. (2012). Recent developments in liquid chromatography–mass spectrometry and related techniques. *Journal of Chromatography A*, *1259*, 3–15. https://doi.org/10.1016/J.CHROMA.2012.08.072

Hopfgartner, G., Varesio, E., Tschäppät, V., Grivet, C., Bourgogne, E., & Leuthold, L. A. (2004). Triple quadrupole linear ion trap mass spectrometer for the analysis of small molecules and macromolecules. *Journal of Mass Spectrometry : JMS*, *39*(8), 845–855. https://doi.org/10.1002/JMS.659

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*, *37*(1), 1–13. https://doi.org/10.1093/NAR/GKN923

Huddleston, M. J., Bean, M. F., & Carr, S. A. (1993). Collisional Fragmentation of Glycopeptides by Electrospray Ionization LC/MS and LC/MS/MS: Methods for Selective Detection of Glycopeptides in Protein Digests. *Analytical Chemistry*, *65*(7), 877–884. https://doi.org/10.1021/AC00055A009

Hunt, D. F., Yates, J. R., Shabanowitz, J., Winston, S., & Hauer, C. R. (1986). Protein sequencing by tandem mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America*, *83*(17), 6233–6237. https://doi.org/10.1073/PNAS.83.17.6233

Hu, Q., Noll, R. J., Li, H., Makarov, A., Hardman, M., & Cooks, R. G. (2005). The Orbitrap: a new mass spectrometer. *Journal of Mass Spectrometry : JMS*, *40*(4), 430–443. https://doi.org/10.1002/JMS.856

Hussain, T., Zhao, D., Shah, S. Z. A., Wang, J., Yue, R., Liao, Y., Sabir, N., Yang, L., & Zhou, X. (2018). MicroRNA 27a-3p Regulates Antimicrobial Responses of Murine Macrophages Infected by Mycobacterium avium subspecies paratuberculosis by Targeting Interleukin-10 and TGF-β-Activated Protein Kinase 1 Binding Protein 2. *Frontiers in Immunology*, *8*(JAN). https://doi.org/10.3389/FIMMU.2017.01915

Hu, X., Wang, Y., Liang, H., Fan, Q., Zhu, R., Cui, J., Zhang, W., Zen, K., Zhang, C. Y., Hou, D., Zhou, Z., & Chen, X. (2017). miR-23a/b promote tumor growth and suppress apoptosis by targeting PDCD4 in gastric cancer. *Cell Death & Disease*, *8*(10), e3059. https://doi.org/10.1038/CDDIS.2017.447

Introduction to mass analyzers : SHIMADZU (Shimadzu Corporation). (n.d.). Retrieved May 5, 2022, from https://www.shimadzu.com/an/service-support/technical-support/analysis-basics/fundamental/mass_analyzers.html

Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., & Mann, M. (2005). Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Molecular & Cellular Proteomics : MCP*, *4*(9), 1265–1272. https://doi.org/10.1074/MCP.M500061-MCP200

Jennings, K. R., & Dolnikowski, G. G. (1990). [2] Mass analyzers. *Methods in Enzymology*, *193*(C), 37–61. https://doi.org/10.1016/0076-6879(90)93410-M

Jordan, M., Schallhorn, A., & Wurm, F. M. (1996). Transfecting mammalian cells: optimization of critical parameters affecting calcium-phosphate precipitate formation. *Nucleic Acids Research*, *24*(4), 596–601. https://doi.org/10.1093/NAR/24.4.596

Kai, M., González, I., Genilloud, O., Singh, S. B., & Svatoš, A. (2012). Direct mass spectrometric screening of antibiotics from bacterial surfaces using liquid extraction surface analysis. *Rapid Communications in Mass Spectrometry*, *26*(20), 2477–2482. https://doi.org/10.1002/RCM.6365

Kaji, H., Saito, H., Yamauchi, Y., Shinkawa, T., Taoka, M., Hirabayashi, J., Kasai, K. ichi, Takahashi, N., & Isobe, T. (2003). Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nature Biotechnology*, *21*(6), 667–672. https://doi.org/10.1038/NBT829

Karas, M., Bachmann, D., & Hillenkamp, F. (1985). Influence of the Wavelength in High-Irradiance Ultraviolet Laser Desorption Mass Spectrometry of Organic Molecules. *Analytical*

Chemistry, *57*(14), 2935–2939. https://doi.org/10.1021/AC00291A042/ASSET/AC00291A042.FP.PNG_V03

Karpievitch, Y. v., Dabney, A. R., & Smith, R. D. (2012). Normalization and missing value imputation for label-free LC-MS analysis. *BMC Bioinformatics*, *13 Suppl 16*(16), 1–9. https://doi.org/10.1186/1471-2105-13-S16-S5/FIGURES/5

Kattula, S., Byrnes, J. R., & Wolberg, A. S. (2017). Fibrinogen and Fibrin in Hemostasis and Thrombosis. *Arteriosclerosis, Thrombosis, and Vascular Biology*, *37*(3), e13–e21. https://doi.org/10.1161/ATVBAHA.117.308564

Kelleher, N. L. (2004). Peer Reviewed: Top-Down Proteomics. *Analytical Chemistry*, *76*(11), 196 A-203 A. https://doi.org/10.1021/AC0415657

Keller, A., Eng, J., Zhang, N., Li, X. jun, & Aebersold, R. (2005). A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Molecular Systems Biology*, *1*. https://doi.org/10.1038/MSB4100024

Kelly, P. S., Breen, L., Gallagher, C., Kelly, S., Henry, M., Lao, N. T., Meleady, P., O'Gorman, D., Clynes, M., & Barron, N. (2015). Re-programming CHO cell metabolism using miR-23 tips the balance towards a highly productive phenotype. *Biotechnology Journal*, *10*(7), 1029–1040. https://doi.org/10.1002/BIOT.201500101

Kelstrup, C. D., Bekker-Jensen, D. B., Arrey, T. N., Hogrebe, A., Harder, A., & Olsen, J. v. (2018). Performance Evaluation of the Q Exactive HF-X for Shotgun Proteomics. *Journal of Proteome Research*, *17*(1), 727–738. https://doi.org/10.1021/ACS.JPROTEOME.7B00602

Kersten, R. D., & Dorrestein, P. C. (2009). Secondary metabolomics: Natural products mass spectrometry goes global. *ACS Chemical Biology*, *4*(8), 599–601. https://doi.org/10.1021/CB900187P/ASSET/IMAGES/LARGE/CB-2009-00187P_0001.JPEG

Keshishian, H., Addona, T., Burgess, M., Mani, D. R., Shi, X., Kuhn, E., Sabatine, M. S., Gerszten, R. E., & Carr, S. A. (2009). Quantification of cardiovascular biomarkers in patient plasma by targeted mass spectrometry and stable isotope dilution. *Molecular & Cellular Proteomics : MCP*, *8*(10), 2339–2349. https://doi.org/10.1074/MCP.M900140-MCP200

Khan, S. R., Baghdasarian, A., Nagar, P. H., Fahlman, R., Jurasz, P., Michail, K., Aljuhani, N., & Siraki, A. G. (2015). Proteomic profile of aminoglutethimide-induced apoptosis in HL-60 cells: Role of myeloperoxidase and arylamine free radicals. *Chemico-Biological Interactions*, *239*, 129–138. https://doi.org/10.1016/J.CBI.2015.06.020

Kingdon, K. H. (1923). A Method for the Neutralization of Electron Space Charge by Positive Ionization at Very Low Gas Pressures. *Physical Review*, *21*(4), 408. https://doi.org/10.1103/PhysRev.21.408

Kirby, T. J., Lee, J. D., England, J. H., Chaillou, T., Esser, K. A., & McCarthy, J. J. (2015). Blunted hypertrophic response in aged skeletal muscle is associated with decreased ribosome biogenesis. *Journal of Applied Physiology*, *119*(4), 321. https://doi.org/10.1152/JAPPLPHYSIOL.00296.2015

Kleifeld, O., Doucet, A., Prudova, A., Auf Dem Keller, U., Gioia, M., Kizhakkedathu, J. N., & Overall, C. M. (2011). Identifying and quantifying proteolytic events and the natural N

terminome by terminal amine isotopic labeling of substrates. *Nature Protocols*, *6*(10), 1578–1611. https://doi.org/10.1038/NPROT.2011.382

Klose, J. (1975). Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik*, *26*(3), 231–243. https://doi.org/10.1007/BF00281458

Koehler, C. J., Arntzen, M., Strozynski, M., Treumann, A., & Thiede, B. (2011). Isobaric peptide termini labeling utilizing site-specific N-terminal succinylation. *Analytical Chemistry*, *83*(12), 4775–4781. https://doi.org/10.1021/AC200229W/SUPPL_FILE/AC200229W_SI_002.PDF

Kong, K. Y., Owens, K. S., Rogers, J. H., Mullenix, J., Velu, C. S., Grimes, H. L., & Dahl, R. (2010). THE MIR-23A MICRORNA CLUSTER INHIBITS B CELL DEVELOPMENT. *Experimental Hematology*, *38*(8), 629. https://doi.org/10.1016/J.EXPHEM.2010.04.004

Koopmans, F., Ho, J. T. C., Smit, A. B., & Li, K. W. (2018). Comparative Analyses of Data Independent Acquisition Mass Spectrometric Approaches: DIA, WiSIM-DIA, and Untargeted DIA. *Proteomics*, *18*(1). https://doi.org/10.1002/PMIC.201700304

Kramer, D. A., Eldeeb, M. A., Wuest, M., Mercer, J., & Fahlman, R. P. (2017). Proteomic characterization of EL4 lymphoma-derived tumors upon chemotherapy treatment reveals potential roles for lysosomes and caspase-6 during tumor cell death in vivo. *PROTEOMICS*, *17*(12), 1700060. https://doi.org/10.1002/PMIC.201700060

Kramer, D. A., Quiroga, A. D., Lian, J., Fahlman, R. P., & Lehner, R. (2018). Fasting and refeeding induces changes in the mouse hepatic lipid droplet proteome. *Journal of Proteomics*, *181*, 213–224. https://doi.org/10.1016/J.JPROT.2018.04.024

Kukulka, N., Bailey, E., Yelam, A., Nagarajan, E., & Govindarajan, R. (2019). Myasthenia Gravis Mimicking Stroke - Implications of Cognitive Errors and Clinical Factors (P5.4-032). *Neurology*, *92*(15 Supplement).

Kurkewich, J. L., Hansen, J., Klopfenstein, N., Zhang, H., Wood, C., Boucher, A., Hickman, J., Muench, D. E., Grimes, H. L., & Dahl, R. (2017). The miR-23a~27a~24-2 microRNA cluster buffers transcription and signaling pathways during hematopoiesis. *PLoS Genetics*, *13*(7). https://doi.org/10.1371/JOURNAL.PGEN.1006887

Kusmierz, J. J., Sumrada, R., & Desiderio, D. M. (1990). Fast atom bombardment mass spectrometric quantitative analysis of methionine-enkephalin in human pituitary tissues. *Analytical Chemistry*, *62*(21), 2395–2400. https://doi.org/10.1021/AC00220A026

Latosinska, A., Vougas, K., Makridakis, M., Klein, J., Mullen, W., Abbas, M., Stravodimos, K., Katafigiotis, I., Merseburger, A. S., Zoidakis, J., Mischak, H., Vlahou, A., & Jankowski, V. (2015). Comparative Analysis of Label-Free and 8-Plex iTRAQ Approach for Quantitative Tissue Proteomic Analysis. *PloS One*, *10*(9). https://doi.org/10.1371/JOURNAL.PONE.0137048

Lazar, C., Gatto, L., Ferro, M., Bruley, C., & Burger, T. (2016). Accounting for the Multiple Natures of Missing Values in Label-Free Quantitative Proteomics Data Sets to Compare Imputation Strategies. *Journal of Proteome Research*, *15*(4), 1116–1125. https://doi.org/10.1021/ACS.JPROTEOME.5B00981

Lee, R. C., Feinbaum, R. L., & Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, *75*(5), 843–854. https://doi.org/10.1016/0092-8674(93)90529-Y

Lee, T., Wang, N., Houel, S., Couts, K., Old, W., & Ahn, N. (2015). Dosage and temporal thresholds in microRNA proteomics. *Molecular & Cellular Proteomics : MCP*, *14*(2), 289–302. https://doi.org/10.1074/MCP.M114.043851

Lefèvre, F., & Gillery, P. (1997). Residual presence of fibrinogen: a common pitfall in the interpretation of serum protein electrophoresis. *Annales de Biologie Clinique*, *55*(3), 238–240.

Leinonen, R., Garcia Diez, F., Binns, D., Fleischmann, W., Lopez, R., & Apweiler, R. (2004). UniProt archive. *Bioinformatics (Oxford, England)*, *20*(17), 3236–3237. https://doi.org/10.1093/BIOINFORMATICS/BTH191

Liebler, D. C., & Zimmerman, L. J. (2013). Targeted quantitation of proteins by mass spectrometry. *Biochemistry*, *52*(22), 3797–3806. https://doi.org/10.1021/BI400110B

Li, J., Zhao, Y., Lu, Y., Ritchie, W., Grau, G., Vadas, M. A., & Gamble, J. R. (2016). The Poly-cistronic miR-23-27-24 Complexes Target Endothelial Cell Junctions: Differential Functional and Molecular Effects of miR-23a and miR-23b. *Molecular Therapy. Nucleic Acids*, *5*(8), e354. https://doi.org/10.1038/MTNA.2016.62

Lindqvist, L. M., Tandoc, K., Topisirovic, I., & Furic, L. (2018). Cross-talk between protein synthesis, energy metabolism and autophagy in cancer. *Current Opinion in Genetics & Development*, *48*, 104–111. https://doi.org/10.1016/J.GDE.2017.11.003

Listgarten, J., & Emili, A. (2005). Statistical and computational methods for comparative proteomic profiling using liquid chromatography-tandem mass spectrometry. *Molecular & Cellular Proteomics : MCP*, *4*(4), 419–434. https://doi.org/10.1074/MCP.R500005-MCP200

Liu, Y., Hüttenhain, R., Surinova, S., Gillet, L. C. J., Mouritsen, J., Brunner, R., Navarro, P., & Aebersold, R. (2013). Quantitative measurements of N-linked glycoproteins in human plasma by SWATH-MS. *Proteomics*, *13*(8), 1247–1256. https://doi.org/10.1002/PMIC.201200417

Lopes, A. M., Miguel, R. N., Sargent, C. A., Ellis, P. J., Amorim, A., & Affara, N. A. (2010). The human RPS4 paralogue on Yq11.223 encodes a structurally conserved ribosomal protein and is preferentially expressed during spermatogenesis. *BMC Molecular Biology*, *11*(1), 1–12. https://doi.org/10.1186/1471-2199-11-33/FIGURES/4

Ludwig, K. R., Dahl, R., & Hummon, A. B. (2016). Evaluation of the mirn23a Cluster through an iTRAQ-based Quantitative Proteomic Approach. *Journal of Proteome Research*, *15*(5), 1497. https://doi.org/10.1021/ACS.JPROTEOME.5B01101

Ly, T., Ahmad, Y., Shlien, A., Soroka, D., Mills, A., Emanuele, M. J., Stratton, M. R., & Lamond, A. I. (2014). A proteomic chronology of gene expression through the cell cycle in human myeloid leukemia cells. *ELife*, *2014*(3). https://doi.org/10.7554/ELIFE.01630

MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., Kern, R., Tabb, D. L., Liebler, D. C., & MacCoss, M. J. (2010). Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics*, *26*(7), 966–968. https://doi.org/10.1093/BIOINFORMATICS/BTQ054

Makarov, A. (2000). Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis. *Analytical Chemistry*, *72*(6), 1156–1162. https://doi.org/10.1021/AC991131P/ASSET/IMAGES/LARGE/AC991131PF00011.JPEG

Mallick, P., Schirle, M., Chen, S. S., Flory, M. R., Lee, H., Martin, D., Ranish, J., Raught, B., Schmitt, R., Werner, T., Kuster, B., & Aebersold, R. (2006). Computational prediction of

proteotypic peptides for quantitative proteomics. *Nature Biotechnology 2006 25:1*, *25*(1), 125–131. https://doi.org/10.1038/nbt1275

Mallick, P., Schirle, M., Chen, S. S., Flory, M. R., Lee, H., Martin, D., Ranish, J., Raught, B., Schmitt, R., Werner, T., Kuster, B., & Aebersold, R. (2007). Computational prediction of proteotypic peptides for quantitative proteomics. *Nature Biotechnology*, *25*(1), 125–131. https://doi.org/10.1038/NBT1275

Malygin, A. A., Babaylova, E. S., Loktev, V. B., & Karpova, G. G. (2011). A region in the C-terminal domain of ribosomal protein SA required for binding of SA to the human 40S ribosomal subunit. *Biochimie*, *93*(3), 612–617. https://doi.org/10.1016/J.BIOCHI.2010.12.005

Mann, M., Hendrickson, R. C., & Pandey, A. (2001). Analysis of proteins and proteomes by mass spectrometry. *Annual Review of Biochemistry*, *70*, 437–473. https://doi.org/10.1146/ANNUREV.BIOCHEM.70.1.437

Mao, Z. F., Yang, L. X., Mo, X. A., Qin, C., Lai, Y. R., He, N. Y., Li, T., & Hackett, M. L. (2011). Frequency of autoimmune diseases in myasthenia gravis: a systematic review. *The International Journal of Neuroscience*, *121*(3), 121–129. https://doi.org/10.3109/00207454.2010.539307

March, R. E. (1997). An Introduction to Quadrupole Ion Trap Mass Spectrometry. *JOURNAL OF MASS SPECTROMETRY*, *32*, 351–369. https://doi.org/10.1002/(SICI)1096-9888(199704)32:4

Martin, I., Kim, J. W., Dawson, V. L., & Dawson, T. M. (2014). LRRK2 pathobiology in Parkinson's disease. *Journal of Neurochemistry*, *131*(5), 554–565. https://doi.org/10.1111/JNC.12949

Marx, V. (2015). Mapping proteins with spatial proteomics. *Nature Methods 2015 12:9*, *12*(9), 815–819. https://doi.org/10.1038/nmeth.3555

Mass Spectrometry Applications Areas | Thermo Fisher Scientific - CA. (n.d.). Retrieved May 4, 2022, from https://www.thermofisher.com/ca/en/home/industrial/mass-spectrometry/mass-spectrometry-learning-center/mass-spectrometry-applications-area.html

McHugh, M. L. (2011). Multiple comparison analysis testing in ANOVA. *Biochemia Medica*, *21*(3), 203–209. https://doi.org/10.11613/BM.2011.029

McRae, E. K. S., Dupas, S. J., Booy, E. P., Piragasam, R. S., Fahlman, R. P., & McKenna, S. A. (2020). An RNA guanine quadruplex regulated pathway to TRAIL-sensitization by DDX21. *RNA (New York, N.Y.)*, *26*(1), 44–57. https://doi.org/10.1261/rna.072199.119

Meyuhas, O. (2015). Ribosomal Protein S6 Phosphorylation: Four Decades of Research. *International Review of Cell and Molecular Biology*, *320*, 41–73. https://doi.org/10.1016/BS.IRCMB.2015.07.006

Miller, P. E., & Denton, M. B. (1986). The quadrupole mass filter: Basic operating concepts. *Journal of Chemical Education*, *63*(7), 617–622. https://doi.org/10.1021/ED063P617

Minamoto, T., Mai, M., & Ronai, Z. (1999). Environmental factors as regulators and effectors of multistep carcinogenesis. *Carcinogenesis*, *20*(4), 519–527. https://doi.org/10.1093/CARCIN/20.4.519

Mirone, L., Altomonte, L., D'Agostino, P., Zoli, A., Barini, A., & Magaro', M. (1996). A study of serum androgen and cortisol levels in female patients with rheumatoid arthritis. Correlation with disease activity. *Clinical Rheumatology*, *15*(1), 15–19. https://doi.org/10.1007/BF02231678

Molin, C. J., Westerberg, E., & Punga, A. R. (2017). Profile of upregulated inflammatory proteins in sera of Myasthenia Gravis patients. *Scientific Reports*, *7*. https://doi.org/10.1038/SREP39716

Muntel, J., Boswell, S. A., Tang, S., Ahmed, S., Wapinski, I., Foley, G., Steen, H., & Springer, M. (2015). Abundance-based classifier for the prediction of mass spectrometric peptide detectability upon enrichment (PPA). *Molecular & Cellular Proteomics : MCP*, *14*(2), 430–440. https://doi.org/10.1074/MCP.M114.044321

Murray, K. K., Boyd, R. K., Eberlin, M. N., Langley, G. J., Li, L., & Naito, Y. (2013). Definitions of terms relating to mass spectrometry (IUPAC Recommendations 2013)*. *Pure Appl. Chem*, *85*(7), 1515–1609. https://doi.org/10.1351/PAC-REC-06-04-06

Nacu, A., Andersen, J. B., Lisnic, V., Owe, J. F., & Gilhus, N. E. (2015). Complicating autoimmune diseases in myasthenia gravis: a review. *Autoimmunity*, *48*(6), 362–368. https://doi.org/10.3109/08916934.2015.1030614

Nahnsen, S., Bielow, C., Reinert, K., & Kohlbacher, O. (2013). Tools for label-free peptide quantification. *Molecular & Cellular Proteomics : MCP*, *12*(3), 549–556. https://doi.org/10.1074/MCP.R112.025163

Neubauer, G., & Mann, M. (1999). Mapping of phosphorylation sites of gel-isolated proteins by nanoelectrospray tandem mass spectrometry: potentials and limitations. *Analytical Chemistry*, *71*(1), 235–242. https://doi.org/10.1021/AC9804902

Niall, H. D. (1973). Automated Edman degradation: the protein sequenator. *Methods in Enzymology*, *27*(C), 942–1010. https://doi.org/10.1016/S0076-6879(73)27039-8

Oda, Y., Huang, K., Cross, F. R., Cowburn, D., & Chait, B. T. (1999). Accurate quantitation of protein expression and site-specific phosphorylation. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(12), 6591–6596. https://doi.org/10.1073/PNAS.96.12.6591/ASSET/89E6B108-00E4-4FEB-8E01-CAB678C7CCE4/ASSETS/GRAPHIC/PQ1290061006.JPEG

Old, W. M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K. G., Mendoza, A., Sevinsky, J. R., Resing, K. A., & Ahn, N. G. (2005). Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Molecular & Cellular Proteomics : MCP*, *4*(10), 1487–1502. https://doi.org/10.1074/MCP.M500084-MCP200

O'Leary, M. N., Schreiber, K. H., Zhang, Y., Duc, A. C. E., Rao, S., Hale, J. S., Academia, E. C., Shah, S. R., Morton, J. F., Holstein, C. A., Martin, D. B., Kaeberlein, M., Ladiges, W. C., Fink, P. J., MacKay, V. L., Wiest, D. L., & Kennedy, B. K. (2013). The ribosomal protein Rpl22 controls ribosome composition by directly repressing expression of its own paralog, Rpl22l1. *PLoS Genetics*, *9*(8). https://doi.org/10.1371/JOURNAL.PGEN.1003708

Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., & Mann, M. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate

approach to expression proteomics. *Molecular & Cellular Proteomics : MCP*, *1*(5), 376–386. https://doi.org/10.1074/MCP.M200025-MCP200

Parks, M. M., Kurylo, C. M., Dass, R. A., Bojmar, L., Lyden, D., Vincent, C. T., & Blanchard, S. C. (2018). Variant ribosomal RNA alleles are conserved and exhibit tissue-specific expression. *Science Advances*, *4*(2). https://doi.org/10.1126/SCIADV.AAO0665/SUPPL_FILE/AAO0665_TABLES8.XLSX

Paša-Tolic, L., Jensen, P. K., Anderson, G. A., Lipton, M. S., Peden, K. K., Martinović, S., Tolić, N., Bruce, J. E., & Smith, R. D. (1999). High throughput proteome-wide precision measurements of protein expression using mass spectrometry. *Journal of the American Chemical Society*, *121*(34), 7949–7950. https://doi.org/10.1021/JA991063O

Peeler, C. E., de Lott, L. B., Nagia, L., Lemos, J., Eggenberger, E. R., & Cornblath, W. T. (2015). Clinical Utility of Acetylcholine Receptor Antibody Testing in Ocular Myasthenia Gravis. *JAMA Neurology*, *72*(10), 1170–1174. https://doi.org/10.1001/JAMANEUROL.2015.1444

Peng, J., Schwartz, D., Elias, J. E., Thoreen, C. C., Cheng, D., Marsischky, G., Roelofs, J., Finley, D., & Gygi, S. P. (2003). A proteomics approach to understanding protein ubiquitination. *Nature Biotechnology 2003 21:8*, *21*(8), 921–926. https://doi.org/10.1038/nbt849

Perry, R. H., Cooks, R. G., & Noll, R. J. (2008). Orbitrap mass spectrometry: instrumentation, ion motion and applications. *Mass Spectrometry Reviews*, *27*(6), 661–699. https://doi.org/10.1002/MAS.20186

Peterson, A. C., Russell, J. D., Bailey, D. J., Westphall, M. S., & Coon, J. J. (2012). Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted

proteomics. *Molecular & Cellular Proteomics : MCP*, *11*(11), 1475–1488. https://doi.org/10.1074/MCP.O112.020131

Pino, L. K., Just, S. C., MacCoss, M. J., & Searle, B. C. (2020). Acquiring and Analyzing Data Independent Acquisition Proteomics Experiments without Spectrum Libraries. *Molecular & Cellular Proteomics : MCP*, *19*(7), 1088–1103. https://doi.org/10.1074/MCP.P119.001913

Pinzón, N., Li, B., Martinez, L., Sergeeva, A., Presumey, J., Apparailly, F., & Seitz, H. (2017). MicroRNA target prediction programs predict many false positives. *Genome Research*, *27*(2), 234–245. https://doi.org/10.1101/GR.205146.116/-/DC1

Piragasam, R. S., Hussain, S. F., Chaulk, S. G., Siddiqi, Z. A., & Fahlman, R. P. (2020). Label-free proteomic analysis reveals large dynamic changes to the cellular proteome upon expression of the miRNA-23a-27a-24-2 microRNA cluster. *Biochemistry and Cell Biology*, *98*(1), 61–69. https://doi.org/10.1139/BCB-2019-0014/SUPPL_FILE/BCB-2019-0014SUPPLA.XLSX

Pitt, J. J. (2009). Principles and Applications of Liquid Chromatography-Mass Spectrometry in Clinical Biochemistry. *The Clinical Biochemist Reviews*, *30*(1), 19. /pmc/articles/PMC2643089/

Polakis, E. S., & Bartley, W. (1966). Changes in dry weight, protein, deoxyribonucleic acid, ribonucleic acid and reserve and structural carbohydrate during the aerobic growth cycle of yeast. *The Biochemical Journal*, *98*(3), 883–887. https://doi.org/10.1042/BJ0980883

Pounds, S., & Morris, S. W. (2003). Estimating the occurrence of false positives and false negatives in microarray studies by approximating and partitioning the empirical distribution

of p-values. *Bioinformatics*, *19*(10), 1236–1242. https://doi.org/10.1093/BIOINFORMATICS/BTG148

Prasad, T. S. K., Kandasamy, K., & Pandey, A. (2009). Human Protein Reference Database and Human Proteinpedia as discovery tools for systems biology. *Methods in Molecular Biology (Clifton, N.J.)*, *577*, 67–79. https://doi.org/10.1007/978-1-60761-232-2_6

Qiu, R., & Regnier, F. E. (2005). Comparative Glycoproteomics of N-Linked Complex-Type Glycoforms Containing Sialic Acid in Human Serum. *Analytical Chemistry*, *77*(22), 7225–7231. https://doi.org/10.1021/AC050554Q

Qu, Z., Huang, J., Yang, F., Hong, J., Wang, W., & Yan, S. (2020). Sex hormone-binding globulin and arthritis: A Mendelian randomization study. *Arthritis Research and Therapy*, *22*(1), 1–7. https://doi.org/10.1186/S13075-020-02202-2/FIGURES/3

Ranjbar, M. R. N., Tadesse, M. G., Wang, Y., & Ressom, H. W. (2015). Bayesian Normalization Model for Label-Free Quantitative Analysis by LC-MS. *IEEE/ACM Transactions on Computational Biology and Bioinformatics / IEEE, ACM*, *12*(4), 914. https://doi.org/10.1109/TCBB.2014.2377723

Rappsilber, J., Ryder, U., Lamond, A. I., & Mann, M. (2002). Large-scale proteomic analysis of the human spliceosome. *Genome Research*, *12*(8), 1231–1245. https://doi.org/10.1101/GR.473902

resolution. (2014). *The IUPAC Compendium of Chemical Terminology*. https://doi.org/10.1351/GOLDBOOK.R05318

resolving power in mass spectrometry. (2008). *The IUPAC Compendium of Chemical Terminology*. https://doi.org/10.1351/GOLDBOOK.R05321

Rigbolt, K. T. G., Prokhorova, T. A., Akimov, V., Henningsen, J., Johansen, P. T., Kratchmarova, I., Kassem, M., Mann, M., Olsen, J. v., & Blagoev, B. (2011). System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation. *Science Signaling*, *4*(164). https://doi.org/10.1126/SCISIGNAL.2001570

Rivera, M. C., Maguire, B., & Lake, J. A. (2015). Isolation of Ribosomes and Polysomes. *Cold Spring Harbor Protocols*, *2015*(3), pdb.prot081331. https://doi.org/10.1101/PDB.PROT081331

Rivers, J., Simpson, D. M., Robertson, D. H. L., Gaskell, S. J., & Beynon, R. J. (2007). Absolute multiplexed quantitative analysis of protein expression during muscle development using QconCAT. *Molecular & Cellular Proteomics : MCP*, *6*(8), 1416–1427. https://doi.org/10.1074/MCP.M600456-MCP200

Roepstorff, P., & Fohlman, J. (1984). Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomedical Mass Spectrometry*, *11*(11), 601–601. https://doi.org/10.1002/BMS.1200111109

Rogers, J. C., & Bomgarden, R. D. (2016). Sample preparation for mass spectrometry-based proteomics; from proteomes to peptides. *Advances in Experimental Medicine and Biology*, *919*, 43–62. https://doi.org/10.1007/978-3-319-41448-5_3/FIGURES/6

Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., & Pappin, D. J. (2004). Multiplexed protein quantitation in Saccharomyces

cerevisiae using amine-reactive isobaric tagging reagents. *Molecular & Cellular Proteomics : MCP*, *3*(12), 1154–1169. https://doi.org/10.1074/MCP.M400129-MCP200

Röst, H. L., Rosenberger, G., Navarro, P., Gillet, L., Miladinoviä, S. M., Schubert, O. T., Wolski, W., Collins, B. C., Malmström, J., Malmström, L., & Aebersold, R. (2014). OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nature Biotechnology 2014 32:3*, *32*(3), 219–223. https://doi.org/10.1038/nbt.2841

Roundtree, I. A., Evans, M. E., Pan, T., & He, C. (2017). Dynamic RNA Modifications in Gene Expression Regulation. *Cell*, *169*(7), 1187–1200. https://doi.org/10.1016/J.CELL.2017.05.045

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592. https://doi.org/10.1093/BIOMET/63.3.581

Saw, Y. O., Salim, M., Noirel, J., Evans, C., Rehman, I., & Wright, P. C. (2009). iTRAQ underestimation in simple and complex mixtures: "The good, the bad and the ugly." *Journal of Proteome Research*, *8*(11), 5347–5355. https://doi.org/10.1021/PR900634C/SUPPL_FILE/PR900634C_SI_002.XLS

Scheiman, J., Jamieson, K. v., Ziello, J., Tseng, J. C., & Meruelo, D. (2010). Extraribosomal functions associated with the C terminus of the 37/67 kDa laminin receptor are required for maintaining cell viability. *Cell Death & Disease*, *1*(5). https://doi.org/10.1038/CDDIS.2010.19

Schilling, B., MacLean, B., Held, J. M., Sahu, A. K., Rardin, M. J., Sorensen, D. J., Peters, T., Wolfe, A. J., Hunter, C. L., MacCoss, M. J., & Gibson, B. W. (2015). Multiplexed, Scheduled, High-Resolution Parallel Reaction Monitoring on a Full Scan QqTOF Instrument with

Integrated Data-Dependent and Targeted Mass Spectrometric Workflows. *Analytical Chemistry*, *87*(20), 10222–10229. https://doi.org/10.1021/ACS.ANALCHEM.5B02983

Schmidlin, T., Garrigues, L., Lane, C. S., Mulder, T. C., van Doorn, S., Post, H., de Graaf, E. L., Lemeer, S., Heck, A. J. R., & Altelaar, A. F. M. (2016). Assessment of SRM, MRM(3) , and DIA for the targeted analysis of phosphorylation dynamics in non-small cell lung cancer. *Proteomics*, *16*(15–16), 2193–2205. https://doi.org/10.1002/PMIC.201500453

Schmidt, A., Kellermann, J., & Lottspeich, F. (2005). A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics*, *5*(1), 4–15. https://doi.org/10.1002/PMIC.200400873

Schoolmeesters, A., Eklund, T., Leake, D., Vermeulen, A., Smith, Q., Aldred, S. F., & Fedorov, Y. (2009). Functional profiling reveals critical role for miRNA in differentiation of human mesenchymal stem cells. *PloS One*, *4*(5). https://doi.org/10.1371/JOURNAL.PONE.0005605

Schwanhüusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., & Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature 2011 473:7347*, *473*(7347), 337–342. https://doi.org/10.1038/nature10098

Searle, B. C., Egertson, J. D., Bollinger, J. G., Stergachis, A. B., & MacCoss, M. J. (2015). Using data independent acquisition (DIA) to model high-responding peptides for targeted proteomics experiments. *Molecular and Cellular Proteomics*, *14*(9), 2331–2340. https://doi.org/10.1074/MCP.M115.051300/ATTACHMENT/A682A283-3314-4C91-A0FA-DD50644F55BB/MMC1.ZIP

Searle, B. C., Pino, L. K., Egertson, J. D., Ting, Y. S., Lawrence, R. T., MacLean, B. X., Villén, J., & MacCoss, M. J. (2018). Chromatogram libraries improve peptide detection and

quantification by data independent acquisition mass spectrometry. *Nature Communications 2018 9:1*, *9*(1), 1–12. https://doi.org/10.1038/s41467-018-07454-w

Sechi, S., & Chait, B. T. (1998). Modification of cysteine residues by alkylation. A tool in peptide mapping and protein identification. *Analytical Chemistry*, *70*(24), 5150–5158. https://doi.org/10.1021/AC9806005

Seitz, H. (2017). Issues in current microRNA target identification methods. *RNA Biology*, *14*(7), 831. https://doi.org/10.1080/15476286.2017.1320469

Shevchenko, A., Tomas, H., Havliš, J., Olsen, J. v., & Mann, M. (2006). In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature Protocols*, *1*(6), 2856–2860. https://doi.org/10.1038/NPROT.2006.468

Shi, Z., Fujii, K., Kovary, K. M., Genuth, N. R., Röst, H. L., Teruel, M. N., & Barna, M. (2017). Heterogeneous ribosomes preferentially translate distinct subpools of mRNAs genome-wide. *Molecular Cell*, *67*(1), 71. https://doi.org/10.1016/J.MOLCEL.2017.05.021

Silva, J. C., Gorenstein, M. v., Li, G. Z., Vissers, J. P. C., & Geromanos, S. J. (2006). Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Molecular & Cellular Proteomics : MCP*, *5*(1), 144–156. https://doi.org/10.1074/MCP.M500230-MCP200

Simsek, D., & Barna, M. (2017). An emerging role for the ribosome as a nexus for post-translational modifications. *Current Opinion in Cell Biology*, *45*, 92–101. https://doi.org/10.1016/J.CEB.2017.02.010

Simsek, D., Tiu, G. C., Flynn, R. A., Byeon, G. W., Leppek, K., Xu, A. F., Chang, H. Y., & Barna, M. (2017). The Mammalian Ribo-interactome Reveals Ribosome Functional Diversity and Heterogeneity. *Cell*, *169*(6), 1051-1065.e18. https://doi.org/10.1016/J.CELL.2017.05.022

Sleno, L., & Volmer, D. A. (2004). Ion activation methods for tandem mass spectrometry. *Journal of Mass Spectrometry : JMS*, *39*(10), 1091–1112. https://doi.org/10.1002/JMS.703

Smith, L. M., & Kelleher, N. L. (2013). Proteoform: a single term describing protein complexity. *Nature Methods 2013 10:3*, *10*(3), 186–187. https://doi.org/10.1038/nmeth.2369

Sonenberg, N., & Hinnebusch, A. G. (2009). Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell*, *136*(4), 731–745. https://doi.org/10.1016/J.CELL.2009.01.042

Spillane, J., Higham, E., & Kullmann, D. M. (2012). Myasthenia gravis. *BMJ*, *345*(7891). https://doi.org/10.1136/BMJ.E8497

Srivastava, N., Manvati, S., Srivastava, A., Pal, R., Kalaiarasan, P., Chattopadhyay, S., Gochhait, S., Dua, R., & Bamezai, R. N. K. (2011). miR-24-2 controls H2AFX expression regardless of gene copy number alteration and induces apoptosis by targeting antiapoptotic gene BCL-2: a potential for therapeutic intervention. *Breast Cancer Research : BCR*, *13*(2). https://doi.org/10.1186/BCR2861

Steinmann, D., & Ganzera, M. (2011). Recent advances on HPLC/MS in medicinal plant analysis. *Journal of Pharmaceutical and Biomedical Analysis*, *55*(4), 744–757. https://doi.org/10.1016/J.JPBA.2010.11.015

Storey, J. D., & Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(16), 9440–9445. https://doi.org/10.1073/PNAS.1530509100/ASSET/25537429-365C-4D06-977A-86C871368513/ASSETS/GRAPHIC/PQ1530509003.JPEG

Student's t-Distribution -- from Wolfram MathWorld. (n.d.). Retrieved May 7, 2022, from https://mathworld.wolfram.com/Studentst-Distribution.html

Stults, D. M., Killen, M. W., Pierce, H. H., & Pierce, A. J. (2008). Genomic architecture and inheritance of human ribosomal RNA gene clusters. *Genome Research*, *18*(1), 13–18. https://doi.org/10.1101/GR.6858507

Suna, G., & Mayr, M. (2018). Proteomics. In *Encyclopedia of Cardiovascular Research and Medicine* (pp. 166–180). Elsevier. https://doi.org/10.1016/B978-0-12-809657-4.99573-5

Sun, R., Hunter, C., Chen, C., Ge, W., Morrice, N., Liang, S., Zhu, T., Yuan, C., Ruan, G., Zhang, Q., Cai, X., Yu, X., Chen, L., Dai, S., Luan, Z., Aebersold, R., Zhu, Y., & Guo, T. (2020). Accelerated Protein Biomarker Discovery from FFPE Tissue Samples Using Single-Shot, Short Gradient Microflow SWATH MS. *Journal of Proteome Research*, *19*(7), 2732–2741. https://doi.org/10.1021/ACS.JPROTEOME.9B00671

Tanaka, N., Yamaguchi, A., Hashizume, K., Araki, M., Wada, A., & Kimata, K. (1986). Separation of isotopic compounds by reversed-phase liquid chromatography. Effect of pressure gradient on isotope separation by ionization control. *Journal of High Resolution Chromatography*, *9*(11), 683–687. https://doi.org/10.1002/JHRC.1240091119

Thadikkaran, L., Siegenthaler, M. A., Crettaz, D., Queloz, P. A., Schneider, P., & Tissot, J. D. (2005). Recent advances in blood-related proteomics. *Proteomics*, *5*(12), 3019–3034. https://doi.org/10.1002/PMIC.200402053

Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., & Hamon, C. (2003). Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Analytical Chemistry*, *75*(8), 1895–1904. https://doi.org/10.1021/AC0262560

Thorlacius, S., Aarli, J. A., Riise, T., Matre, R., & Johnsen, H. J. (1989). Associated disorders in myasthenia gravis: autoimmune diseases and their relation to thymectomy. *Acta Neurologica Scandinavica*, *80*(4), 290–295. https://doi.org/10.1111/J.1600-0404.1989.TB03881.X

Toby, T. K., Fornelli, L., & Kelleher, N. L. (2016). Progress in Top-Down Proteomics and the Analysis of Proteoforms. *Annual Review of Analytical Chemistry (Palo Alto, Calif.)*, *9*(1), 499–519. https://doi.org/10.1146/ANNUREV-ANCHEM-071015-041550

Todd, J. F. J. (1991). Recommendations for nomenclature and symbolism for mass spectroscopy (including an appendix of terms used in vacuum technology). *Pure and Applied Chemistry*, *63*(10), 1541–1566. https://doi.org/10.1351/PAC199163101541/MACHINEREADABLECITATION/RIS

Trinh, H. v., Grossmann, J., Gehrig, P., Roschitzki, B., Schlapbach, R., Greber, U. F., & Hemmi, S. (2013). iTRAQ-Based and Label-Free Proteomics Approaches for Studies of Human Adenovirus Infections. *International Journal of Proteomics*, *2013*, 1–16. https://doi.org/10.1155/2013/581862

Tukey, J. (1977). *Exploratory data analysis*. http://theta.edu.pl/wp-content/uploads/2012/10/exploratorydataanalysis_tukey.pdf

Uzawa, A., Kanai, T., Kawaguchi, N., Oda, F., Himuro, K., & Kuwabara, S. (2016). Changes in inflammatory cytokine networks in myasthenia gravis. *Scientific Reports*, *6*. https://doi.org/10.1038/SREP25886

Välikangas, T., Suomi, T., & Elo, L. L. (2018). A systematic evaluation of normalization methods in quantitative label-free proteomics. *Briefings in Bioinformatics*, *19*(1), 1–11. https://doi.org/10.1093/BIB/BBW095

Venable, J. D., Dong, M. Q., Wohlschlegel, J., Dillin, A., & Yates, J. R. (2004). Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nature Methods*, *1*(1), 39–45. https://doi.org/10.1038/NMETH705

Vincent, A., & Newsom-Davis, J. (1985). Acetylcholine receptor antibody as a diagnostic test for myasthenia gravis: results in 153 validated cases and 2967 diagnostic assays. *Journal of Neurology, Neurosurgery, and Psychiatry*, *48*(12), 1246. https://doi.org/10.1136/JNNP.48.12.1246

Vizcaíno, J. A., Deutsch, E. W., Wang, R., Csordas, A., Reisinger, F., Ríos, D., Dianes, J. A., Sun, Z., Farrah, T., Bandeira, N., Binz, P. A., Xenarios, I., Eisenacher, M., Mayer, G., Gatto, L., Campos, A., Chalkley, R. J., Kraus, H. J., Albar, J. P., … Hermjakob, H. (2014). ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nature Biotechnology*, *32*(3), 223–226. https://doi.org/10.1038/NBT.2839

Vogel, C., & Marcotte, E. M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews. Genetics*, *13*(4), 227. https://doi.org/10.1038/NRG3185

Wang, H., Alvarez, S., & Hicks, L. M. (2012). Comprehensive comparison of iTRAQ and label-free LC-based quantitative proteomics approaches using two chlamydomonas reinhardtii strains of interest for biofuels engineering. *Journal of Proteome Research*, *11*(1), 487–501. https://doi.org/10.1021/PR2008225/SUPPL_FILE/PR2008225_SI_006.XLS

Wang, W., Becker, C. H., Zhou, H., Lin, H., Roy, S., Shaler, T. A., Hill, L. R., Norton, S., Kumar, P., & Anderle, M. (2003). Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. *Analytical Chemistry*, *75*(18), 4818–4826. https://doi.org/10.1021/AC026468X

Wang, Y., Liang, H., Zhou, G., Hu, X., Liu, Z., Jin, F., Yu, M., Sang, J., Zhou, Y., Fu, Z., Zhang, C. Y., Zhang, W., Zen, K., & Chen, X. (2017). HIC1 and miR-23~27~24 clusters form a double-negative feedback loop in breast cancer. *Cell Death and Differentiation*, *24*(3), 421–432. https://doi.org/10.1038/CDD.2016.136

Wasinger, V. C., Cordwell, S. J., Cerpa-Poljak, A., Yan, J. X., Gooley, A. A., Wilkins, M. R., Duncan, M. W., Harris, R., Williams, K. L., & Humphery-Smith, I. (1995). Progress with gene-product mapping of the Mollicutes:Mycoplasma genitalium. *Electrophoresis*, *16*(1), 1090–1094. https://doi.org/10.1002/elps.11501601185

Watrous, J., Hendricks, N., Meehan, M., & Dorrestein, P. C. (2010). Capturing bacterial metabolic exchange using thin film desorption electrospray ionization-imaging mass spectrometry. *Analytical Chemistry*, *82*(5), 1598–1600.

https://doi.org/10.1021/AC9027388/ASSET/IMAGES/LARGE/AC-2009-027388_0001.JPEG

Webb-Robertson, B. J. M., Wiberg, H. K., Matzke, M. M., Brown, J. N., Wang, J., McDermott, J. E., Smith, R. D., Rodland, K. D., Metz, T. O., Pounds, J. G., & Waters, K. M. (2015). Review, evaluation, and discussion of the challenges of missing value imputation for mass spectrometry-based label-free global proteomics. *Journal of Proteome Research*, *14*(5), 1993–2001. https://doi.org/10.1021/PR501138H

Wegler, C., Ölander, M., Wiśniewski, J. R., Lundquist, P., Zettl, K., Asberg, A., Hjelmesæth, J., Andersson, T. B., & Artursson, P. (2020). Global variability analysis of mRNA and protein concentrations across and within human tissues. *NAR Genomics and Bioinformatics*, *2*(1). https://doi.org/10.1093/NARGAB/LQZ010

WELCH, B. L. (1947). THE GENERALIZATION OF 'STUDENT'S' PROBLEM WHEN SEVERAL DIFFERENT POPULATION VARLANCES ARE INVOLVED. *Biometrika*, *34*(1–2), 28–35. https://doi.org/10.1093/BIOMET/34.1-2.28

Wiese, S., Reidegeld, K. A., Meyer, H. E., & Warscheid, B. (2007). Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. *Proteomics*, *7*(3), 340–350. https://doi.org/10.1002/PMIC.200600422

Wilf-Yarkoni, A., Alkalay, Y., Brenner, T., & Karni, A. (2020). High κ free light chain is a potential biomarker for double seronegative and ocular myasthenia gravis. *Neurology(R) Neuroimmunology & Neuroinflammation*, *7*(5). https://doi.org/10.1212/NXI.0000000000000831

Wilkins, M. R., Gasteiger, E., Gooley, A. A., Herbert, B. R., Molloy, M. P., Binz, P. A., Ou, K., Sanchez, J. C., Bairoch, A., Williams, K. L., & Hochstrasser, D. F. (1999). High-throughput mass spectrometric discovery of protein post-translational modifications. *Journal of Molecular Biology*, *289*(3), 645–657. https://doi.org/10.1006/JMBI.1999.2794

Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K., Golaz, O., Sanchez, J.-C., Yan, J. X., Gooley, Andrew. A., Hughes, G., Humphery-Smith, I., Williams, K. L., & Hochstrasser, D. F. (1996). From Proteins to Proteomes: Large Scale Protein Identification by Two-Dimensional Electrophoresis and Arnino Acid Analysis. *Nature Biotechnology*, *14*(1), 61–65. https://doi.org/10.1038/nbt0196-61

Williams, M. E., & Sussex, I. M. (1995). Developmental regulation of ribosomal protein L16 genes in Arabidopsis thaliana. *The Plant Journal*, *8*(1), 65–76. https://doi.org/10.1046/J.1365-313X.1995.08010065.X

Wilm, M. (2009). Quantitative proteomics in biological research. *PROTEOMICS*, *9*(20), 4590–4605. https://doi.org/10.1002/PMIC.200900299

Wolf-Yadlin, A., Hu, A., & Noble, W. S. (2016). Technical advances in proteomics: new developments in data-independent acquisition. *F1000Research*, *5*. https://doi.org/10.12688/F1000RESEARCH.7042.1

Wong, J. W. H., & Cagney, G. (2010). An overview of label-free quantitation methods in proteomics by mass spectrometry. *Methods in Molecular Biology (Clifton, N.J.)*, *604*, 273–283. https://doi.org/10.1007/978-1-60761-444-9_18

Worboys, J. D., Sinclair, J., Yuan, Y., & Jørgensen, C. (2014). Systematic evaluation of quantotypic peptides for targeted analysis of the human kinome. *Nature Methods 2014 11:10*, *11*(10), 1041–1044. https://doi.org/10.1038/nmeth.3072

Xu, C. M., Zhang, J. Y., Zhang, W., & Xie, H. W. (2013). Data Acquisition Strategy for Mass Spectrometers Applied to Bottom-up-Based Protein Identification. *Chinese Journal of Analytical Chemistry*, *41*(7), 1120–1128. https://doi.org/10.1016/S1872-2040(13)60742-2

Yamada, E. A., & Sgarbieri, V. C. (2005). Yeast (Saccharomyces cerevisiae) Protein Concentrate: Preparation, Chemical Composition, and Nutritional and Functional Properties. *Journal of Agricultural and Food Chemistry*, *53*(10), 3931–3936. https://doi.org/10.1021/JF0400821

Yamashita, M., & Fenn, J. B. (1984). Electrospray ion source. Another variation on the free-jet theme. *Journal of Physical Chemistry*, *88*(20), 4451–4459. https://doi.org/10.1021/J150664A002/ASSET/J150664A002.FP.PNG_V03

Zhang, F., Ge, W., Ruan, G., Cai, X., & Guo, T. (2020). Data-Independent Acquisition Mass Spectrometry-Based Proteomics and Software Tools: A Glimpse in 2020. *PROTEOMICS*, *20*(17–18), 1900276. https://doi.org/10.1002/PMIC.201900276

Zhang, H., Liu, Q., Zimmerman, L. J., Ham, A. J. L., Slebos, R. J. C., Rahman, J., Kikuchi, T., Massion, P. P., Carbone, D. P., Billheimer, D., & Liebler, D. C. (2011). Methods for peptide and protein quantitation by liquid chromatography-multiple reaction monitoring mass spectrometry. *Molecular & Cellular Proteomics : MCP*, *10*(6). https://doi.org/10.1074/MCP.M110.006593

Zhang, Y., Fonslow, B. R., Shan, B., Baek, M. C., & Yates, J. R. (2013). Protein analysis by shotgun/bottom-up proteomics. *Chemical Reviews*, *113*(4), 2343–2394. https://doi.org/10.1021/CR3003533/ASSET/IMAGES/CR3003533.SOCIAL.JPEG_V03

Zhang, Y., Wolf-Yadlin, A., Ross, P. L., Pappin, D. J., Rush, J., Lauffenburger, D. A., & White, F. M. (2005). Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Molecular & Cellular Proteomics : MCP*, *4*(9), 1240–1250. https://doi.org/10.1074/MCP.M500089-MCP200

Zhang, Z., Wu, S., Stenoien, D. L., & Paša-Tolić, L. (2014). High-throughput proteomics. *Annual Review of Analytical Chemistry (Palo Alto, Calif.)*, *7*, 427–454. https://doi.org/10.1146/ANNUREV-ANCHEM-071213-020216

Zhu, W., Smith, J. W., & Huang, C. M. (2010). Mass spectrometry-based label-free quantitative proteomics. *Journal of Biomedicine & Biotechnology*, *2010*. https://doi.org/10.1155/2010/840518