2018

Mass Spectrometry-Based Proteomics Analysis of Secreted Proteins

Emmanuel K. Cudjoe Jr
Virginia Commonwealth University

MASS SPECTROMETRY-BASED PROTEOMICS ANALYSIS OF SECRETED PROTEINS

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

By

Emmanuel Kenneth Cudjoe Jr,

Master of Philosophy (MPhil.), Bachelor of Pharmacy (B.Pharm.)

Kwame Nkrumah University of Science and Technology, Ghana

Director: Adam M. Hawkridge, PhD

Assistant Professor, Department of Pharmaceutics

School of Pharmacy

Virginia Commonwealth University

Richmond, Virginia

May 2018

Dedication

This work is dedicated to my family who have been my support and motivation throughout my life. I will not be here today but for the financial support of my father, Mr. Emmanuel Ken Cudjoe, the prayers of my mother, Mrs. Mary Cudjoe, and the love of my siblings, Mavis, Raymond, and Benjamin.

Acknowledgements

I would like to express, first and foremost, my sincerest appreciation to God for his infinite mercies and protection throughout these past 4 years. I would like to show great appreciation to my advisor, Dr. Adam Hawkridge, whose advisorship, and feedback have been invaluable for my professional and personal growth throughout my time at VCU. For all the things I have learned through him and from him, I am eternally grateful. I would also like to acknowledge and thank my committee members, Drs. David Gewirtz, Patricia Slattum, Douglas Sweet, MaryPeace McRae, and Matthew Halquist, all of whom have been mentors to me over the years. I appreciate your advice, time, and support through this journey.

Being the first student to join the Hawkridge lab, I was fortunate to have Dr. John B. Mangrum, who joined a few months later as a postdoc to help me in the early stages of my career in mass spectrometry. My thanks also go to my former lab mate Moral Randeria and current lab mate Zaneera Hassan, for their support and encouragement. I would also like to express my gratitude to past and present members of the Gewirtz lab where I gained my cell and tissue culture skills. Their support and camaraderie during my time at

VCU has meant a lot to me and for that, I will always be grateful. I was also fortunate to be a member of the Bioanalytical Research Group in the department of Pharmaceutics and would like to express my gratitude to everyone, past and present, for constructive criticism and valuable feedback during my presentations. I learned a lot from everyone.

This section will not be complete if I do not show my utmost appreciation for my family and friends, in and out of the US for their immense and immeasurable contributions to my growth and sanity through the course of these 5 years of my PhD life. My parents; Emmanuel Ken Cudjoe and Mary Cudjoe; my siblings Mavis, Raymond and Benjamin, my roommate of 2 years, Samuel Obeng; close friends, Daniel Afosah, Tsemre Tessema, Derrick Afful, Kevin Affram, Daniel Marfo, Matthew Amissah, Benjamin Doi, Eugene Yemofio, Ekow Ewusie-Amissah, Wisdom Gbediame, Kwaku Asiedu, Yoofi Boham, Kasalina Kiwanuka, and Muloongo Simuzingili have all been very instrumental in my growth and development during this period of my life. The journey has been long and arduous, but their support never wavered and their generosity knew no bounds. I shall never be able to pay them back for all their prayers and gifts, in kind and in cash, that I was fortunate enough to enjoy. To them I simply say: God bless you and repay you in greater measure than you have given me.

Table of Contents

Dedication ii

Acknowledgements iii

Table of Contents v

List of Figures xii

List of Tables xiv

List of Abbreviations xv

Abstract xxi

Chapter 1: General Introduction 1

1.1 Mass Spectrometry-Based Proteomics 1

1.2 Instrumentation 2

1.2.1 Ionization 3

1.2.1.1 Electrospray Ionization (ESI) 6

1.2.1.2 Matrix-Assisted Laser Desorption Ionization (MALDI) 7

1.2.2 Mass Analyzers 9

1.2.2.1 Orbitrap-based Mass Analyzers 12

1.2.3 High Performance Liquid Chromatography (HPLC) 14

1.2.1.1 Principles of Nanoscale Liquid Chromatography (nano-LC) 16

1.3 Approaches to Protein Identification 19

1.3.1 Bottom-Up Approach 21

1.3.2 Fragmentation 22

1.3.3 Database Searching for Protein Identification 23

1.4 Quantitative Approaches 28

1.4.1 Label-free Strategies 29

1.4.1.1 Spectral Counting 31

1.4.1.2 Intensity-based Methods 32

1.4.2 Labeling Strategies 32

1.4.2.1 Metabolic Labeling 33

1.4.2.2 Chemical Labeling 36

1.5 The Secretome 38

1.5.1 Classical/Conventional Protein Secretory (CPS) Pathway 39

1.5.2 Non-Classical/Unconventional Secretory Pathways 39

1.5.3 Approaches and Challenges in Secretome Analysis 40

1.6 Protein Glycosylation 43

1.6.1 Glycans 44

1.6.1.1 O-Linked Glycans 47

1.6.1.2 N-Linked Glycans 47

1.6.2 Mass Spectrometry Analysis of Glycoproteins 50

1.6.2.1 Derivatization of Glycans for MS Analysis 51

1.6.2.2 Enrichment Strategies 52

1.7 Dissertation Objectives 53

1.7.1 H1299 Study 53

1.7.2 HepG2 Study 55

Chapter 2: Proteomics Insights into Autophagy 56

2.1 Introduction 56

2.2 Proteomics Approaches to Study Autophagy 61

2.2.1 Whole Cell Proteomics Analysis 61

2.2.2 Subcellular Fractionation 71

2.2.2.1 Autophagosomes 72

2.2.2.2 Lysosomes 75

2.2.2.3 Other Subcellular Fractionation studies 80

2.3 The Secretome: Proteomics Biomarker Discovery 82

2.3.1 Autophagy and Cellular Secretion 83

2.3.2 Proteomics of Autophagy-Associated Secretion 84

2.4 Autophagy and Cancer Therapy: How Proteomics Could Help 86

2.5 Conclusions 88

Chapter 3: Secretome Analysis of Non-Small Cell Lung Cancer in Response to Ionizing Radiation and p53-status 99

3.1 Lung Cancer Statistics: US and Worldwide 99

3.2 Types of Lung Cancer 100

3.2.1 Small Cell Lung Cancer 100

3.2.2 Non-Small Cell Lung Cancer 101

3.3 NSCLC Treatment Options 102

3.3.1 Chemotherapy 102

3.3.1.1 Advances in Chemotherapy 103

3.3.2 Radiation Treatment 104

3.3.3 Limitations of Current Treatment 105

3.3.4 NSCLC Tumor Response to Therapy 106

3.3.4.1 Apoptosis 107

3.3.4.2 Autophagy and Senescence 109

3.4 TP53 and Autophagy 112

3.5 Justification and Objectives of the Study 113

3.6 Experimental 115

3.6.1 Materials and Reagents 115

3.6.2 Cell Culture Conditions and Ionizing Radiation Treatment 115

3.6.3 Western Blot 116

3.6.4 Proteomics Sample Preparation 116

3.6.5 LC-MS/MS Analysis 117

3.6.6 Data Analysis 118

3.7 Results 119

3.7.1 Secretome Analysis of Radiation-treated H1299 cells 119

3.7.2 Quantitative Analysis of the Secretome 120

3.7.3 p53 Expression Status Promotes Differential Protein Secretion Before and After Ionizing Radiation Treatment 125

3.7.3.1 p53wt-IR vs p53-null –IR 125

3.7.3.2 p53wt +IR vs p53-null +IR 127

3.7.4 Association between Significant Proteins and Patient Survival 127

3.8 Discussion 127

3.8.1 Cathepsin D (CTSD) 130

3.8.2 Chromogranin B and Secretogranin 2 (CHGB and SCG2) 131

3.8.3 Glucose-6-phosphate isomerase (GPI) 133

3.8.4 Thioredoxin reductase 1 (TXNRD1) 134

3.8.5 Protein FAM3C 135

3.8.6 Calnexin (CANX) 135

3.8.7 U6 snRNA-associated Sm-like protein LSM8 136

3.8.8 Eukaryotic translation initiation factor 5A-1 (EIF5A)..... 137

3.9 Conclusions 138

Chapter 4: Glycoproteomics Characterization of the SILAC-Labeled HepG2 Secretome as a Platform for the Generation of Stable Isotope Labeled Plasma Proteins..... 140

4.1 Introduction 140

4.2 Experimental..... 142

4.2.1 Reagents 142

4.2.2 Cell Culture Conditions..... 143

4.2.3 Annexin V/Propidium Iodide staining 143

4.2.4 Proteomics Sample Preparation 144

4.2.5 Gel Electrophoresis 147

4.2.6 LC-MS/MS Method 148

4.2.7 LC-MS/MS Protein Identification, Quantification, and Data Analysis 149

4.3 Results & Discussions..... 150

4.3.1 Quantitative Analysis of the SILAC HepG2 Secretome 153

4.3.2 N-Glycosylation of Plasma Proteins..... 158

4.3.3 SIL HepG2 secretome protein levels versus plasma 160

4.3.4 Differential glycosylation of SIL HepG2 secretome versus plasma glycoproteome 164

4.4 Conclusion 168

Chapter 5: Overall Conclusions 170

5.1 H1299 NSCLC Study 171

5.1.1 Future Studies 172

5.2 HepG2 Study 173

5.2.1 Future Studies 174

References 178

Vita 229

List of Figures

Figure 1 - Schematic of a nanoLC coupled to a Q-Exactive Mass Spectrometer 13

Figure 2 - Tandem MS fragmentation of a generic dipeptide 23

Figure 3 - Intact Peptide Identification and Fragmentation 24

Figure 4 - Quantitative Proteomics Approaches 33

Figure 5 - Conventional and unconventional protein secretion 41

Figure 6 - Different forms of D-glucose 45

Figure 7 - N-linked glycans 48

Figure 8 - Types of N-glycans 49

Figure 9 – The Process of Autophagy 58

Figure 10 - Overview of proteomics characterization of Autophagy with respect to cellular and sub-cellular compartments. 62

Figure 11 - Functional forms of autophagy 113

Figure 12 - H1299 Secretome Experimental Design 119

Figure 13 - Preliminary Qualitative Data Analysis of H1299 Secretome 121

Figure 14 - Hierarchical Clustering of ANOVA significant proteins 122

Figure 15 - Volcano Plots showing IR and TP53 effects on H1299 Secretome 126

Figure 16 - Kaplan Meier survival plots for selected proteins 128

Figure 17 – H1299 secretory profile of selected proteins across the different conditions 132

Figure 18 - Comparative Venn Diagrams 138

Figure 19 - Protein Content of the HepG2 Secretome over 72 hours 150

Figure 20 - Annexin V/PI Apoptosis Assay 151

Figure 21 - Protein Distribution in the HepG2 Secretome 152

Figure 22 - Gene Ontology Cellular Component (GOCC) Analysis 154

Figure 23 - Differential protein secretion over 72 hours in HepG2 cells. 155

Figure 24 – Hierarchical Clustering of Secreted Proteins in HepG2 Cells 157

Figure 25 - Plasma N-glycoproteins in the HepG2 Secretome 159

Figure 26 - Plasma Protein Secretion Changes over 72 hours in HepG2 Cells 160

Figure 27 - Peptides from alpha-1-acid-glycoprotein 2 (ORM2) 162

Figure 28 - The effect of lectin enrichment on N-glycoprotein levels between proteins in the HepG2 secretome and plasma 163

Figure 29 - Deglycosylated Peptides from serotransferrin (TF) 166

Figure 30 - Differential enrichment of 68 SIL HepG2 glycoproteins relative to human plasma 167

List of Tables

Table 1 - Summary of reviewed proteomics studies on autophagy 90

Table 2 - Proteins that were found to be significantly differentially expressed (using ANOVA) in the secretomes of p53wt and p53-null H1299 cells before and after ionizing radiation treatment. 124

List of Abbreviations

^{13}C	Carbon 13
^{15}N	Nitrogen 15
^{235}U and ^{238}U	Uranium 235 and Uranium 238
2D-GE	Two-dimensional gel electrophoresis
3-MA	3-methyladenine
Å	Angstroms
ABC	ATP-Binding Cassette
AGC	Automatic Gain Control
AHA	Azidohomoalanine
AMPK	AMP-activated Protein Kinase
APCI	Atmospheric Pressure Chemical Ionization
APPI	Atmospheric Pressure Photo-Ionization
ATG	Autophagy-related Gene
BONCAT	Biorthogonal Non-Canonical Amino Acid Tagging
C-18	Octadecyl carbon
CE	Capillary electrophoresis

CI	Chemical Ionization
CID	Collision Induced Dissociation
CMA	Chaperone-mediated autophagy
Con A	Concanavalin A
COO^-	Carboxylate ion
CPS	Classical/Conventional Protein Secretion
CRM	Charged Residue Model
CV	Coefficient of Variation
DiART	Deuterium isobaric Amine Reactive Tag
DiLeu	Dimethyl leucine
DMEM	Dulbecco's modified Eagle's medium
DMSA	Dimercatosuccinic Acid
DNA	Deoxyribonucleic Acid
Dol-P	Dolichol Phosphate
Dol-P-P-GlcNAc	N-acetylglucosaminylpyrophosphoryldolichol
DSB	Double Strand Breaks
ECD	Electron Capture Dissociation
EDTA	Ethylenediaminetetraacetic acid
EI	Electron-Impact Ionization
ELISA	Enzyme Linked Immunosorbent Assay
ER	Endoplasmic Reticulum
ESI	Electrospray Ionization
ETD	Electron Transfer Dissociation

eV	Electron Volts
EV	Extracellular vesicle
FAB	Fast Atom Bombardment
FBS	Fetal Bovine Serum
FCS	Fetal Calf Serum
FDA	Food and Drug Administration
FDR	False Discovery Rate
FT-ICR	Fourier Transform Ion Cyclotron Resonance
Fuc	Fucose
FWHM	Full Width Half Maximum
Gal	Galactose
GalNAc	N-acetylgalactosamine
GC	Gas chromatography
Glc	Glucose
GlcA	Glucuronic Acid
GlcNAc	N-acetylglucosamine
GlcNAc-P	N-acetylglucosamine Phosphate
GRASP	Golgi Re-Assembly and Stacking Protein
HCD	Higher-Energy Collision-Induced Dissociation
HILIC	Hydrophilic Interaction Chromatography
HPLC	High performance liquid chromatography
HPLC	High Performance Liquid chromatography
HR/AM	High Resolution and Accurate Mass

ICAT	Isotope Coded Affinity Tags
IdoA	Iduronic Acid
IEC	Ion Exchange Chromatography
IEM	Ion Evaporation Model
IPI	International Protein Index
IR	Infrared spectroscopy
iTRAQ	Isobaric Tagging for Relative and Absolute Quantification
kDa	KiloDalton
kV	KiloVolt
LC	Liquid chromatography
LC3	Microtubule-Associated Protein Light Chain 3
LC-MS/MS	Liquid chromatography – Tandem Mass Spectrometry
LE/LYS	Late Endosome or Lysosomes
LINCL	Late Infantile Neuronal Ceroid Lipofuscinoses
LTQ-Orbitrap	Linear Ion Trap – Orbitrap
m/z	Mass-to-Charge ratio
MALDI	Matrix-Assisted Laser Desorption Ionization
Man	Mannose
mM	Millimolar
mRNA	Messenger Ribonucleic Acid
MS	Mass Spectrometer/Spectrometry
mTOR	Mammalian Target of Rapamycin
MV	Microvesicle

NCBI	National Center for Biotechnology Information
Neu5Ac	N-acetylneuraminic Acid
Neu5Gc	N-glycolylneuraminic Acid
NH_2-	Amino functional group
NMR	Nuclear magnetic resonance
NSCLC	Non-Small Cell Lung Cancer
N-X!P-S/T	Asparagine-X-Serine/Threonine
N-X-C	Asparagine-X-Cysteine
^oC	degrees Celcius
PEEK	Polyetheretherketone
PGC	Porous Graphitized Carbon
PMF	Peptide Mass Fingerprinting
PtdIns3P	PhosphatidylInositol-3-Phosphate
PTM	Post-translational modification
Q-TOF	Quadrupole/Time-of-Flight
RCA	*Ricinus communis* agglutinin
RNA	Ribonucleic Acid
ROS	Reactive Oxygen Species
SBRT	Stereotactic Body Radiation Therapy
SCLC	Small Cell Lung Cancer
SEC	Size Exclusion Chromatography
SEER	Surveillance, Epidemiology, and End Results
Sia	Sialic Acid

SID	Surface Induced Dissociation
SILAC	Stable Isotope Labeling by Amino Acids in Cell Culture
SMERs	Small Molecule Enhancers
SMIRs	Small Molecule Inhibitors
SNP	Single Nucleotide Polymorphism
SPIONs	Superparamagnetic Iron Oxide Nanoparticles
SSB	Single Strand Break
TLR	Toll-like receptor
TMT	Tandem Mass Tags
TOF	Time-of-Flight
TOR	Target of Rapamycin
Tris-HCl	Trihydroxymethylaminomethane hydrochloride
UDP-GlcNAc	Uridine diphosphate N-acetylglucosamine
ULK1/2	Unc-51-Like Kinase 1/2
UPLC	Ultra Performance Liquid chromatography
UPS	Unconventional Protein Secretion
UVRAG	UV Radiation Resistance-Associated Gene
Vps34	Vacuolar Sorting Protein 34
WGA	Wheat Germ Agglutinin
Xyl	Xylose

Abstract

MASS SPECTROMETRY-BASED PROTEOMICS ANALYSIS OF SECRETED PROTEINS

By Emmanuel Kenneth Cudjoe Jr, MPhil, BPharm

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 2018.

Director: Adam M. Hawkridge, PhD
Assistant Professor, Department of Pharmaceutics
School of Pharmacy

Secreted proteins play important roles in many cellular functions and molecular processes. Because secreted proteins potentially enter the blood stream, they can serve as valuable measures of health and disease useful for disease diagnosis and prognosis, therapeutic target identification, and patient stratification in personalized medicine. Consequently, significant interest exists in secreted protein analysis within complex biospecimens, particularly blood but significant bioanalytical challenges including the wide protein dynamic range >10 orders of magnitude remain. The cellular secretome therefore represents a viable alternative to direct biomarker discovery in biofluids. Finally, cellular systems are amenable to labeling for the production of intact stable isotope labeled (SIL) proteins that can be used as global internal standards for quantitative proteomics. In this dissertation, two secretome-focused studies were undertaken.

The first study involving candidate biomarker discovery in radiation-induced autophagy utilized the p53-null and inducible H1299 non-small cell lung cancer (NSCLC) secretome. The study identified 364 secreted proteins that were mainly associated with exosomes (N=224) and chaperone activity (N=21). CHGB and SCG2 were identified as potential population-based biomarkers (for patient stratification) due to their consistent overexpression in p53-null H1299 cell secretomes compared to p53-wt cells before and after radiation. FAM3C, CANX, EIF5A, GPI, and TXNRD1 were identified as candidate biomarkers for patient prognosis following radiotherapy due to their differential expression only in response to radiation treatment.

In the second study, a comprehensive glycoproteomics characterization of the SILAC-labeled HepG2 secretome was undertaken. 1635 SIL proteins, 492 of which were major plasma proteins including 192 cancer biomarkers were identified with high sequence

coverage spanning six orders of magnitude. EDTA plasma spiked with the SIL secretomes yielded 63 proteins that were quantified with H/L ratios in all samples out of 1405 total proteins identified. Additionally, LC-MS/MS analysis of the Con A and WGA enriched 72h secretome:plasma sample afforded an opportunity to clearly distinguish between glycoproteins in plasma and the HepG2 secretome that share/differ in N-glycan structures.

Collectively, the two studies reveal the suitability of the H1299 cancer cell secretome as an experimental model for biomarker studies and support the HepG2 secretome as a viable platform for producing SIL glycoproteins.

Chapter 1: General Introduction

1.1 Mass Spectrometry-Based Proteomics

Mass spectrometry is generally viewed as the use of the mass spectrometer in the determination of the molecular weight of various species via the measurement of the mass-to-charge ratios (m/z) of ions [1]. Mass spectrometry has also been defined by John Fenn, the developer of the electrospray ionization technique, as the measurement of atomic mass or molecular weight, which may be enough, usually necessary and always important for identification of different species [2]. Mass spectrometry is one of the most versatile analytical techniques with varied applications in physics, biology, chemistry, as well as medicine. The applications of mass spectrometry range from the analysis of chemicals and the identification of trace amounts of impurities in drug or biological samples to the analysis of biomolecules of which proteins form an important part.

In the past decade or two, mass spectrometry has become the analysis technique of choice not only in research or industry but also in clinical analysis [3]. Mass spectrometry offers a number of advantages over other analytical techniques. Amongst its advantages are its high molecular specificity in the determination of identity, speed of analysis, wide dynamic signal range, quantitative ability and the possibility of coupling with different

separation platforms including the more popular chromatographic systems such as gas chromatography (GC) and liquid chromatography (LC), and capillary electrophoresis (CE) [4, 5].

Proteomics, coined as an analogous term to genomics, refers to the study of the entire protein complement of a cell or tissue or plasma collected at a given time usually to determine cellular function [6, 7]. It is a broad field that encompasses many disciplines including microscopy, biochemical imaging, and immunoassays [7, 8]. When looked at as such, the development of the two-dimensional gel electrophoresis (2D-GE) technique that allowed for the display and identification of multiple proteins in complex matrices from cells, and/or tissues can be viewed as the early beginnings of proteomics [9, 10]. Proteomics may be classified into qualitative proteomics involving the identification of proteins and post-translational modifications (PTMs), quantitative proteomics involving the determination of protein levels, and lastly, protein-protein interaction [9].

Mass spectrometry (MS)-based proteomics may therefore be defined as the large-scale qualitative and quantitative study of proteins using the mass spectrometer. The mass spectrometer has gained increasing importance as the most widely used instrument platform for proteomics analyses involving complex proteomes [11, 12].

1.2 Instrumentation

Pioneering work that formed the basis of mass spectrometry began in the 1880's with the discovery of a new type of radiation that was named Kanalstrahlen, by Eugen Goldstein [13, 14]. Sir J. J. Thomson, whose work resulted in the discovery of electrons in 1897, developed the first mass spectrometer years later in 1913 by building on the

developments of earlier years involving cathode rays [14, 15]. He won the Nobel Prize in Physics for measuring the mass of the electron in 1906 [16]. Thomson's mass spectrometer, referred to at the time as a parabola spectrograph, separated ions according to differences in the paths of travel in electromagnetic fields and the ions were detected by a photographic plate [15]. In the ensuing years, continuous research resulted in several developments to the mass spectrometer. For instance, one of the first improvements in mass spectrometry came by way of Francis W. Aston, a student of J. J. Thomson's at the University of Cambridge, improving upon the resolving power of Thomson's mass spectrometer. This led to the ability to study isotopes of the same element. This happened sometime after World War I around which same time the first electron-impact ionization source was developed by A. J. Dempster of the University of Chicago in a mass spectrometer with a magnetic analyzer that also improved resolution. The work of these scientists formed the basis of modern mass spectrometry and the instrumentation.

The current mass spectrometer consists primarily of three parts, namely an ion source, a mass analyzer, and a detector [7]. The ion source is responsible for the conversion of molecules to ions that are then separated based on their mass-to-charge (m/z) ratios in the mass analyzer and finally detected by the detector [1].

1.2.1 Ionization

Ionization is the process of converting molecular analyte species into charged ions and is necessarily the most critical step in mass spectrometry especially as it relates to macromolecules such as peptides and proteins. It may also involve the transfer of already

charged molecules into the gaseous phase [2]. The conversion of neutral molecules into charged species may occur via protonation or deprotonation, electron removal or addition, or by cationization [2].

Ionization of species affords the ability to obtain information on the m/z of the intact molecule (precursor), and fragments of the precursor ion. These two pieces of information are critical in the identification and quantification of peptides and proteins in MS-based proteomics measurements. There have been many developments in the ionization of molecules since the popularization of mass spectrometry as an analytical technique. The revolution of ionization has gone from hard ionization sources such as electron-impact ionization (EI - 1918), and chemical ionization (CI - 1966) to the so-called soft ionization techniques such as field desorption MS of organic molecules (1969), plasma desorption (1974), fast atom bombardment (FAB - 1981), matrix-assisted laser desorption ionization (MALDI – 1983), and electrospray ionization (ESI - 1984) [15]. The “soft ionization” techniques are so called because of the minimum internal energy transferred to the analyte ions during the process. Other ionization techniques including atmospheric pressure chemical ionization (APCI) and atmospheric pressure photo-ionization (APPI) exist and have application in mass spectrometry analysis involving small molecules. These are often used in the analysis of small biomolecular species unlike ESI and MALDI despite being considered “soft ionization” techniques.

The first and most characterized ionization method is the electron-impact ionization developed by A. J. Dempster. In EI, an electron beam is applied to the neutral analyte molecules in the gas phase causing electrons to be knocked off the neutral molecules, which leads to a positively charged analyte ion [13]. Alternatively, in CI, a reagent gas

molecule (e.g. methane, isobutene, ammonia) is ionized by electrons and the resulting reagent gas ions or radicals react with neutral reagent molecules to cause ionization. Together with electron ionization, CI has a practical mass range of up to 0.5 kDa. Samples amenable to EI and CI must be thermally volatile and stable. Mass spectra obtained using EI are very informative due to the presence of sufficient fragment information. Being also that EI mass spectra are very reproducible, EI is applied in the generation of spectral libraries for many small molecules. As a hard ionization technique, EI may result in either a molecular ion of the neutral molecule or fragment ions. The molecular ion is however hardly observed. Compared to EI, CI produces relatively simpler mass spectra due to less fragmentation.

Mass spectrometers were originally employed in the study of elements and their naturally occurring isotopes. Hard ionization techniques were therefore sufficient for the applications of the instrument. Organic compound studies using mass spectrometers did not happen until the late 1950s and in 1959 peptides and oligonucleotides were sequenced [17]. By the 1980s, mass spectrometry had become routine for the analysis of smaller organic biomolecules. The existing ionization methods could however, not be used for the mass spectrometry analysis of macromolecular species such as proteins due to the inability to analyze intact species without extensive fragmentation. It was not until the development of ESI that macromolecules such as proteins were analyzed successfully. For these thermolabile samples (proteins and oligopeptides), successful proteomics analysis required “soft ionization” techniques such as FAB, ESI and MALDI in order to produce ions with little fragmentation [1, 18].

FAB, among the early “soft ionization” techniques to be developed, involves the ionization of analytes applied onto a target and bombarded with a continuous high current of atoms, neutral molecules, or ions [2]. The analyte is dissolved in a liquid matrix (e.g. glycerol, thioglycerol, m-nitrobenzyl alcohol) prior to bombardment by the beam of particles which results in the desorption of the analyte. Owing to the ability to generate multiply charged species, later “soft ionization” techniques such as ESI and MALDI have become the predominant techniques used in LC-MS/MS-based proteomics with ESI being the most widely used and preferred mode of ionization [18–20].

1.2.1.1 Electrospray Ionization (ESI)

The introduction and development of electrospray ionization is credited primarily to Yamashita and John B. Fenn, but the efforts of the Aleksandrov group are recognized [21, 22]. Fenn’s efforts led to the joint award of half of the 2002 Nobel Prize in Chemistry with Koichi Tanaka who first reported his work with laser desorption ionization of proteins [23]. The pioneering work that set the stage for the possibility of electrospray ionization began with Malcolm Dole as Fenn himself attested to [24, 25].

ESI represented a huge breakthrough in the efforts to extend the applications of mass spectrometry to biomolecules especially proteins and large peptides. In electrospray ionization, a high voltage (~ 1.5 to 5 kV) is applied to a stream/aerosol of molecules emanating from a narrow tip resulting in charged droplets. It occurs at atmospheric pressure [26]. Solvents used for ESI typically combine water and a volatile organic solvent, which improves the sensitivity of the analysis. Low concentrations of acids such as formic acid may be used to enhance ionization of analytes.

Two major theories exist to explain the mechanism of ion production from the charged droplets namely the ion evaporation model (IEM) and the charge residue model (CRM) [21, 24–28]. In the IEM, an applied high temperature then results in the evaporation of the droplets until the concentration of charges on the droplets reaches a limit greater than the Rayleigh limit [26]. Small charged droplets are then formed and from these droplets, ions emerge and are directed into the mass spectrometer [21, 26]. The charged residue model proposes that droplets from the electrospray tip undergo shrinkage due to solvent evaporation. Successive fission events that ultimately result in very small droplets containing single ions of the analyte [21]. It is generally believed that, in ESI, the IEM provides a better explanation for gas phase ion production whereas for very large molecules, the CRM is more likely [24, 26].

As a “soft ionization” technique, ESI has a number of advantages including most importantly, the compatibility with liquid chromatography coupled to the mass spectrometer. Again, the formation of multiply charged ions helps to extend the mass range of fixed range analyzers. Furthermore, being a “soft-ionization” technique, the ions produced retain the structural properties of the original molecule permitting structural analysis. ESI is also useful in the analysis of many non-covalent complexes due to the relatively low energy transferred to molecules during ionization [18]. Very low detection limits in the femtomole range have been achieved with ESI [2, 29]. Amongst its disadvantages, ESI is known to be affected by ion suppression due to matrix effects and may not be sufficient for the efficient ionization of neutral and non-polar or low polarity species such as lipids [18, 29, 30].

1.2.1.2 Matrix-Assisted Laser Desorption Ionization (MALDI)

MALDI, like ESI, is a “soft ionization” method used for the mass spectrometry analysis of macromolecules. The development of MALDI is erroneously credited to Koichi Tanaka citing the Nobel Prize even though the award to Tanaka and Fenn was "for their development of soft desorption ionization methods for mass spectrometric analyses of biological macromolecules" [16]. It is important to note that Karas and Hillenkamp contributed significantly towards what is now referred to as MALDI [31]. In Tanaka’s laser desorption ionization experiment, he used glycerol and an ultrafine cobalt powder (300 Å diameter) mixture as the matrix and showed the ionization of protein molecules up to 100 kDa [32, 33]. Tanaka’s work was very important in showing that large biomolecules could, in fact, be analyzed using laser desorption. Alternatively, for their ultraviolet-laser MALDI work, Karas and Hillenkamp used nicotinic acid, an organic compound, as the matrix which was later proven to be more sensitive and produced more stable ions than Tanaka’s approach [16, 32]. The Karas and Hillenkamp laser desorption approach was more widely accepted and used by the mass spectrometry community owing to its superiority.

In its currently used configuration, MALDI mass spectrometry involves the mixture of the protein or peptide sample to be analyzed with a matrix consisting of a compound able to absorb ultraviolet light from a laser pulse. Organic matrices such as nicotinic acid, sinapinic acid, α-cyano-4-hydroxycinnamic acid, or 2,5-dihydrobenzoic acid are typically used with sinapinic acid and α-cyano-4-hydroxycinnamic acid being the most commonly used matrices for protein and peptide analysis respectively [17, 30]. The mixture of the analyte and the matrix is transferred onto a metal plate and dried, after which the laser pulse is applied. The wavelength of the laser used is dependent upon the organic matrix being used. The matrix molecules absorb the laser and the mixture is desorbed from the

surface of the plate carrying analyte molecules with them into the gas phase [30]. Excited matrix molecules then transfer charges to the analyte molecules resulting in the ionization of the analyte. In addition to UV lasers, infrared lasers also find use in some MALDI applications. MALDI differs from FAB in that in MALDI, a pulsed laser is used to desorb analyte molecules from the target whereas a continuous beam of particles is used in FAB [2]. The use of pulsed lasers and the subsequent generation of ions in pulses has made MALDI most suited to time-of-flight (TOF) mass spectrometers and as a result, the majority of MALDI applications now are used in tandem with TOFs [32]. MALDI, as an ionization technique, is however capable of being coupled with other kinds of mass spectrometers.

Various advantages are associated with the use of MALDI in LC-MS/MS. Unlike ESI, MALDI is not sensitive to the presence of salts, detergents, and contaminants at low concentrations [13]. MALDI also has a practical mass range up to 300 kDa but higher molecular weight species have been observed with MALDI and a high current detector [2]. In MALDI, the use of lasers may result in photodegradation of the analyte and the matrix may interfere with the identification of small molecules up to 0.7 kDa [2].

1.2.2 Mass Analyzers

In the mass analyzer, ions from the ion source are separated and detected. The degree of separation and detection of ions with very close m/z ratios is denoted by the resolution R, of the instrument. Resolution R is calculated as the width of the ion signal peak (Δm – typically taken at half of the peak height) divided by the mass M of the species [30]. Enormous challenges occur in the analysis of various samples by mass spectrometry

owing to the complexity of mixtures analyzed in proteomics [7, 34, 35]. In response to the increasing complexity of samples in proteomics analysis, mass spectrometers continue to see improvements in sensitivity, selectivity, speed and resolution through development of mass analyzers [34, 36].

The development of the mass-resolution double-focusing mass spectrometer by Alfred Nier, an engineer and physicist, contributed immensely to the advancement and widespread adoption of mass spectrometry. In collaboration with Enrico Ferni, Nier successfully separated the uranium isotopes ^{238}U and ^{235}U using his high-mass-resolution double-focusing mass spectrometer [15, 16]. Further advancements resulted in TOF analyzers which were developed in the 1940s but only made public after World War II due to issues of secrecy during the war [37]. TOFs operate under the principle that ions of different masses or m/z will travel at different velocities through a known distance towards a detector; larger ions move slower relative to smaller ions of the same charge. Theoretically, TOF analyzers have limitless mass range even though this is not seen in practice [15].

The development of quadrupole mass analyzers and later quadrupole ion traps in the 1950s was championed by Von Wolfgang Paul and Helmut Steinwedel [1, 15]. Quadrupoles currently constitute the most common mass analyzers in mass spectrometry. They consist of four carefully engineered parallel rods equidistant from one another. Quadrupoles may function as ion traps where they store ions and as mass analyzers for the resolution and measurement of m/z ratios of the stored ions [38]. Von Wolfgang Paul received a Nobel Prize in Physics in 1989 for his research on ion trapping. When functioning as mass analyzers, ions passing centrally through quadrupoles are separated

according to their m/z ratios using electric fields generated by voltages applied to the rods. Quadrupoles may be used to permit the passage of only ions of a given mass-to-charge ratio by controlling the magnitude of fixed direct current and alternating radio frequency voltages applied to the rods. These ions subsequently move into a detector while ions with different m/z ratios simply collide with the rods, as they are unable to traverse the trajectory created by the applied voltages. The stability of quadrupoles and their ability to do tandem MS saw them adopted and used more than Nier's double-focusing mass spectrometer even though the latter was more accurate [15].

Marshall and Comisarow, colleagues at Stanford and then at the University of British Columbia in Canada, were the pioneers of Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometry in 1974 [16, 39]. ICR had originally been used by J. A. Hipple and colleagues in the late 1940s but Marshall and Comisarow applied FT to the technique, truly transforming mass spectrometry [15, 16, 40]. ICR instruments operate by the application of a radio frequency electric field (same frequency as the ions cyclotron) under a uniform magnet field, which ultimately results in the oscillation of these charged species in the electromagnetic field [15, 16]. Detection is recorded as the current produced as rotating ions strike a collector. FT-ICR instruments are highly sensitive and accurate with powerful resolution and large dynamic range of detection and quantitation [7]. Resolution in the sub-ppm range is possible with FT-ICR [15]. ICR instruments and quadrupoles represent different forms of ion traps and are also known respectively as Penning and Paul ion traps [39].

Increasing complexity of samples analyzed (e.g. cell and tissue lysates, plasma, urine, cerebrospinal fluid) in mass spectrometry has necessitated continuous developments in

mass analyzers for increased sensitivity, speed, and resolution [34]. Recent improvements in mass spectrometry have mainly come by way of analyzers involving different combinations of the four main separation principles currently in existence: quadrupole mass filters, time-of-flight, linear ion traps, and Orbitrap™ mass analyzers [18, 34]. Combinations of two or more mass analyzer types in instruments such as the triple-quadrupole, and quadrupole/time-of-flight (Q-TOFs) mass spectrometers represent the early efforts at improving resolution compared to single analyzer instruments [4].

1.2.2.1 Orbitrap-based Mass Analyzers

The orbitrap represents the latest development in mass analyzers and was first described by Makarov in 2000 [41]. These mass analyzers came after the ion traps. While very important in mass spectrometry, ion traps have been shown to have disadvantages including limited range and charge capacity as well as inadequate mass accuracy.

Developed by Makarov, the orbitrap represents an improvement in mass accuracy, reduced size and cost, and increased linear dynamic range as well as charge capacity. The orbitrap operates on the principle of orbital ion trapping where injected ions are trapped between a coaxial outer electrode and an axial inner spindle-shaped electrode [41, 42]. The trapped ions rotate around the spindle-shaped central electrode while oscillating along the horizontal axis. The frequency of the harmonic oscillations is directly proportional to the $\sqrt{z/m}$ [41]. The time-domain signal current produced on the outer electrodes is converted to frequency-domain and then to m/z ratio via Fourier transform [34, 43]. Modern orbitrap instruments are capable of resolving powers of up to 1M and mass accuracies in the low parts per million (ppm) [42, 44]. Different hybrid combinations of

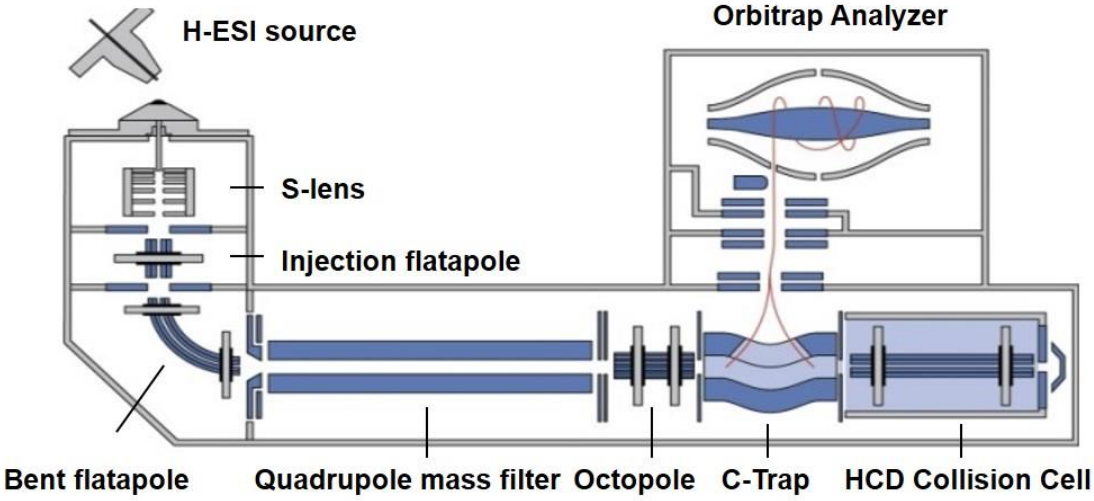


Figure 1 - Schematic of a nanoLC coupled to a Q-Exactive Mass Spectrometer

orbitrap have been developed since the single orbitrap analyzer instrument was first developed and described including the linear ion trap-orbitrap (LTQ-Orbitrap) configuration [34]. The Q-Exactive is a tandem mass spectrometer that combines a quadrupole mass filter with Orbitrap mass analysis **(Figure 1)**. The unique configuration of the Q-Exactive affords it the ability to develop new proteomics methods based on high resolution and accurate mass (HR/AM) analysis including targeted analysis in both MS and MS/MS modes [45]. The Q-Exactive differs from the earlier LTQ-Orbitrap instruments due to the presence of an S-lens, which ensures the transfer of higher number of ions into the MS [11, 34]. Resolutions of 140K full width half maximum (FWHM) at m/z 200 can be achieved on the Q-Exactive [11].

In the Q-Exactive, protein/peptide molecules ionized as they emerge from the ESI tip are transferred into the MS via an ion transfer tube through the S-lens. Ions then pass through an injection multipole into a bent flatapole configured to allow droplets to exit easily. From the bent flatapole, by means of a lens, ions enter the quadrupole where all ions may be

transmitted directly into the adjoining octapole or only a given m/z window is isolated and allowed to go through. Another lens at the exit of the quadrupole transfers ions to the octapole from where they are sent into the C-trap. In MS mode, ions are transmitted from the C-trap to the orbitrap for detection; in MS/MS mode, ions are transferred into the higher-energy collision-induced dissociation (HCD) cell where they are fragmented by manipulating the radiofrequency rods and then sent back into the C-trap, and transmitted into the orbitrap for detection [34]. Trapping of fragmented ions in the HCD cell requires that the offset of the radiofrequency rods in the HCD cell make the cell negative relative to the C-trap and the HCD exit lenses [34]. This also permits the entry of other precursor ions for fragmentation and transfer of all fragment ions to the C-trap and then to the orbitrap for detection.

The complexity of the proteomics samples analyzed using mass spectrometers such as the Q-Exactive necessitates the separation of the component species prior to MS analysis to allow for maximum identification and quantification. Consequently, mass spectrometers are coupled to many separation platforms including high performance liquid chromatography, gas chromatography, and capillary electrophoresis [1, 5, 18]. The majority of mass spectrometry-based proteomics measurements however employ reverse-phase high-performance liquid chromatography coupled to a high-performance tandem mass spectrometer (MS/MS) due to a combination of factors such as high peak capacity, decreased ion suppression, and compatibility of solvents with ESI [46, 47]. Collectively this instrument platform is referred to as LC-MS/MS.

1.2.3 High Performance Liquid Chromatography (HPLC)

The Russian botanist Mikhail Tswett is acknowledged as the pioneer of liquid chromatography based on his separation of various plant pigments from a petroleum-ether extract by passing the extract through a powdered calcium carbonate glass column [48]. Tswett's work did not gain much interest until the 1930s when chromatography was employed for biochemical separations.

Liquid chromatography now represents one of the most popular and important analytical tools in science. The technique involves the separation of samples based on their differential interactions between the components of the sample and two phases: a solid support called the stationary phase and a liquid called the mobile phase. Csaba Horváth's work in the 1960s with capillary columns of 276 µm internal diameter (I.D.) to separate nucleotides formed the foundation for the development of high performance liquid chromatography which greatly improved the separation capability and analysis time of liquid chromatography [24, 49]. Previously referred to as high-pressure liquid chromatography, HPLC involves the use of high pressures to generate the flow of liquid necessary to allow for chromatographic analysis involving packed columns. Developments in instrumentation (such as increased pressure up to 6,000 psi) that consequently resulted in higher performance of the technique prompted the change of the name to high performance liquid chromatography.

Chromatographic sample separation can be done via one of three major separation mechanisms. These mechanisms include polarity, charge, and size. In HPLC/UPLC, four modes of separation are employed namely normal phase (polarity), reverse phase (polarity), ion exchange (charge), and size exclusion/gel permeation (size). Size exclusion chromatography is usually employed for the separation of species based on weight.

Stationary phases have pores into which smaller species go and are retained, therefore eluting much later than larger species, which move around the particles and are not retained. In ion exchange chromatography, samples are separated based on their charge. The stationary phase is usually a charged resin that attracts and retains species of opposite charge. The mobile phase comprises solutions with increasing concentration of salts with oppositely charged species that displace tightly retained or bound species.

Normal and reverse phase HPLC, the most common separation technique used in LC-MS/MS, involve the separation of species based on polarity (hydrophilicity or hydrophobicity). In normal phase, a polar (hydrophilic) stationary phase, usually free silica, and a non-polar (hydrophobic) mobile phase are employed. Hydrophilic samples have higher affinity for and partition more into the hydrophilic stationary phase. They are therefore bound more tightly to the stationary phase and elute later than hydrophobic species that partition more into the mobile phase, and are eluted much earlier. Reverse phase HPLC employs the opposite configuration where the stationary phase is usually a hydrophobic support of alkyl-chain molecules (e.g. butyl, pentyl, octadecylsilane, octyl, cyclohexyl etc.) bonded to silica, and the mobile phase is hydrophilic [43]. Long-chain carbon, usually, C-18 (octadecyl) is used for most proteomics applications involving peptide and oligopeptide analysis while shorter chains such as butyl and octyl are typically used in intact protein analysis [43]. This is the most commonly employed mode of HPLC for tandem mass spectrometry-based proteomics analysis.

1.2.1.1 Principles of Nanoscale Liquid Chromatography (nano-LC)

Conventional HPLC typically involves the use of large columns with internal diameters ranging from 3.5 – 4.6 mm and the flow rates employed in separation using these columns are usually in the mL per minute range [50]. In nano-LC, internal column diameters range from 10 – 150 µm with 75 µm being the most common while typical flow rates in the 10 – 1000 nL per minute range are routinely used [50–52]. The use of micro- and nanoscale columns and flow rates have become increasingly routine, due primarily to applications in biomolecule analysis carried out with MS-based proteomics due to a combination of many advantages of nano-flow chromatography systems. These include the increased efficiency of separation, increased sensitivity, considerable decrease in sample quantity requirement, and decrease in stationary and mobile phase volumes necessary for effective separation [47, 50, 53]. Importantly, the compatibility with nano-ESI introduced by Wilmar and widely employed in mass spectrometry has been very instrumental in the current widespread adoption of nanoscale liquid chromatography in proteomics [47, 50]. This compatibility has contributed to the detection and quantitation of very low concentrations of peptides possible [21, 22, 46].

In liquid chromatography, the efficiency of separation and length of time that a component molecule spends on the column interacting with the stationary phase are affected by a number of factors. These include the dimensions of the column (length, internal diameter I.D.), stationary phase particle size, porosity of the stationary phase packing, and the composition and flow rate of the mobile phase [54]. The efficiency of the chromatographic separation process is dependent on the degree of dilution of the analyte by the mobile phase (chromatographic dilution) during the separation process. Chromatographic dilution (D) is related to the column I.D. as shown in the equation below [50].

$$D = \frac{C_o}{C_{max}} = \frac{\pi d_c^2 \varepsilon (1+k)\sqrt{2LH\pi}}{4V_{inj}}$$

where C_o and C_{max} are the initial and final concentrations of the analyte respectively, k is the retention factor (k = 0 for a non-retained analyte), L the column length, H is the theoretical plate height, V_{inj} is the volume of sample injected, and d_c is the internal diameter (I.D.) of the column.

From the equation, it is evident that decreasing column diameter and length results in a consequent decrease in chromatographic dilution, which ultimately leads to increased analyte concentration and a consequent increase in instrument sensitivity. The use of increasing mobile phase strength in gradient elution may compensate for the loss of sensitivity that results from high chromatographic dilution but this is less than the effect of column diameter on the sensitivity [52]. Furthermore, in nano-LC, peak broadening which is related to both the column I.D. and length may present issues with separation efficiency and lead to poor chromatographic resolution. Peak broadening decreases with decreasing column I.D. and length which is positive for resolution [50]. This also has implications on quantification of analytes since better-resolved peaks allows for more accurate quantification of the component species of the sample. Other factors that may affect resolution due to peak broadening include pre- and post-column dead volume [50, 52]. The effect of dead volume on peak broadening is more pronounced in shorter columns, all other factors being kept constant. To decrease the effect of dead volume on the efficiency of nano-LC chromatographic separation, low volume tubing together made of fused silica with short and tight fittings, usually made of polyetheretherketone (PEEK) are employed [50, 52]. Alternatively, some applications employ nano-ESI emitters with stationary phase material which considerably decreases dead volume [52].

Other factors that influence the degree and quality of separation achieved in chromatographic separation and hence peak capacity are column length and gradient time. In principle, with all other factors kept constant, longer gradients results in less interaction time between sample and stationary phase resulting in faster elution and consequently decreased retention time.

1.3 Approaches to Protein Identification

There are four different technological platforms necessary for mass spectrometry-based protein identification. These include platforms for the isolation, extraction, or separation of proteins from complex mixtures/matrices and instrumental analysis of the separated proteins to obtain structural information [7, 55]. The results of the instrumental analysis then need to be compared to an available gene and/or protein database, and lastly bioinformatics platforms usually involving computer programs and complex algorithms for matching raw data to the database information to give protein identification and quantification [56].

Before the widespread use of mass spectrometry for protein identification and quantitation, techniques such as one and two dimensional gel electrophoresis (1DE/2DE), Western blots, were the major methods for molecular weight determination as well as comparative proteomics [30]. Edman degradation was used in sequencing peptides and proteins by cleaving amino acids from the N-terminus [57]. The techniques use protein migration differences resulting from differences in size and charge to separate proteins; the same protein from different samples may then be compared using different visualization strategies including fluorescence staining. 2DE is now used for protein

separation following which isolated proteins may be analyzed either via top-down or bottom-up proteomics [43, 55]. To extract proteins, gels are excised and solvents applied to the gel to dissolve the protein of interest.

There are two main approaches in mass spectrometry-based proteomics namely top-down and bottom-up strategies. In top-down proteomics, intact proteins are isolated or extracted and analyzed by mass spectrometry whereas in bottom-up proteomics, proteolytic peptides that result from digestion of intact proteins are analyzed by mass spectrometry [55, 58]. The limited fixed mass range of most commercial mass spectrometers makes top-down proteomics rather limited in its routine application in many proteomics labs. While large proteins may be capable of carrying more charge, it is more difficult to ionize them [55]. Again, separation of macromolecules is more challenging compared to the bottom-up approach using conventional modes of separation such as reverse phase chromatography. Furthermore, the effect of analyzer resolution is more apparent with higher molecular weights making it necessary for instruments used in top-down proteomics to be capable of very high resolution [55]. Higher energy is also needed for fragmentation of higher molecular weight species. In spite of the stated challenges, top-down proteomics is useful for a number of reasons. It allows for the characterization of many properties of a protein relating to structure e.g. PTMs [43]. Moreover, since the analysis is of intact proteins, major issues associated with inferring protein presence from peptides are avoided. Top-down proteomics may also be better suited for the characterization of protein-protein interactions. Advancements in mass spectrometry from sample preparation to analyzer m/z range, sensitivity, and resolution present hope for the imminent adoption and routine use of the top-down approach for proteomics studies.

Of the two approaches, bottom-up proteomics also called shotgun proteomics is currently the most widely used by proteomics researchers for protein identification and quantification [6].

1.3.1 Bottom-Up Approach

In bottom-up or shotgun proteomics, proteins are digested into peptides using chemicals or enzyme; enzymes used are specific in their proteolytic activity [43]. The most commonly used enzyme is trypsin, which being specific in its cleavage of the peptide bond, cleaves only at the C-terminal of arginine and lysine residues unless these are C-linked to proline or N-linked to aspartic acid. Trypsin is also used due to the ability of tryptic peptides to be charged positively at both N- and C-termini. Due to the specificity of the enzyme, an *in silico* digestion of all proteins in the database may be carried out using the specific rules of cleavage and a list of the theoretical m/z ratios of the resulting peptides generated. In the identification of proteins, the m/z values should fall within a given range of the theoretical or *in silico* generated m/z values. It is therefore very essential that the charge states of the ions be accurately determined as this provides an accurate starting point for the identification process. High-throughput protein identification relied on the use of the Peptide mass fingerprinting (PMF) for quite a long time. PMF is a technique in which a single, usually unknown, protein is digested and the masses of the resulting peptides determined experimentally and compared to a database of proteins and their corresponding peptides [56].

Despite being the most commonly used proteomics approach, various issues associated with the use of peptides to make inferences on the presence and quantities of proteins in

an analyte mixture remain. Peptides that are not specific to a given protein may be wrongly assigned which affects protein identification. In addition to potential wrong assignment of peptides to proteins, not all peptides coming from a digest protein may be observed or identified in the mass spectrometry analysis resulting in lost information which may be important, necessary even, for the identification of possible PTMs, sequence variation, or domains for binding etc. [43].

1.3.2 Fragmentation

Tandem MS information obtained from fragmentation of peptides is the backbone of some of the search algorithms used in protein identification. Many fragmentation methods exist but the most predominantly used method in many mass spectrometers including the triple quadrupoles, the Q-TOF, and the LTQ-Orbitrap is collision induced dissociation (CID) [34, 59]. In CID, precursor molecular ions are accelerated and made to collide with the molecules of an inert gas (e.g. nitrogen, helium, or argon) which results in bond cleavage through the conversion of kinetic energy into internal energy [59]. CID usually involves the use of low energy gas molecules (less than 100 eV) for the collision. This produces cleavage of amino acids predominantly along the peptide bond and yields b-ions, y-ions, and neutral water or ammonia losses. Many search algorithms therefore use b- and y-ion information in identification of peptides **(Figures 2 & 3)** [59].

A higher energy variation of CID known as higher energy collisional dissociation (HCD) is used in orbitrap-based systems including the Q-Exactive [34]. Other fragmentation methods used in mass spectrometry include electron capture dissociation (ECD), surface induced dissociation (SID), and electron transfer dissociation (ETD) [59, 60].

1.3.3 Database Searching for Protein Identification

Proteins are identified by one of three main methods including database searching using a database of proteins of the relevant species, de-novo sequencing, or tag-based

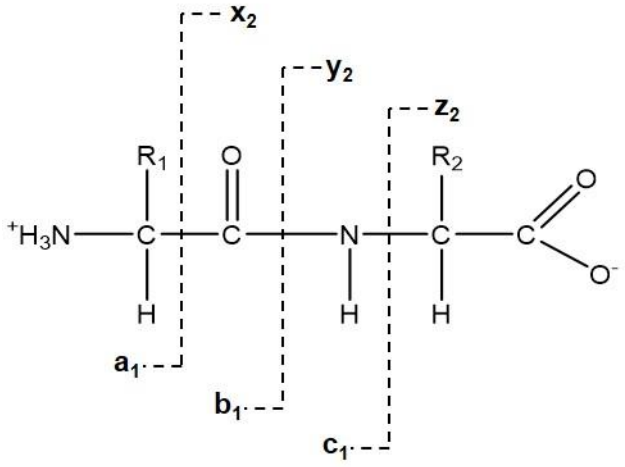


Figure 2 - Tandem MS fragmentation of a generic dipeptide

algorithms [61]. Database searching is by far the most popular method.

There are three main database searching approaches to protein identification including the use of only peptide (precursor) masses, the use of tandem MS information from one or more peptides of a protein, or the use of both mass data and amino acid sequence information or physicochemical data directly related to the amino acid composition [59, 62].

The main aim of database searching in protein identification is to correctly identify proteins in the experimental data that are in the database or in the case of proteins which are not in the experimental database, identify the proteins that are closest in sequence homology to the experimental proteins [62]. There are a number of protein databases, the most popular amongst them being the Uniprot, human IPI, and NCBI databases. The Uniprot database uses the Fasta format, which comprises amino acid sequence represented by

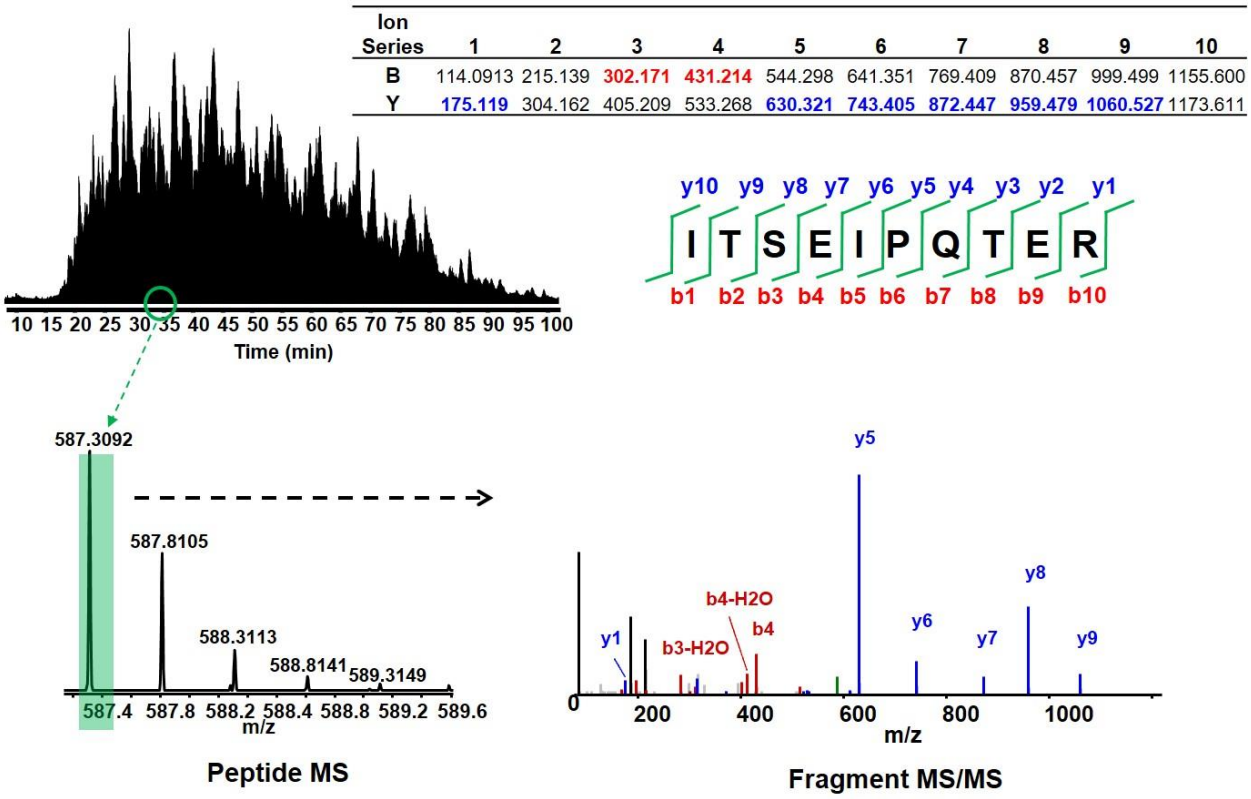


Figure 3 - Intact Peptide Identification and Fragmentation

The monoisotopic peak 587.3092 of the intact peptide ITSEIPQTER was identified around ~ 35 mins and picked for MS/MS fragmentation and analysis. Many y-ions and a couple of b-ions were identified from the MS/MS spectrum of the fragmented peptide.

single letters and written as a string of characters usually 60 characters per line and never exceeding 80. The databases consist of all experimentally identified and validated proteins (and sometimes, theoretical proteins based on mRNA expression).

The acquisition software of an LC-MS/MS system produces a chromatogram of both peptide and fragment ions **(Figure 3)** from the raw data obtained from the instrument. Within the chromatogram are the spectra (represented by single peaks positioned at the center of the mass to charge ratio distribution and referred to as centroids) of all identified precursor and fragment ions. The spectra are identified and converted to m/z ratio data by complex deconvolution algorithms within the various software packages. The accurate estimation of the m/z ratio of a precursor ion is predicated upon the correct determination of the charge of that precursor. Using the isotopic peak distribution of a precursor ion, deconvolution algorithms are used to determine the charge states of a peptide and hence, the m/z ratio [59].

In-silico peptides are generated through the digestion of all proteins in the database using computer algorithms. The in-silico peptides are then fragmented by the program and the similarities between these theoretical computer-generated spectra compared to the experimental spectra generated on the instrument [61]. The quality of the spectra generated by the mass spectrometer therefore plays a key role in correct matching of sequences and subsequent identification. The quality of the experimentally generated spectra is therefore very important; the poorer the quality of a spectrum, the higher the likelihood of false identifications based purely on chance. Algorithms assessing spectral quality therefore exist to help eliminate poor quality spectra. Some of these algorithms use statistical regressions to determine the quality of spectra and hence determine which spectra to eliminate [59, 63].

Various models also exist which are used by different programs or algorithms to predict the fragment ions and spectra of a given peptide against which to search the experimental

data [59]. The fragment ions are usually predicted to be mostly composed of b- and y-ions, which result from breaking the amide bond between amino acids in the peptide structure. Other ions such as a-, x-, c-, and z-ions may arise but these are not normally used [59]. The MS/MS spectra of all experimental peptides that were within the allowed mass range of a given precursor mass are matched to theoretical fragment spectra and scores assigned to all the matching spectra using a model [59, 64]. The score is a function of the calculated probability that the match is a true match or happened by chance [59, 62, 64]. A program such as Mascot uses this probability scoring to filter out the candidate precursor masses [64]. The highest scoring matches, which have the lowest probability of being chance events, are then used to estimate the protein scores. The precursor m/z values are compared to the in-silico or predicted values and a list of candidate proteins generated by the algorithm. The candidate precursors (peptides) are validated according to a number of criteria which may include, for instance, the number of fragment ion matches [59, 65]. Unique peptides are more easily matched to their respective proteins. For the peptides that are not unique, a program like MaxQuant employs the “Occam’s Razor” rule and assigns non-unique peptides to the protein with the most identified peptides. Mascot uses the “Principle of Parsimony” and chooses the protein ID with the simplest and most reasonable justification [64]. Some algorithms (such as Percolator in Proteome Discoverer) employ a target-decoy search tool in the protein identification sequence, which reveals the number of false identifications and hence sets the false discovery rate (FDR). The decoy databases comprise ‘nonsense’ protein sequences (reverse or random protein sequences of the actual database). The decoy database may be concatenated with the actual database and searched as one or the actual and decoy databases may be

searched individually against the experimental data. Any hits in the decoy database represent false identifications or false positives which are then used to determine the FDR calculated as the ratio of false positives to total number of identifications [61]. Percolator employs an iterative machine learning algorithm to learn different features of correct and incorrect identifications in the actual and decoy databases respectively [61]. The target decoy database searching which provides the FDR may be a limitation in analyzing lots of different datasets of the same proteome as opposed to one or a few [58]. It is important to note however, that when analyzing many datasets, there is a disproportionate increase in true positives versus false positives owing to the duplication or repeated sampling of the same proteome. The independence of false positives on the proteome against which the data is searched may give rise to this [58]. Again, some search algorithms such as Mascot employ probability scoring and rely on the size of the database to set the FDR threshold. Accordingly, the use of concatenated decoy databases increases the size of the database and therefore raises the FDR which translates into fewer peptide identifications or hits [61]. A workaround is to keep target and decoy databases separate and searching the data against them separately.

The identification of non-unique proteins by the programs is a limitation of current methods. Proteins with significant sequence homology, such as isoforms of the same protein, may not be correctly separately identified by the program and such proteins are usually put into one protein group when they may in fact be different proteins. Again, the accuracy of protein identifications may be further complicated by the presence of modifications on the proteins. Modifications may be PTMs or have occurred during the sample processing steps. This introduces a complexity into the identification process

seeing as the modifications may increase exponentially the number of possible hits for a given mass to charge ratio. The software and the user may not be able to provide a list of all these modifications to accurately determine the exact sequence matches for the proteins in the experimental data. One other limitation of protein identification using database searching is the limited number of proteins available in the protein databases. However, until there is a comprehensive and exhaustive list of all possible proteins, there is always the possibility of missing important information in the mass spectra generated by the instrument. An exhaustive protein database seems an impossible feat right now considering all the possible PTMs and SNP/splice variants that may exist for each protein. In organisms with incomplete genome sequences where proteomics studies may seek to identify novel proteins, this is particularly problematic.

De-Novo sequencing algorithms attempt to circumvent this problem of limited databases but that has its own challenges. In de-novo sequencing, algorithms extract information on the (partial) sequence of peptides using tandem MS information. Fragmentation produces successive fragment ions (and corresponding peaks) from which the sequence may be constructed. For two consecutive fragments, the difference in mass (distance between the peaks on the m/z scale) represents the mass of an amino acid residue. This can be done for all peaks in the series and used to generate the (partial) sequence that may be blasted against the genome database of better-characterized species to find sequences with appreciable homology.

1.4 Quantitative Approaches

Quantitative proteomics is an essential component of mass spectrometry-based proteomics studies and provides information to augment identification of proteins/peptides in complex mixtures. Many approaches exist but these are broadly grouped into label-free and labeling strategies. The label-free approach involves the quantitation of proteins/peptides without the need for tags whereas labeling approaches use different labels/tags to enable comparative analysis between samples in the same workflow. Quantitative approaches may be absolute or relative [66]. **Figure 4** is a summary of the major quantitative strategies currently used in the majority of proteomics studies.

1.4.1 Label-free Strategies

Label-free proteomics is possibly the simplest and fastest form of quantitative proteomics owing to the absence of extra labeling steps [67–69]. Here, different samples are analyzed separately and not mixed together with each sample going through the same processing steps. There is inherently less sample 'complexity' in label-free quantitation relative to labeled samples, which are always a combination of differentially labeled proteomes. By virtue of its simplicity and the fact that no special reagents have to be purchased, label-free strategies are also the least costly protein quantification strategies [67].

Due to the direct analysis of proteomes without the use of labels, label-free proteomics can be applied to all samples whether *in vitro* or *in vivo*. In regards to the instrument analysis, more peptides are analyzed and detected in label-free workflows than in labeled workflows for the same instrument duty cycle. There is maximum amount of scanning to obtain more microscans per second for the same ion species than in labeling methods. This affords the ability to obtain the maximum amount of data in label-free analysis since

the instrument scans multiple species (e.g. heavy and light labeled) for the same ion in labeling strategies and ultimately results in longer scan time and less tandem mass spectra [60]. It can be assumed that in the time that it scans two or more differentially labeled species (multiplex) of the same ion in labeled workflows, it will scan two or more different ions in label-free workflow. This maximum scanning therefore potentially results in greater depth of proteome coverage in label-free proteomics [46].

Label-free approaches are however, beset with a number of issues principal among them being the inherent differences in the final sample to be analyzed due to separate processing for different samples. There is less quantitative precision compared to labeling techniques due to the susceptibility of label-free quantification to pre- and post-analytical variability. Owing to the fact that samples are prepared separately, there is the likelihood of introducing bias from the analyst, or the materials used. Different tubes in which sample processing is carried out may have different properties, which may affect sample losses, for instance, through differential protein/peptide binding to the walls of the tubes. Again, instrument performance differences from sample to sample during analysis may affect quantitation. This may consequently affect reproducibility of analysis in label-free workflows, which significantly impacts quantification. Consequently, high reproducibility from sample run to run is necessary in label-free quantitation. The development of different software with improved algorithms for normalization and data analysis such as LFQ in the MaxQuant platform seek to combat the effect of pre-analytical variability in label-free workflows [67].

Label-free quantification of peptides and/or proteins is done using one of two approaches: spectral-based methods or intensity-based methods [69–71].

1.4.1.1 Spectral Counting

In spectra-based methods, the number of experimental MS/MS spectra of a given peptide/protein that match with known spectra in the database is used to quantify peptides/proteins. The number of matched spectra is referred to as a spectral count. Spectral count methods rely on the assumption that the more abundant a protein is in a mixture, the greater the number of peptides it will produce and hence the more spectra it will generate [60, 72]. Various studies have demonstrated a strong relationship between the spectral counts and relative abundance of peptides/proteins [60]. There are however, problems with spectral counting methods that should prompt caution in interpreting results. For example, the size of a protein largely determines how many peptides can be produced from it. Consequently, for two proteins with equimolar amounts in a matrix, the larger protein tends to produce more tryptic peptides and hence, more MS/MS spectra [60]. Further, a number of factors including ionizability of peptides and efficiency of peptide fragmentation affect MS signal intensity, which affects the selection of precursor masses for fragmentation [60]. These issues tend to introduce bias into peptide/protein quantification. Spectral counting methods have also been found in some studies to provide less precise quantitative information and be less likely to detect small changes in protein abundance compared to labeled methods [73].

To attenuate the problem arising from large proteins having more spectra, normalized spectral counts have been developed and used with success [6, 74]. This notwithstanding, spectral methods are generally not as commonly used as intensity-based methods for protein quantification.

1.4.1.2 Intensity-based Methods

Intensity-based methods use the height or area under the curve in the chromatogram as a quantitative measure of how much peptide or protein there is present in a sample. There are a host of intensity-based label-free quantification tools currently in use in proteomics amongst which is MaxLFQ [67], an algorithm proposed and used by the Mann group at the Max Planck Institute of Biochemistry in Germany. In MaxLFQ, the problem of differences resulting from separate sample treatments is solved by a concept the authors term 'delayed normalization'. This concept involves keeping normalization coefficients (factors) as free variables under the assumption that the proteome of a cell or organism remains largely unchanged between conditions [67]. The response (intensity) of a peptide in a fraction is multiplied by an unknown normalization coefficient. This is done for all peptides in all fractions. The differences between the logarithms of each product for all possible fraction pairs is then squared for each peptide and the results summed up. The algorithm then determines the values of the coefficients as those values, which produce the least value of the sum of the logarithmic fold changes. These values consequently translate into the least differences in peptide abundances across the various fractions, which constitutes the basic assumption underlining the LFQ algorithm.

1.4.2 Labeling Strategies

Sample labeling is one of the more important strategies in quantitative mass spectrometry. Labeling approaches do not only enable the multiplexing of samples, which reduces analysis times, but also ensure the elimination of the effects of bias in the sample preparation steps. There are two major approaches to labeling in proteomics studies including metabolic labeling and chemical labeling.

1.4.2.1 Metabolic Labeling

Metabolic labeling involves the introduction of stable isotope labeled amino acids into protein structure using the metabolic machinery of cells and/or whole organisms. This approach to stable isotope labeling has achieved high rates of success, particularly in cell culture but also in whole organisms such as *Caenorhabditis elegans* and the fruit fly, *Drosophila melanogaster* [75–77]. Among the currently used metabolic labeling approaches

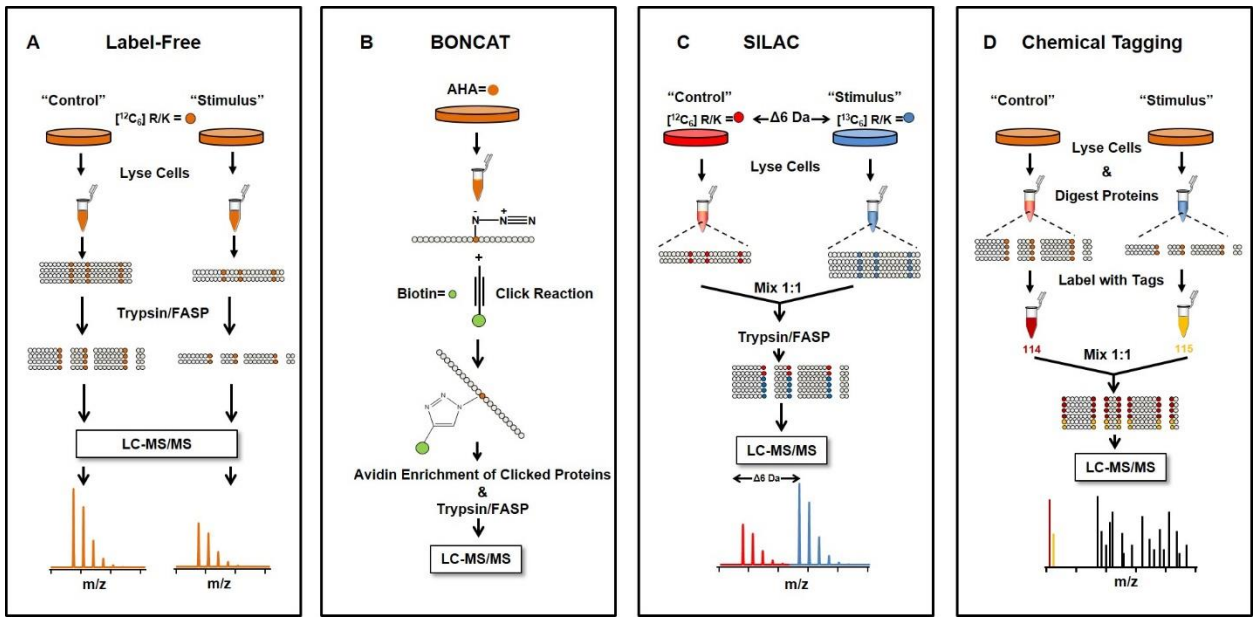


Figure 4 - Quantitative Proteomics Approaches

A) Label-free quantitative proteomics involving separate sample prep for each sample. B) Click Reaction involving metabolic labeling using the methionine mimetic, azidohomoalanine (AHA). C) Typical workflow for Stable Isotope Labeling by Amino Acids in Cell Culture involving 1:1 w/w sample mixing before sample prep. D) Chemical tagging approach (TMT, iTRAQ) involving chemical labeling of samples following trypsin digestion

(i.e. Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC), Biorthogonal Non Canonical Amino Acid Tagging (BONCAT), and ^{15}N labeling), SILAC is the most commonly used.

SILAC, together with the other stable isotope-based labeling strategies, currently represent the gold standard in quantitative proteomics [67]. SILAC involves the supplementation of cell culture media with ^{13}C- and ^{15}N-labeled amino acids (arginine and lysine) which are taken up by cells for use in protein formation [78, 79]. Two sets of cells (or more depending on the number of treatments and cell types being studied) are therefore cultured, one with regular (light) arginine and lysine and the other with the 'heavy' labeled forms. Lysates from the different cell cultures are mixed in equal proportion, digested, and analyzed via LC-MS/MS. Heavy and Light labeled peptides, having similar physicochemical properties, elute from the HPLC column at the same time and are analyzed by the mass spectrometer but are differentiated by the mass spectrometer by a constant mass difference between these peptides due to the different stable isotope labels.

More recently, the applications of SILAC have been extended to biospecimen, including patient tumor tissues, in what has been termed Super SILAC [80, 81]. In the Super SILAC approach, a master sample of proteins is generated by combining lysates from different SILAC heavy-labeled cell lines of a given cancer or tumor type. The master sample contains a near-exhaustive mixture of proteins expected to be present in the tumor cells. This is then spiked into different samples (e.g. patient tumor before and after radiation treatment) and serves as an internal standard to measure changes in protein types and quantities in the samples. In a recent proof of concept study, our lab has been able to demonstrate the applicability of the Super SILAC approach to study the plasma proteome in human plasma samples [82].

BONCAT, another metabolic labeling technique uses the copper-catalyzed azide-alkyne cycloaddition reaction known mostly as click chemistry [83, 84]. In the application of this technique in cell culture, an amino acid analog of methionine bearing an azide group (that can be 'clicked' with an alkyne functional group) is fed to cells as part of the culture media. In the absence of methionine in the culture media, cells incorporate this analog into protein formation resulting in proteins having this azide functional group, which can then be pulled down with the alkyne (clicking reagent). This strategy is particularly useful in secretome studies in which only proteins secreted by the cells are wanted. An added advantage of this strategy in secretome studies is the use of serum in the labeling media, which ensures that cells have requisite nutrients to grow as they normally would. Unlike other labeling strategies, BONCAT is not inherently quantitative. BONCAT is typically used to facilitate protein identification rather than quantification.

Metabolic labeling including SILAC, has the advantage of avoiding bias due to pre-analytical variability since samples are mixed early in the sample preparation protocol and undergo essentially the same treatment steps [85]. Other advantages of SILAC include being robust, simple to perform, and the ability to fully label whole proteomes without the need for chemical reactions [79, 85].

Labeling strategies, however, are generally more expensive than label-free strategies and require more steps before sample analysis [67]. Relative to chemical labeling strategies, metabolic labeling also has the disadvantage of being limited in the number of labels that can be used (usually 3) as well as being unsuitable for labeling clinical and animal-based samples or cells that do not grow very well in labeled media [60, 67, 70]. Researchers, including our lab, have faced a problem with BONCAT where the cells were unable to

grow in the media containing the labeled methionine (azidohomoalanine, AHA) past a few hours. The workaround for this method is therefore to employ short labeling times [86, 87] where the cells are grown in regular media till desired confluency and exchanged with AHA containing media for 30 – 60 minutes and the media collected after. This makes the BONCAT approach mainly useful for studying short-term secretory profiles instead of long-term secretion.

1.4.2.2 Chemical Labeling

Chemical labeling techniques, unlike metabolic labeling, involve the labeling of proteins or peptides following protein extraction and/or digestion. Isotope Coded Affinity Tags (ICAT) have been used in the past but Isobaric Tagging for Relative and Absolute Quantification (iTRAQ) and Tandem Mass Tags (TMT) represent the two main isobaric chemical labeling tags currently used routinely in quantitative proteomics [66, 88]. N,N-Dimethyl leucine (DiLeu) and Deuterium isobaric Amine Reactive Tag (DiART) are two less commonly known and less frequently used isobaric tags [89].

Isobaric tags are chemical species of identical mass and chemistry that permit these tags to co-elute at the same retention times when bound to peptides to be analyzed by mass spectrometry [90]. There are three regions of isobaric tags, four with the peptide included. The regions are the reporter region which is the tag (consists of varying ^{13}C substitutions depending on the intended tag mass), a mass normalization region that is synthesized to balance the tag mass and ensure that all tags have the same mass regardless of label, and a peptide/protein reactive group that reacts with the peptide [90]. Tags are manufactured to ensure that cleavage from CID yields reporter ions of different masses

that are used to quantify the peptides to which they were bound and from which they were cleaved.

Isobaric tags have the advantage of being used simultaneously with multiple samples; up to 8 or 10 samples can be multiplexed in iTRAQ or TMT respectively and this affords a very precise way to determine expression differences between large sample numbers [91]. This ability to combine many unique samples increases sample throughput and can permit the analysis of many more samples per unit time compared to label-free strategies. In isobaric tagging, same peptides (from different samples) labeled with different tags have the same mass regardless of which sample they come from unlike SILAC where labeled peptides from different samples have different masses. Peptides labeled with the different isobaric tags therefore arise as one peak in the MS precursor ion scan [89]. This, in fact, increases the precursor ion intensity and decreases sample complexity relative to SILAC, which may further increase depth of proteome coverage. Duty cycles may thereby be higher than in SILAC. Quantification is done by a direct comparison of the reporter ion intensities in the MS/MS spectrum. Peptide identifications are done with the fragment ion information obtained in the high m/z region of the product ion spectrum. Peptides which did not pick up the tags/labels are not quantified as there are no reporter ions from the CID events in MS/MS [90]. The labeling efficiency of the tagging therefore plays a very important role in protein identification and by extension differential abundances observed. Isobaric tags can be used on *in vitro* as well as *in vivo* samples because the peptides are labeled after digestion unlike in classical SILAC where proteins are labeled inside the cells.

Pre-analytical variability may not be entirely avoided in isobaric tagging techniques because samples are labeled after enzymatic digestion. Any variations resulting from reagents, materials, or analyst may therefore affect the results. In spite of these potential issues, iTRAQ has been shown to give accurate and very reproducible results with mean CVs of 0.09 in one study [88, 92]. Furthermore, because the samples are labeled before analysis in the mass spectrometer, post-analytical variability, including differential instrument performance from sample to sample, is avoided. However, pre-analytical variability may persist owing to the separate tryptic digestion of samples; differences in digestion efficiency may bias results. The overall reproducibility of isobaric tagging workflows may not be as good as with SILAC but certainly more so than label-free approaches.

1.5 The Secretome

First publicly used by Tjalsma et al., the term secretome represents the complement of all proteins secreted into the extracellular environment by cells, tissues, or organisms [93, 94]. It is estimated that about 10 – 15% of known and predicted human proteins are soluble and may be secreted. Protein secretion constitutes a part of the natural mechanism of homeostasis. Cells also secrete proteins and other macromolecules in the body as a response to injury. For example, inflammatory cells secrete various factors as part of the immune response while tumor cells may secrete factors to promote their growth or in response to treatment [95].

Due to the vast amounts of information contained in cell secretomes, they have generated interest as important resources for many MS-based proteomics studies. The cancer

secretome in particular has been shown to contain microvesicles and exosomes which contain proteins that play vital roles in tumor progression *in vivo* [96, 97]. Protein secretion may occur via one of two broad secretory mechanisms – the classical/conventional secretory pathway or the non-conventional pathways.

1.5.1 Classical/Conventional Protein Secretory (CPS) Pathway

Conventional protein secretion refers to the secretion of proteins involving transport through the endoplasmic reticulum (ER) to the Golgi Apparatus and ultimately to the cell membrane from where the proteins are secreted into the extracellular environment [98]. Proteins secreted via the CPS pathway have N-terminal secretory signal peptides, usually about 16-30 amino acids, or transmembrane domains that direct them to the membrane of the ER where they are translocated into the lumen [96, 98–100]. In the lumen of the ER, signal peptides are cleaved off and the proteins transported to the Golgi Apparatus in coat protein II (COPII)-coated vesicles [99].

1.5.2 Non-Classical/Unconventional Secretory Pathways

The long held belief in the scientific community was that protein secretion only occurred via the classical pathway and a signal peptide was necessary for protein targeting to the secretory pathway. It is now widely believed, however, that alternative pathways exist for protein secretion, broadly classified as unconventional protein secretion (UPS). Recent research in different labs have delineated the molecular processes involved in different aspects of UPS.

Proteins that go through the UPS pathway may be subdivided into two subgroups including those that have a signal peptide and enter the ER but do not go through the

Golgi Apparatus prior to secretion and those that have no signal peptide popularly referred to as leaderless proteins [98]. Proteins without signal peptides may reach the extracellular environment through one of three mechanisms namely pore-mediated translocation (Type I), ABC transporter mediated transfer (Type II), or autophagosome/endosome-associated transport (Type III) [98]. Types I and III involve translocation across the cell and autophagosomal membranes respectively while Type II typically involves acylated peptide and yeast mating peptides. The secretion of proteins with secretory signal peptides via the UPS pathway is referred to as Type IV UPS.

UPS pathways are believed to be stress-induced especially in situations where stress results in an impairment in the classical pathway. For instance, Type IV protein secretion via the Golgi bypass mechanism involving proteins of the Golgi re-assembly and stacking protein (GRASP) family constitutes one of the major consequences of ER stress induction. Similarly, autophagosome-mediated protein secretion (Type III) of proteins such as interleukin 1-β (IL1- β) and acyl-CoA-binding protein (ACBP) has been shown to be GRASP-dependent [98, 99, 101].

1.5.3 Approaches and Challenges in Secretome Analysis

The use of immortalized cancer cells as experimental models for studying cancer has become the mainstay of cancer research *in vitro* [102]. Cell lines represent a readily available model that can be studied under carefully controlled experimental conditions. Cell lines are also relatively easy to grow and manipulate genetically. In spite of the merits of using cell lines as experimental models for basic or applied research, limitations exist including the lack of a tumor microenvironment associated with 2D cell cultures. Howe

ver, the imminent rise of 3D cell cultures seeks to make cell cultures more representative of what goes on *in vivo* while retaining the ease of growth and amenability to manipulation[102].

Cell secretomes have emerged as viable alternatives to the use of plasma in the discovery of proteins that may serve as important biomarkers of disease. This is in part due to the difficulties observed in direct plasma-based biomarker studies including the vast

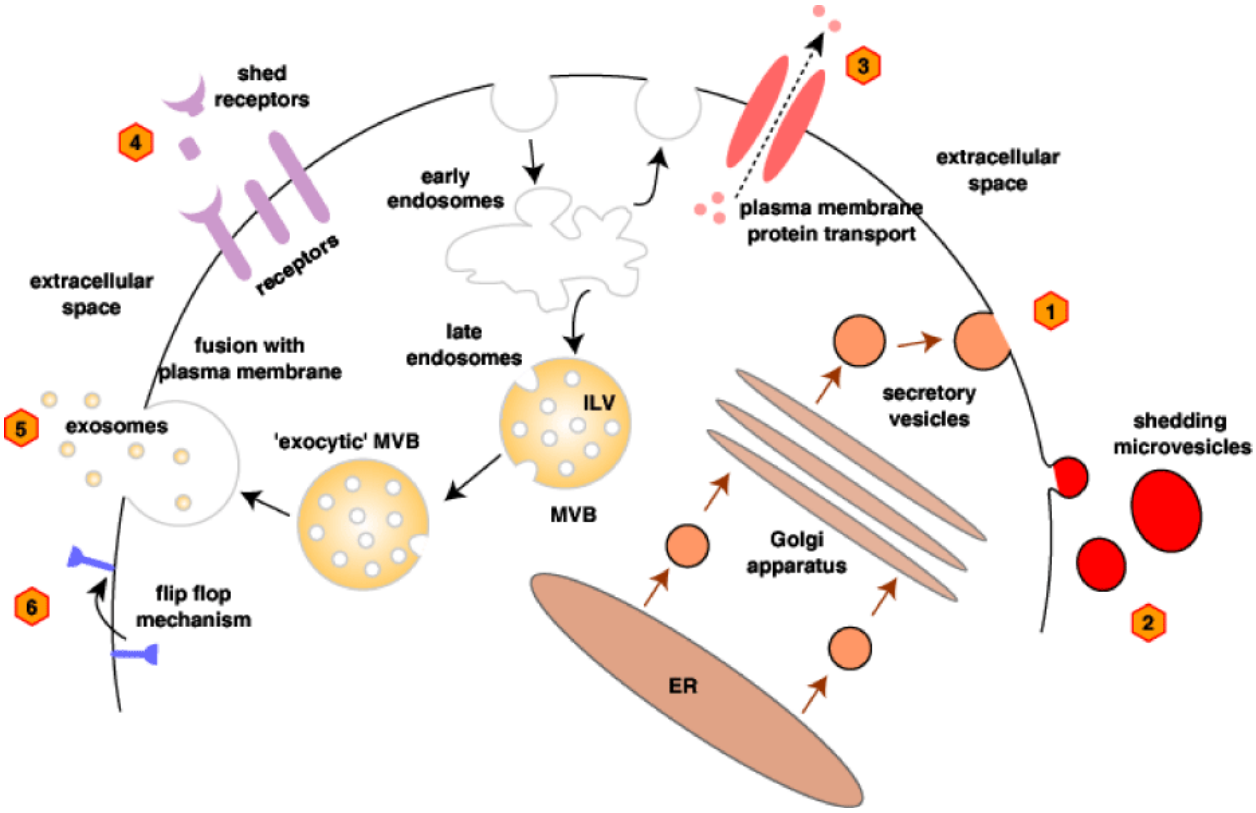


Figure 5 - Conventional and unconventional protein secretion

Classical secretion is depicted as 1; processes 2 – 6 depict the different types of unconventional secretion. Image taken from https://www.omicsonline.org/articles-images/JPB-05-Editorial16-g001.html

abundance of only a minority of proteins and the large dynamic range (~10 orders of magnitude) therein [102]. Furthermore, being representative of proteins likely to be found in biological fluids (e.g. plasma, cerebrospinal fluid), the secretome represents an important surrogate with the potential for use in biomarker discovery studies [103]. Cancer

cell secretomes are very likely to contain proteins involved in key functional activities *in vivo* owing to the dysregulation of various pathways in cancer.

Typically, cell and tissue culture media is supplemented with up to 10% serum (e.g. fetal bovine serum - FBS, or fetal calf serum - FCS). The high protein and amino acid content of serum necessitates the use of serum-free media in efforts to circumvent the high protein background. Cells are therefore grown in serum containing media for some time following which the media is removed and the cells washed extensively to remove the serum. However, it is known that despite the washing, some serum persists. Being necessary therefore to be able to distinguish between proteins secreted from cells and proteins already present in culture media from serum, labeling strategies such as SILAC and/or Click Chemistry are sometimes employed.

In SILAC, proteins made by the cell and secreted into the media will contain labels (^{13}C and/or ^{15}N) not present in FBS-derived proteins. Similarly, in Click Chemistry, secreted proteins from the cell will expectedly contain either an azide or an alkyne group (depending on the labeling agent used) that should be absent in serum-derived proteins. Most labeled secretome studies also use dialyzed serum, which consists of serum in which small molecules particularly amino acids together with hormones and cytokines have been significantly depleted. Depletion of amino acids ensures that protein production in cells only utilizes supplemented labeled amino acids used.

The use of dialyzed serum and/or serum-free media may present biological issues in proteomics studies. It is believed that cell culture in serum-free media for extended periods may result in changes in cell protein dynamics as well as apoptosis, senescence, or autophagy. Where apoptosis is induced by the use of serum-free media, secretion of

proteins from the lysed or dead cells may contribute to the secretome measured. To ensure that the contribution of apoptotic cells to the secretome is insignificant or minimal, we and other labs conduct apoptosis assays to determine the percentage of apoptotic cells following secretome collection. Using the apoptosis assays, different labs may decide how long it is appropriate to culture cells in serum-free conditions while maintaining the integrity of the secretome. Some groups have used very low serum (~1% FBS) to avoid substantial cell death in their secretome studies.

Another issue often encountered in secretome studies is the lack of proteins with equal rates of secretion in "all" cells to be used for normalization of samples similar to the housekeeping proteins used in Western Blotting. This is important especially following identification of candidate potential biomarkers from secretome proteomics experiments where further tests using Western blots, ELISAs etc. are warranted [103].

1.6 Protein Glycosylation

Glycosylation refers to the addition of glycan groups to different residues on proteins through the activity of glycosyltransferases [104]. Typically, glycosylation occurs in the ER/Golgi but it has been known to sometimes take place in the cytoplasm or nucleus [105]. Glycosylation is the most predominant protein PTM and has been demonstrated to have very pertinent effects on protein function (e.g. receptor interactions, immune response mechanisms, secretion, and transport) and physicochemical properties (e.g. solubility, stability, and folding) [104, 106, 107]. For instance, glycoproteins with a mannose-6-phosphate group are known to be targeted to lysosomes and one such glycoprotein, tripeptidyl-peptidase 1, is implicated in classical late-infantile neuronal ceroid lipofuscinoses (LINCL)

[96, 108]. In addition, cancers have been found to commonly present with changes in glycosylation [105].

Up to 50% of all proteins are estimated to have one or more glycoforms particularly proteins that are secreted or targeted to the cell surface [105, 109]. Among the notable plasma proteins known to be glycosylated are antithrombin III, fibrinogen, alpha-1-acid glycoprotein, alpha-1-antitrypsin, apolipoproteins B-100, D, and F [106, 110]. The abundance of glycoproteins coupled with the many roles they play in cellular function and disease makes glycoproteins well suited for investigation as potential biomarkers of early diagnosis or treatment prognosis [111].

Glycoproteins may have more than one site of glycosylation and that site may be occupied by more than one glycan resulting in different glycoforms of the protein. While glycosylation patterns may differ between people depending on genetic differences, diet, disease, and lifestyle, it has proven to be very stable in individuals [106, 111, 112]. Additionally, age and sex have been shown to correlate with the glycosylation status such as the increased bisection and decreased galactosylation of immunoglobulin G (IgG) with age [106, 113].

The central dogma of molecular biology is the formation of RNA from DNA and the subsequent formation of protein from the RNA template. While this process of protein manufacture in cells follows a template (i.e. DNA), glycosylation is very complex and not template-driven, making glycoprotein analysis such as sequence determination a huge analytical challenge.

1.6.1 Glycans

Glycans are carbohydrate moieties typically found in the body attached to proteins or lipids, as mono-, di-, or oligosaccharides, but they may also exist as free molecules. The molecule to which the glycan (glycone) is bonded is referred to as an aglycone. Glycans have been known to be involved in different cellular functions facilitating cell-cell, cell-matrix, or cell-molecule interactions such as pathogen recognition and inflammation [111, 114].

Glycans exist as polyhydroxy molecules of aldehydes or ketones and may be monosaccharides referring to the simplest glycan consisting of only one sugar unit (e.g. glucose shown in **Figure 6**) or oligosaccharides comprising many sugar units. Multiple oligosaccharide units bonded together are referred to as polysaccharides. An aldose refers to a monosaccharide with an aldehyde group whereas one with a ketone group is referred to as a ketose. In the free state, monosaccharides exist as ring or open chain

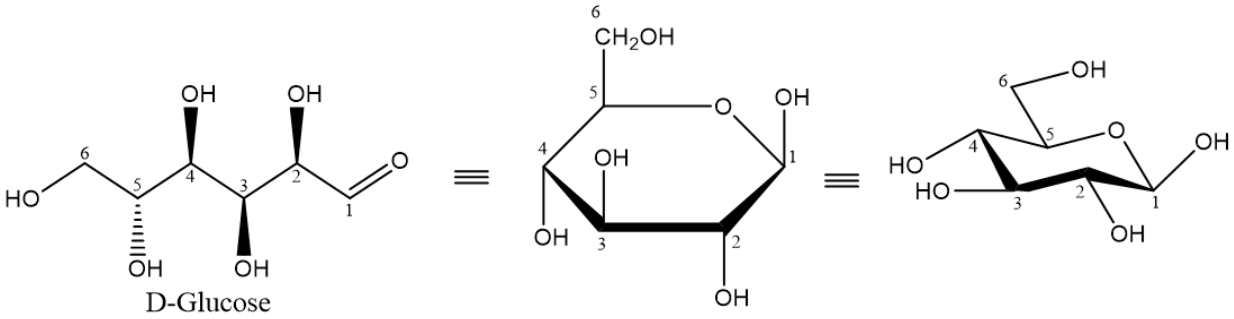


Figure 6 - Different forms of D-glucose

The image on the left panel is the open form of D-glucose whereas the middle image represents the ring form. The ring form may take two conformations – the boat or the chair conformation (shown in the image on the right panel). Substitution of the hydroxyl groups around the structure result in different molecules with distinct properties.

molecules whereas oligosaccharides exist primarily as ringed structures.

Glycans have two ends – the reducing end containing the aldehyde or ketone functional group and the opposite non-reducing end. Single glycan units may be bonded to each

other via two similar but distinct linkages – α or β linkages. The two linkages have been demonstrated, as in the case of starch (α1-4 linkages) and cellulose (β1-4 linkages), to result in different structural and biological properties even though they are both polymers of glucose [114]. Glycans are highly complex and diverse species in that the monosaccharides may be linked to each other via several different hydroxyl groups in either the α or β conformation. For instance, three six-carbon monosaccharide units (hexoses) may be linked to each other in up to 27,648 different possible ways compared to the six tripeptides or trinucleotides possible from three amino acids or nucleotides respectively [114]. Increasing the number of units of a polysaccharide unit increases the possible number of products geometrically e.g. six hexose units can form > 1 trillion possible hexasaccharide structures [114, 115].

The most commonly occurring monosaccharide units are five-carbon (e.g. xylose Xyl) and six-carbon (e.g. glucose Glc, galactose Gal, and mannose Man) neutral or non-neutral species [114]. Glucosamine and galactosamine are six-carbon units with the hydroxyl group at position 2 substituted with an amino (NH_2-) group. The amino group may be acetylated to give hexosamines such as N-acetylglucosamine (GlcNAc) and N-acetylgalactosamine (GalNAc). The hydroxyl group at position 6 may be removed to give a deoxyhexose such as fucose (Fuc) or replaced with a carboxylate (COO^-) group to give uronic acids such as glucuronic acid (GlcA) and iduronic acid (IdoA). Sialic acids (Sia) such as N-acetylneuraminic acid (Neu5Ac) and N-glycolylneuraminic acid (Neu5Gc) comprising nine carbon atoms are also common in nature.

Glycans are generally classified into four major types namely O-linked glycans, N-linked glycans, glycolipids/glycosphingolipids, and glycosaminoglycans, depending on the type of linkage to the aglycone [105].

1.6.1.1 O-Linked Glycans

O-linked glycans are formed by the glycosidic linkage of glycans to the oxygen of the hydroxyl group on serine (S) or threonine (T) residues. The process of O-glycosylation begins in the Golgi apparatus. It is the predominant form of glycosylation found in mucins in the body. The glycan is covalently bonded through an α-linkage to the protein by N-acetylgalactosamine. The resulting product is referred to as a mucin-O-glycan or O-GalNAc [105, 116]. Mucins are found in secretions of mucosa in different epithelial cell surfaces and body fluids. O-glycans may also be bonded to the aglycone via a β-linkage using N-acetylglucosamine resulting in O-GlcNAc [105]. The GalNAc residue is extended by different monosaccharide units including Gal, Fuc, or Sia but not Glc, Man, or Xyl residues [116].

While there is no consensus motif for O-glycosylation, the presence of a proline at the N-terminal side of the S/T group or three amino acid residues away from the S/T group has been shown to favor O-linked glycan formation.

1.6.1.2 N-Linked Glycans

N-glycoproteins are formed from the covalent bonding of an N-glycan group to an asparagine (N) residue of a polypeptide or protein [117]. Studies have revealed an Asn-X-Ser/Thr (N-X!P-S/T) consensus sequence that is most commonly found at N-glycosylated asparagine sites **(Figure 7)** in a polypeptide chain where X is any amino acid except proline. Other sequences such as the N-X-C sequon have been observed in nature to be N-glycosylated [117]. It must be noted however, that there are several N-X-S/T sequons in the human proteome that have not been shown to harbor a glycan.

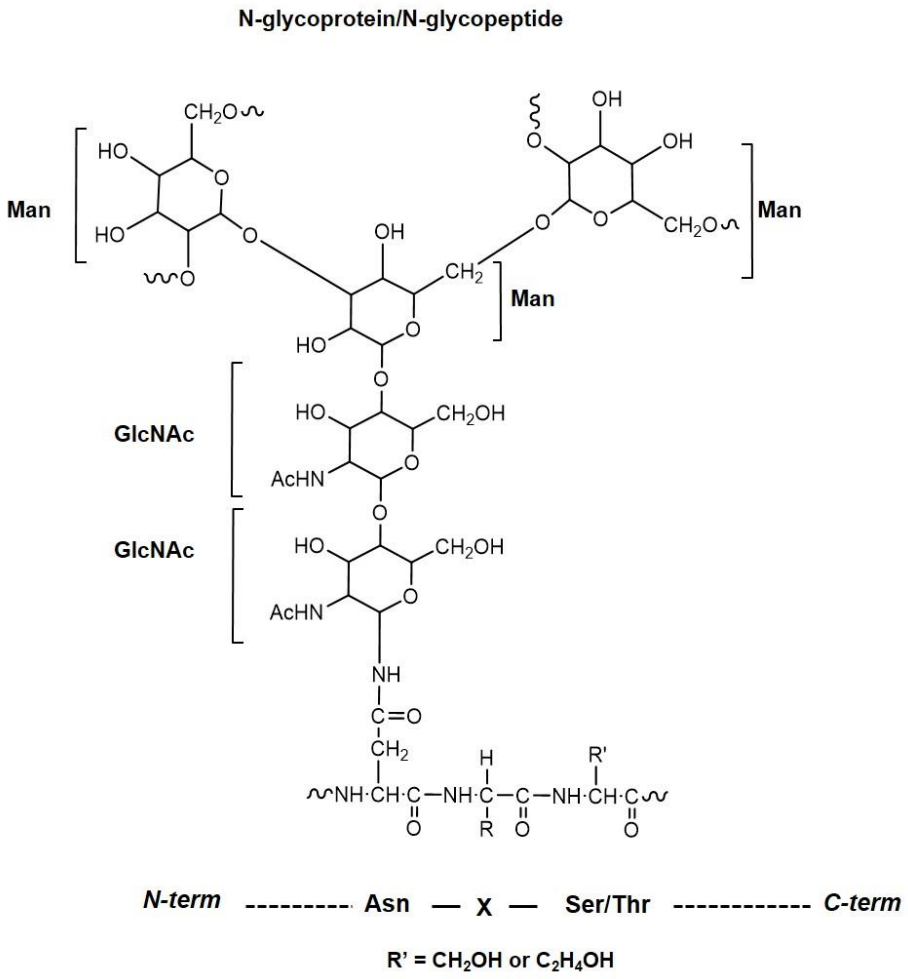
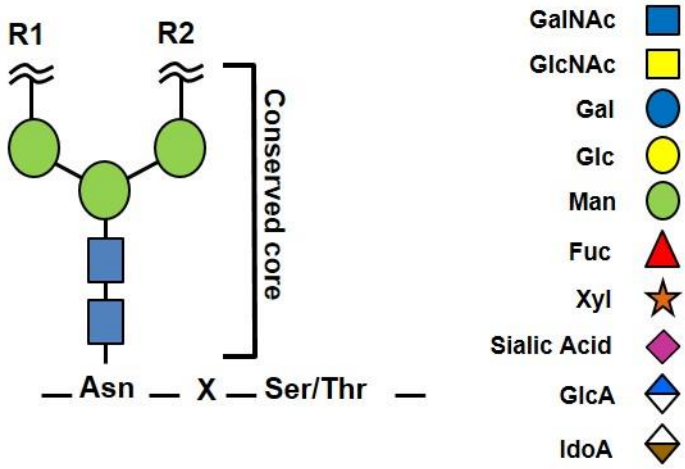


Figure 7 - N-linked glycans

About five different monosaccharide molecules may be found linked to the asparagine residue but the most common group is GlcNAc. The formation of N-glycans has been shown to begin with dolichol phosphate (Dol-P), a lipid polymer of five-carbon isoprene molecules. The entire oligosaccharide glycan molecule is synthesized on Dol-P and then transferred onto the aglycone unit. Synthesis of the oligomannose begins on the



	R1	R2
Oligomannose	Mannose residues	Mannose residues
Complex	Combination of hexoses, hexosamines, uronic acids, sialic acids	Combination of hexoses, hexosamines, uronic acids, sialic acids
Hybrid	Combination of hexoses, hexosamines, uronic acids, sialic acids	Mannose residues

Figure 8 - Types of N-glycans

cytoplasmic side of the ER membrane where Dol-P accepts an N-acetylglucosamine-phosphate (GlcNAc-P) group from UDP-GlcNAc to form Dol-P-P-GlcNAc catalyzed by GlcNAc-1-phosphotransferase [117]. Successive additions of GlcNAc and Man residues results in the formation of a pentasaccharide core that is common to all N-glycans. The common core is typically made up of two GlcNAc units and three mannose sugars [111].

The addition reactions are catalyzed by different saccharyltransferases. The N-glycan molecule is then translocated to the lumenal side of the ER where additional monosaccharide units may be added before transfer of the glycan molecule onto the protein structure occurs while ribosomal translation is taking place. Importantly, Tunicamycin, a nucleosidic antibiotic, possesses inhibitory activity against GlcNAc-1-phosphotransferase and has been used in studies to inhibit protein glycosylation [107].

As depicted in **Figure 8**, N-glycans may be broadly divided into three types. These are 1) oligomannose consisting only of mannose units joined to the common core, 2) complex N-glycans consisting of two to three antennae starting with GlcNAc joined to the common core plus different monosaccharides including sialic acids, hexoses, or other hexosamines, and 3) hybrid N-glycans consisting of two or more "antennae" with one antenna having only mannose units [111, 117].

1.6.2 Mass Spectrometry Analysis of Glycoproteins

The importance of protein glycosylation on cellular development and function cannot be overstated. From protein folding to cell proliferation and signaling, glycans have proven to be pertinent mediators of biological function. Furthermore, their role in cancer can means they can serve as biomarkers of disease where they can facilitate early disease diagnosis and treatment prognosis. Known to be influenced by age and being involved in the etiology of such diseases as Alzheimer's, Huntington's, and some autoimmune conditions, glycans may be used for the detection of these age-related diseases [111, 118].

Mass spectrometry-based analysis has been instrumental in the recent rise in interest in the study of protein glycosylation and glycan differences. Together, the lack of standards,

high variability between different subjects, vast number of possible isomers per glycan structure per glycosite, and the paucity of automated tools/software for interpreting complex MS/MS data from fragmentation of glycoproteins/glycopeptides make the identification and quantification of glycans for use as disease biomarkers or therapeutic targets an enormous analytical challenge. The absence of chromophores or fluorophores also means that mass spectrometry represents one of the best analytical tools for glycan analysis [115].

MALDI-MS and ESI-MS are the two most commonly used MS platforms for glycan analysis. Fragmentation events in MS/MS analysis allows for structural elucidation of glycopeptides/glycoproteins. In ESI-MS, the separation method chosen may be reverse phase, normal phase, hydrophilic interaction (HILIC), porous graphitized carbon (PGC) chromatography, or capillary electrophoresis. In MALDI-MS analyses of glycans, labile sialic acid residues may be cleaved off making derivatization in order to stabilize the monosaccharide units essential [111].

1.6.2.1 Derivatization of Glycans for MS Analysis

The standard modus operandi in glycoprotein (glycan) analysis by mass spectrometry involves the enzymatic (N-glycans) or chemical (O-glycans) release of the glycan group, identification of the site of glycosylation via an analysis of the deglycosylated protein/peptide, and identification and quantification of the released glycans. Released glycans may be derivatized to make them more amenable to commonly used hyphenated chromatographic and mass spectrometry techniques. Glycan derivatization, extensively reviewed by David Harvey [115], may be done by reactions with the hydroxyl group (e.g.

permethylation, acetylation), reducing end (e.g. reductive amination, hydrazone formation), or sialic acid residues. Permethylation is the most popular derivatization approach involving the complete or near-complete addition of methyl groups to all hydroxyl and where present carboxylic acid groups [115]. The INLIGHT glycan tagging approach developed by David Muddiman and coworkers at the North Carolina State University is an example of the hydrazone formation strategy [119].

1.6.2.2 Enrichment Strategies

In order to increase glycoprotein and glycan identification and quantification, different enrichment strategies may be employed. Different strategies including immunoprecipitation using antibodies, covalent hydrazide derivatization, and lectin affinity binding exist for enriching glycoproteins/glycopeptides [118]. The most popular enrichment strategy currently used in glycoprotein/glycopeptides analysis is however the lectin enrichment approach and its applicability has been demonstrated in analysis of lung, breast, and liver cancers [120–122]. In hydrazide-based enrichment, glycoproteins/glycopeptides are immobilized by reacting the reducing end of the glycan (aldehyde) with a solid hydrazide support ensuring that only glycosylated proteins are retained on the column. Peptide-N-Glycosidase F (PNGase F) may then be used to cleave the glycan off and the deglycosylated protein analyzed [118].

Lectins are proteins known to have high affinities for carbohydrates or glycan species. This property of lectins is therefore exploited in the enrichment of glycoproteins in complex protein mixtures. A wide variety of lectins have been isolated from many different sources including mushrooms, legumes like beans and peanuts, garden snails, potatoes,

wheat, and tomatoes among others. Concanavalin A (Con A), wheat germ agglutinin (WGA), and *Ricinus communis* agglutinin (RCA) represent some of the more commonly used lectins in glycoproteomics studies.

Con A and WGA have been demonstrated to have higher affinities for certain glycan units more than others. Con A is generally associated with higher affinities for the high mannose N-glycan type whereas WGA preferentially binds glycoproteins with sialic acid residues (e.g. N-acetylneuraminic acid) and N-acetylglucosamine residues [81, 123]. Hydrogen bonding and hydrophilic interactions between lectins and glycans bound to glycoproteins/glycopeptides accounts for the selective enrichment of glycoproteins/glycopeptides using lectins.

1.7 Dissertation Objectives

This dissertation comprises two projects dealing with the use of mass spectrometry-based proteomics to analyze secreted proteins. The first project is a clinical applications study in pursuit of candidate biomarkers for diagnostic or prognostic use in NSCLC. The second project discussed in this dissertation is a more fundamental proteomics study in which we carry out a comprehensive characterization of glycoproteins in the HepG2 cell secretome.

1.7.1 H1299 Study

NSCLC is responsible for more deaths worldwide than any other form of cancer due to a combination of late stage diagnosis and relapse following chemo/radiation combination therapy. One primary mechanism of post-treatment cancer cell survival is thought to be that of autophagy. Previous studies including studies in the Gewirtz lab have indicated a relationship between p53 status and the functional form of autophagy induced in response

to radiotherapy [124–126]. In tumor cells with functional p53, autophagy has been determined to be cytoprotective in response to ionizing radiation whereas in tumor cells without p53, non-protective autophagy is known to be induced. Where the cytoprotective form of autophagy is present, it can be pharmacologically inhibited to sensitize cancer cells to chemotherapy and ionizing radiation treatment. However, this approach is unlikely to be effective when the autophagy is non-protective and there are currently no available biomarkers to stratify patients according to whether autophagy is protective, non-protective or, in fact, cytotoxic to the tumor cell. This also provides a plausible explanation for the potential failure of clinical trials exploring the inhibition of autophagy as a cancer treatment enhancement strategy as patients could not be stratified based on the nature of the autophagy exhibited by their tumor [127]. In this study, we used p53-null and p53-inducible H1299 NSCLC cell lines as models of non-protective and protective autophagy, respectively, in an effort to identify candidate biomarkers.

The overarching aim of this study is to understand the proteomic signatures involved in radiation-induced autophagy in NSCLC in an effort to identify candidate proteins that may be developed into diagnostic and/or prognostic biomarkers, as well as therapeutic targets. This will help to distinguish between NSCLC patient groups in which autophagy manipulation may or may not be a beneficial therapeutic strategy. The specific aims of this study are:

1. To characterize the response of the H1299 NSCLC cells to ionizing radiation treatment as a function of p53 status
2. To identify the proteomic signatures that differentiate between cytoprotective and non-protective autophagy.

1.7.2 HepG2 Study

One of the limitations of label-free quantitative proteomics is the sensitivity to pre-analytical variability, which affects sample-to-sample reproducibility and quantitative precision. The SILAC approach, initially developed as a quantitative strategy for cell culture analysis, was recently extended to clinical samples by Matthias Mann and colleagues and other groups including our lab [80, 82, 128, 129]. In this extended application, a library of labeled intact proteins pooled from different samples are used as spike-in internal standards to analyze clinical samples including tissue biopsies and plasma. With plasma representing the primary, and perhaps most important, clinical biospecimen [130], our previously published proof-of-principle study [82] demonstrated the use of the HepG2 cell line to produce labeled plasma proteins that can then be spiked into plasma for comparative quantitative analysis. The overall goal of this research is therefore to develop the HepG2 cell secretome as an expression platform for the generation of a library of intact stable isotope labeled proteins to serve as internal standards in quantitative proteomics applications. To do this effectively, it is essential to characterize the SILAC labeled HepG2 secretome especially relating to plasma protein glycosylation.

The specific aims of this study include:

1. Quantitative analysis of temporal plasma protein secretion in HepG2 cells.
2. Qualitative and quantitative analysis of plasma protein glycosylation in the HepG2 secretome.
3. Comparative plasma proteomics analysis of HepG2 secretome spiked plasma.

Chapter 2:
Proteomics Insights into Autophagy

This chapter is drawn from a published review article by *Cudjoe EK et al. 2017*

2.1 Introduction

Autophagy is one of the most widely studied cellular mechanisms owing to the myriad of roles it plays in the cell. Dr. Ohsumi Yoshinori, one of the pioneers in the study of autophagy, recently received the 2016 Nobel Prize in Physiology or Medicine for his seminal role and contributions to the discovery of major signaling and molecular pathways in autophagy [131]. Autophagy, as the name implies, is a mechanism whereby cells "eat" or digest the components of their cytoplasm by presenting these subcellular organelles to lysosomes for degradation by hydrolases [132–136]. The term was coined by Christian de Duve and the process is a highly conserved mechanism observed in single and multicellular eukaryotes [137]. Autophagy occurs in cells at basal levels and in response to various stimuli, notably stress triggers such as starvation, chemotherapy, and radiotherapy. As a basal cellular mechanism, autophagy plays a homeostatic role, maintaining levels of proteins and energy in the cell [134] whereas in stress-induced autophagy the cells use breakdown products for energy and metabolic precursor generation, and survival. The importance of autophagy in both health and disease, has

generated significant interest in defining related signaling mechanisms, sub-cellular location of signaling elements (e.g. proteins), and extracellular signals that affect cell-cell communication. Autophagy represents a primarily homeostatic mechanism. In the process of autophagy, cellular components that may be damaged, dysfunctional, or excessive are captured into double-membrane vesicles known as autophagosomes from where they are transported to lysosomes that degrade the contents of the autophagosome **(Figure 9)**. As a homeostatic mechanism, basal autophagy is involved in the maintenance of cellular integrity and the generation of building blocks of energy and metabolism that facilitate cell survival in response to starvation and/or stress.

As a cellular process, autophagy is primarily classified into macro-autophagy, micro-autophagy, or chaperone-mediated autophagy based on how the substrate cargo is delivered to the degradation machinery [134, 137–139]. Macro-autophagy, ubiquitously used to mean autophagy, refers to the process wherein a double membrane phagophore is formed that sequesters the substrates and matures into the autophagosome. In micro-autophagy, however, the cargo is recognized and sequestered directly by the invagination of the lysosomal membrane for degradation [134]. Chaperone-mediated autophagy (CMA) involves the delivery of selected substrates, soluble proteins, to the lysosome via chaperone proteins. CMA is considered to be very selective whereas micro-autophagy and macro-autophagy may be selective or non-selective [140, 141]. In the context of cancer, different forms of autophagy have been identified and have been classified as either cytoprotective, cytotoxic, cytostatic or non-protective [125, 142–144]. This classification applies primarily to the case of autophagy induced in response to therapy in the tumor cell. Basal autophagy is however, virtually always cytoprotective. Functionally, autophagy

is said to be cytoprotective when its induction, as a response to stress including chemo- and radiotherapy, results in cell survival [125]. Cytotoxic autophagy results in cell death, either directly or by facilitation of apoptosis and is considered to be a desirable outcome of treatment in cancer cells [145]. This form has been referred to as autophagic cell death in some literature and is believed to occur when autophagy is induced to excessive levels [134]. Cytostatic autophagy refers to induced autophagy that results in the arrest of cell

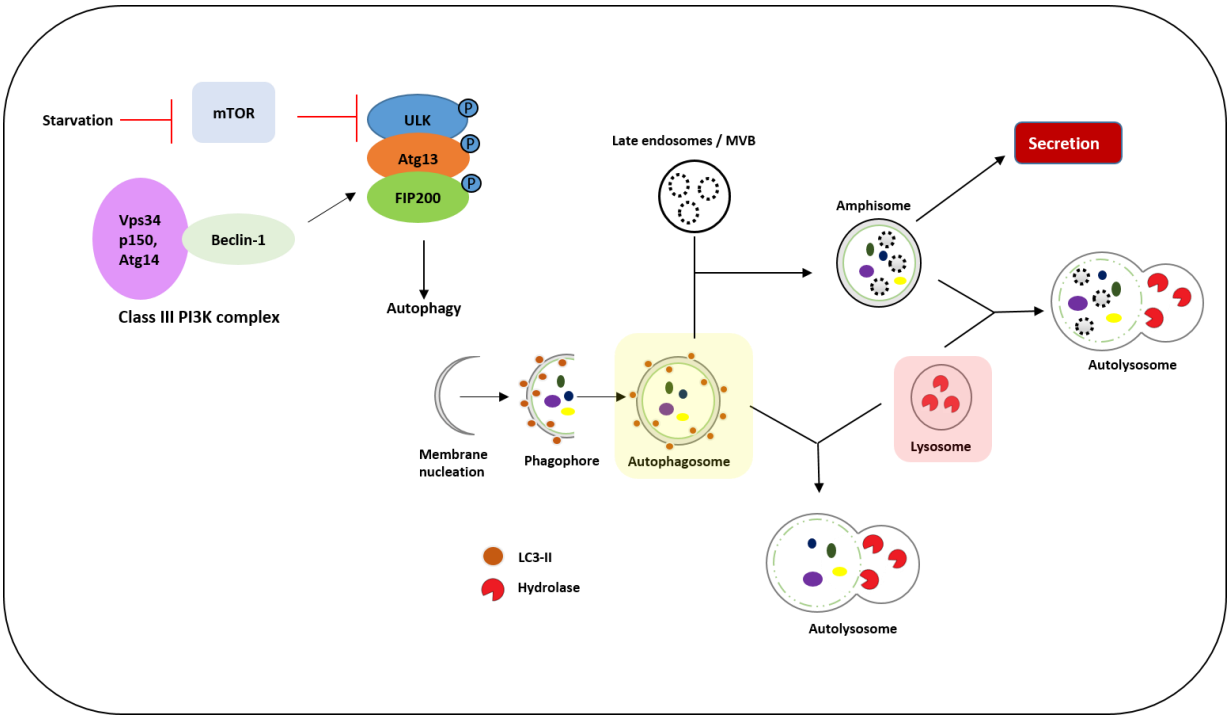


Figure 9 – The Process of Autophagy

The process of autophagy involves the formation of a double membrane structure that matures into the autophagosome and fuses with the lysosome. The autophagosome may also fuse with late endosomes (from the process of endocytosis) into an amphisome that may fuse with a lysosome or result in protein secretion.

growth but not cell death. Currently, more than 36 autophagy-related (Atg) genes have been identified to be involved in some capacity in the autophagic machinery [146]. Many of these genes have been discovered through yeast genetic screening efforts. Being a conserved mechanism, most of the Atg genes discovered in yeast have homologs in

mammalian cells. Autophagy begins with the formation of the double-membrane autophagosome from the phagophore and continues with the fusion of the autophagosome with the lysosome to form the autolysosome. Autophagosome formation, which represents the initiation of autophagy, takes place in four main steps namely induction, nucleation, elongation and completion [141, 147]. Basal autophagy under normal nutrient conditions generally occurs at very low levels, with some exceptions such as pancreatic cancer cells, which tend to have endogenously high levels of autophagy [148, 149]. Stress-induced autophagy is generally observed under nutrient or oxygen starvation or when the cell is challenged by chemotherapy or radiotherapy. In yeast, the target of rapamycin (TOR) is a key autophagy inhibitor and inhibits Atg1, a serine/threonine kinase which is necessary for the formation of the Atg1-Atg13-Atg17 complex [148] under normal nutrient conditions. In mammalian cells, mTOR inhibits the Atg1 homologs, Unc-51-like kinases ULK1 and ULK2. However, during nutrient and amino acid starvation, AMP-activated protein kinase (AMPK) is activated, which inhibits mTOR [147, 148]. It should be noted that autophagy induction by rapamycin (dependent on mTOR signaling) is not the exclusive pathway for this process. As reported by Sarkar et al. there are a number of small molecule enhancers (SMERs) and inhibitors (SMIRs) of the cytostatic effects of rapamycin that may induce autophagy independently of rapamycin and mTOR [147, 150]. This induction step results in the activation of ULK1 and ULK2, which are known to phosphorylate Atg13 and FIP200, the mammalian homolog of yeast Atg17. A ULK-Atg13-FIP200 complex localized to the phagophore is then formed; Atg 101 is believed to bind to and stabilize Atg13 in the complex [148, 151, 152].

Mercer et al. found that Atg101, a novel autophagy protein, also interacts with ULK1 in an Atg13-dependent manner and is a part of the ULK-Atg13-FIP200 complex [151]. This marks the nucleation step, which requires the phosphatidylinositol 3-kinase (PtdIns3K) complex comprising vacuolar sorting protein 34 (Vps34), p150, Atg14, and Beclin-1 [147, 148]. Vps34, the only class III PI3 kinase in mammalian cells, is responsible for the phosphorylation of phosphatidylinositol to phosphatidylinositol-3-phosphate (PtdIns3P) [147, 153]. This phosphorylation process is facilitated by Beclin-1. Several proteins bind to Beclin-1 and may result in either promotion or inhibition of autophagy. Among the Beclin-1 binding partners that promote autophagy are Atg14L and the UV radiation resistance-associated gene (UVRAG) whereas binding to Bcl-2 and Bcl-XL results in inhibition of autophagy [147].

Conjugation of two ubiquitin-like (Ubl) protein complexes – Atg5-Atg12-Atg16 and Atg8–Phosphatidylethanolamine mediates the elongation and completion phases of autophagy which involves Atg3, Atg5, Atg7, Atg10, Atg12 and Atg16L as well as microtubule-associated protein light chain 3 (LC3) [141, 148, 154, 155]. Yeasts have only one Atg8 protein whereas in mammals there are many homologs in the Atg8 family of proteins including LC3, GABARAP, and GATE-16 [155]. The completion of autophagosome formation is associated with the formation of an Atg5-Atg12-Atg16 complex that results in the conversion of cytosolic LC3 to the lipidated membrane-bound isoform, LC3-II [154]. It was previously believed that Atg5 and Atg7 are necessary for autophagy in mammalian cells but an alternative pathway independent of Atg5/Atg7 has been identified [156]. Autophagy enters into completion via the fusion of the autophagosome with the lysosome to form the autolysosome where the cargo (contents of the autophagosome) are broken down by

hydrolases to generate energy, metabolic precursors, or amino acids for protein synthesis.

2.2 Proteomics Approaches to Study Autophagy

Mass spectrometry-based proteomics has proven invaluable for studying protein structure, expression, sub-cellular location, protein-protein interactions and PTMs [157, 158]. In recent years, mass spectrometry-based proteomics technology has played a major role in advancing our knowledge relating to autophagy and increased our understanding of the content of autophagic structures such as autophagosomes and lysosomes. While some of these studies have focused on autophagy-related mechanisms, other studies reflect global proteomics analysis to investigate cells and disease states, which provide insight into known and/or novel functions of various proteins linked to the autophagic machinery. A variety of proteomics approaches, including labeling (e.g. Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC), biorthogonal noncanonical amino acid tagging (BONCAT), and isobaric tagging for relative and absolute quantification (iTRAQ)) and label-free methods, have been used in combination with immuno-based techniques, among others, in exploring autophagy [149, 159–165]. In this review, we will cover proteomics studies focused on autophagy in the context of the cell compartments analyzed and shown in **Figure 10**. A brief summary of each study including the cell and/or animal models, treatment conditions, and analytical approaches used as well as key regulated proteins is provided in **Table 1**.

2.2.1 Whole Cell Proteomics Analysis

Proteomics has been used successfully to screen whole cell lysates for differential protein expression and PTMs as a function of different stress stimuli including drug treatments and radiation [160, 162, 164, 166–170]. Cells exposed to exogenous stimuli are subsequently lysed to obtain the entire complement of proteins that may have been affected, but without regard to cellular compartment or function. In global studies, comparative data analysis reveals the effects of the stimuli on biological function relating to autophagy. Proteomics studies have served to reveal novel proteins or novel functions of known proteins and their involvement in the process of autophagy.

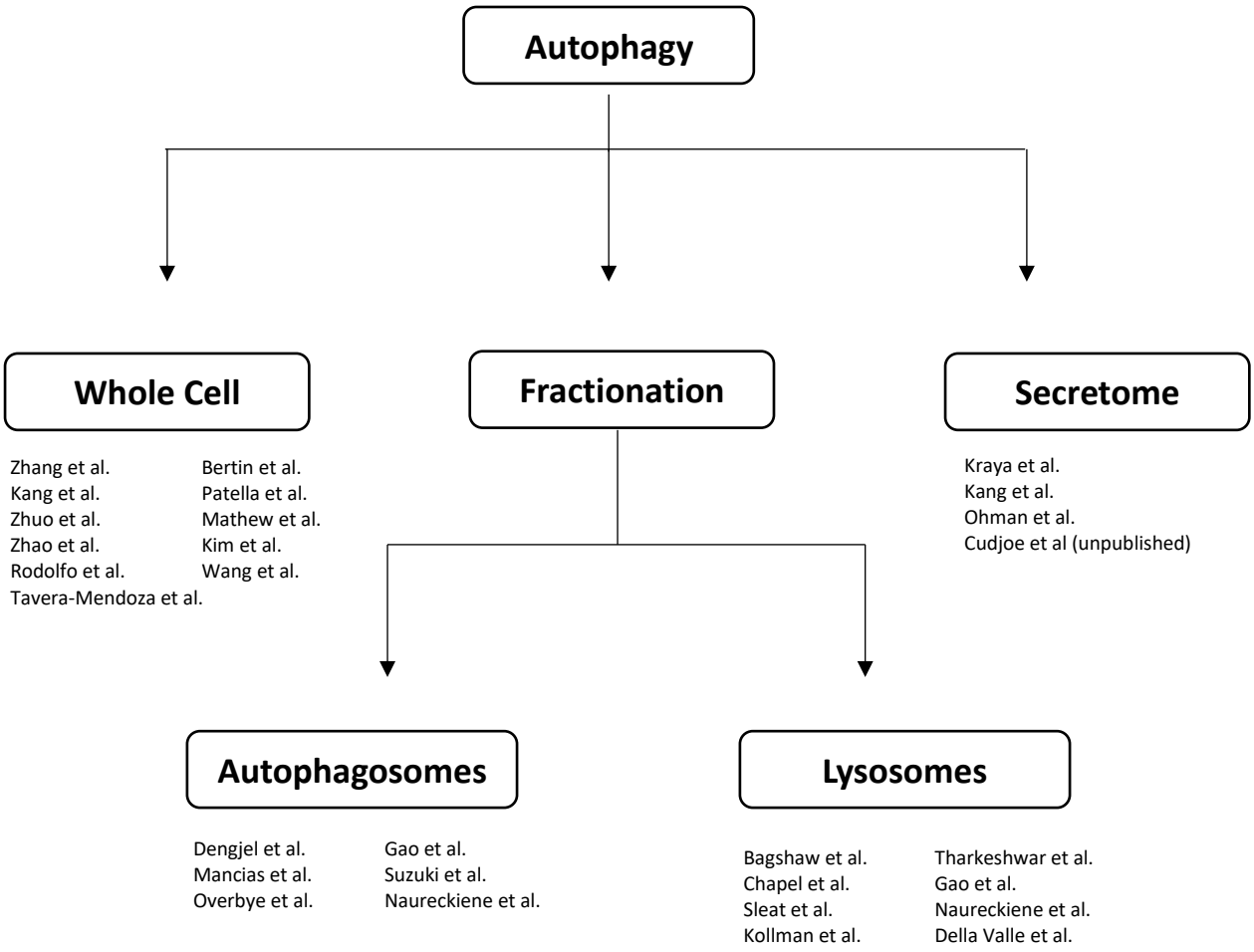


Figure 10 - Overview of proteomics characterization of Autophagy with respect to cellular and sub-cellular compartments.

In two recent publications, Zhang et al describe a proteomics method involving a combination of BONCAT labeling and iTRAQ for the detection and quantification of de novo protein synthesis in autophagy [166, 167]. The BONCAT approach involved introducing azidohomoalanine (AHA), an azide methionine mimetic, into the culture media, which subsequently is incorporated into newly synthesized cellular proteins. Autophagy was induced in AHA-treated HeLa cells via amino acid starvation (in amino acid-free media) followed by total cell lysis and enrichment of newly synthesized AHA-containing proteins via click reaction coupling of biotin-linked alkyne groups and subsequent avidin bead capture. Bound AHA-containing proteins were then digested and treated with iTRAQ reagents followed by LC-MS/MS proteomics identification and quantification. Zhang et al profiled 711 proteins using this method. Among the proteins identified and verified were ATP synthase, H+ transporting, mitochondrial F1 complex, β polypeptide (ATP5B), heat shock protein family E [Hsp10] member 1 (HSPE1) and solute carrier family 25 member 3 (SLC25A3), receptor for activated C kinase 1 (RACK1/GNB2L1) and PNP (purine nucleoside phosphorylase) [166]. Gene knockdown (using siRNA) of ATP5B, RACK1, or SLC25A3 resulted in decreased autophagic flux indicating a role for these proteins in the promotion of autophagy. The approach used in the study is applicable to many multiple experimental designs and can be very useful in understanding protein synthesis during stress-induced autophagy.

Kang et al. performed proteomics analysis on starvation-induced autophagy in HCT116 cells and identified Annexin A1 (ANXA1) as a potential regulator of autophagic degradation [41]. 2D-gel electrophoresis and subsequent MALDI-TOF MS analysis of the proteins identified various proteins that were up or downregulated after autophagy

induction. Kang et al. reasoned that downregulated proteins following autophagy induction may be targets of autophagic degradation whereas upregulated proteins may be involved in the process of autophagy itself. ANXA1 levels increased by ~3 fold following the induction of autophagy and when it was knocked down by siRNA, there was no difference in LC3 degradation between ANXA1 siRNA and wild type cells after starvation. Notably, a modest decrease in p62 levels was observed which the authors argue may be indicative of a later role related to autophagic degradation for ANXA1 [171]. Zhuo et al. in an iTRAQ comparative proteomics study of autophagy in mouse embryonic fibroblasts (MEFs) with and without the autophagy gene Atg7 revealed a connection between F-actin and autophagy [172]. Functional enrichment and network analysis of the results indicated an upregulation of proteins in the F-actin network [172]. Further analysis revealed the F-actin fibers in the Atg7 $^{-/-}$ cells were highly disorganized compared to wild type cells. The functional relevance of F-actin in autophagy was investigated by inducing F-actin depolymerization in Atg7 wild type cells using cytochalasin D (CD) which resulted in increased LC3 conversion and p62 degradation following starvation. However, in CD-treated cells following starvation, more LC3 degradation but decreased p62 degradation was seen implying a decrease in autophagosome degradation [172]. F-actin was therefore shown to be important in the maturation of autophagosomes and completion of the autophagy process.

Toll-like receptors (TLRs) are immune system proteins that recognize certain microbial molecules. TLRs have been shown to be involved in the regulation of autophagy but the role of specific Toll-like receptors such as TLR-9 in autophagy remains controversial [173–175]. Delgado et al. showed, through their comprehensive study of Toll-like receptors, that

CpG – a TLR-9 ligand, did not induce autophagy while Sanjuan et al. showed that CpG activated autophagy [173, 174, 176]. A comparative global proteomics study by Bertin et al., however, showed that CpG is involved in the autophagic process in tumor cells as well as human embryonal kidney HEK293 cells expressing murine TLR9 [175]. In their study, CpG-treated tumor cells showed autophagy induction characterized by LC3I conversion to LC3II. They also observed changes in autophagy regulated proteins such as annexin A1 (ANXA1). In another study, Li et al. investigated changes in the host proteome following infections with the rabies virus in mice brains and determined factors that may be responsible for different pathogenicity [177]. They used two strains of the virus and found one strain, CVS-11, to be more virulent than the SRV9 strain and that SRV9 infection was attenuated between 10 – 14 days post infection. Autophagy, together with associated pathways such as mTOR signaling, was one of the top 5 canonical pathways differentially dysregulated between CVS-11 infection and SRV9 infection in the Li et al. study [177]. Li et al also demonstrated that autophagy induction was not simply a result of viral infection but a function of viral replication. NA cells were infected with UV-inactivated rabies virus and they found no significant induction of LC3 conversion compared to mock-infected cells. Increased autophagosome accumulation was observed in attenuated SRV9-infected samples compared to CVS-11-infected samples. The presence of viral genomic RNA material in the SRV9-infected cells after attenuation of infection when viral titers fell below detectable levels may indicate that autophagy induction is more a response to viral RNA than to viral particles.

Patella et al. sought to identify changes in the proteome of endothelial cells (ECs) associated with the transition from sub-confluence to confluence [178]. Cultured ECs are

motile and proliferate when sub-confluent whereas confluent cells establish cell-cell adhesion and secrete various autocrine and paracrine factors into the extracellular space that facilitate decreased cell proliferation and permeability and promote quiescence [178, 179]. These investigators identified proteins that were differentially regulated in sub-confluent and confluent cells. Most of the proteins upregulated in sub-confluent cells compared to confluent cells were related to cell division and DNA/RNA-processes, suggesting decreased growth in confluent cells, which is consistent with our understanding of the confluent state. The proteins upregulated in confluent cells were mostly enriched for extracellular matrix organization, metabolic processes and vesicle organization. Vacuole organization was the most enriched and within this group, several lysosomal proteins were present. Their results also indicated that p62/SQSTM1, a protein that is generally reduced via autophagic degradation, was downregulated in confluent cells. Patella et al. further sought to identify the impact of autophagy on EC function and found that inhibition of autophagy via Atg5 silencing resulted in decreased cell proliferation while both Atg5 silencing and bafilomycin treatment impaired EC barrier integrity and increased EC permeability [178]. Pathway analysis was used to predict the factors driving the global changes in proteome induced by autophagy inhibition. Among the identified factors was hydrogen peroxide, a widely known inducer of reactive oxygen species (ROS). In autophagy-deficient cells, Rho-related GTP-binding protein (RHOB), which has been shown to be upregulated in response to oxidative stress in pulmonary endothelial cells, was found to be upregulated following autophagy inhibition suggesting that autophagy inhibition results in oxidative stress [178]. Treatment of bafilomycin-treated cells with the ROS scavengers, N-acetyl-L-cysteine and ascorbic acid, resulted in a partial

restoration of EC barrier integrity. This showed that the increased EC permeability following autophagy inhibition was a consequence of ROS generation.

Mathew et al. carried out a comparative proteomics study to identify differences in the proteomes of autophagy-proficient and –deficient Ras-driven cancer cells [162]. Autophagy was induced in SILAC-labeled baby mouse epithelial (iBMK) cells via starvation followed by LC-MS/MS analysis. A total of 3181 proteins were found to be differentially expressed before and after starvation in the autophagy-proficient and –deficient cell comparison. This finding reflects the extent of protein turnover during autophagy and further provides insight into proteins that serve as cargo for the autophagic machinery. Autophagy was also determined to be largely targeted to proteins that promote cell survival. In autophagy-deficient cells, poly(ADP-ribose) polymerase (PARP), known to be involved in single-strand DNA repair, and STAT1, known to be involved in interferon-mediated cell death, were upregulated [162].

The EVA1A gene is necessary for autophagosome formation during autophagy via its interaction with Atg16L [180] and it is also reported to induce cell death [181]. Proteomics studies have subsequently revealed that EVA1A regulates genes that are involved in the pathogenesis of neurodegenerative diseases including Alzheimer's and Huntington's, thereby shedding some light on the potential involvement of autophagy in neurodegeneration [164]. Autophagy has also been determined to be involved in the metabolism of amyloid beta as well as its secretion [182]. In the study by Zhong et al, EVA1A homozygous knockout mice ($Eva1a^{-/-}$) showed higher levels of neurodegeneration compared to EVA1A wild type mice [164]. This was detected by determining both mRNA and protein levels of the neuronal marker, β-tubulin III, which

showed a marked decrease in knockout mice. Fluorescence detection of β-tubulin III showed a similar result. The effect of EVA1A deficiency was only apparent in neurons but not other cell types. The proteomics analysis identified 5438 proteins of which 4462 were quantified in all three replicates. 28 proteins were differentially regulated by EVA1A knockdown and pathway analysis determined an enrichment of neurodegenerative disease-relevant proteins [164].

Further insight into the actions of certain drugs using proteomics tools have revealed the central role played by autophagy in drug sensitivity. Zhao et al. found that the neurotoxicity of bupivacaine is strongly linked to the generation of excess reactive oxygen species and activation of autophagy. It is also related to its inhibitory effects on PI3K [160]. The phosphoinositide-3-kinases (PI3Ks) represent one of the more important family of genes involved in autophagy. The class 1 PI3Ks are known to be inhibitory while the class III PI3Ks are known to be stimulatory in autophagy [148]. A phosphoproteomics study was carried out by Zhang et al. using 2 lung cancer cell lines H3255 and H1975 [168]. The H3255 cells carry a single mutation in the EGFR gene which makes them susceptible to EGFR-targeted tyrosine kinase inhibitors (TKIs) such as erlotinib while the H1975 cells carry a double mutation in the EGFR gene which makes them resistant to erlotinib. Zhang et al. found that in erlotinib sensitive cells but not erlotinib resistant cells, the key autophagy signaling protein ULK1, is not phosphorylated upon treatment. This implies that the sensitivity of the lung adenocarcinoma cells used in the study to erlotinib may be autophagy dependent [168]. The phosphorylation action of kinases is known to be regulated in the body by ATP binding and hydrolysis [183, 184]. Kinases are also thought to be involved in the ability of cancer cells to evade chemotherapy. Kim et al. therefore

conducted a proteomics study to determine the effects of MEK-inhibitor drug treatment on the ATP-binding proteome in KRAS mutant lung cancer cell lines with different p53/LKB1 mutation statuses including A427, A549 (p53 wild type/LKB1 mutant), Calu-1, Calu-6 (p53 mutant/LKB1 wild type); and H157 (p53 mutant/LKB1 mutant). Cells were treated with 2 MEK-inhibitors, lysed, and the ATP-binding kinome enriched, using a kinase enrichment kit, for subsequent digestion and LC-MS/MS analysis. The study found that changes in the ATP-binding proteome were cell type dependent and not KRAS mutation dependent [184]. Furthermore, p53/LKB1 status was not determined to be a major factor in changes in the ATP-binding proteome. MEK-inhibition was also determined to effect changes in glucose metabolism in A427, A549, and Calu-1 cells. Kim et al. found that ULK1, ULK3 and AMPK were upregulated in cells treated with MEK-inhibitors compared to untreated cells. The findings of the study by Kim et al. align with MEK-inhibitors having been previously reported as inducing protective autophagy in KRAS mutant non-small cell lung cancer (NSCLC), and suggests that upregulation of autophagy in KRAS mutant lung cancer may be kinase dependent [184, 185]. When validated, this may provide a reasonable basis for combination chemotherapy with kinase inhibitors.

Some anticancer agents including Ophiobolin A (OP-A), a fungal toxin, have been determined to induce autophagy [169]. Rodolfo et al. investigated the anticancer activity of OP-A using a human melanoma cell model (A375, and CHL-1 cell lines) [169]. A comparative proteomics experiment was used to determine differentially regulated proteins due to OP-A treatment. In the study, OP-A induced autophagy based on increased LC3 conversion with increasing drug concentrations as well as activating apoptosis. Following 2D gel electrophoresis, Rodolfo et al. used LC-MS/MS to identify 24

protein spots that were significantly affected by OP-A treatment. Among the differentially regulated proteins were fructose 1,6 bisphosphate aldolase A (ALDOA) and triose phosphate isomerase (TPI) [169]. These enzymes, found to be associated with autophagy, have been shown to be important enzymes in glycolysis in a cell type dependent manner [186, 187]. Wang et al [170] used proteomics to identify potential targets of the natural product curcumin in HCT116 colon cancer cells via an alkyne-labelled curcumin probe. The curcumin bound-proteins were biotinylated via azide-linked biotin tags, enriched with avidin beads, digested with trypsin and then labeled with iTRAQ tags, and analyzed by LC-MS/MS. Wang et al. identified 197 proteins with unique cellular locations and biological functions [170]. Among the biological pathways enriched in their data were eIF2 signaling, eIF4 and p70S6K signaling, and the mTOR signaling pathway which is directly related to autophagy. Wang et al. went on to investigate curcumin treatment effect on protein synthesis using AHA. Newly synthesized AHA-containing proteins were fluorescent labeled via a click reaction and the fluorescence intensity of cells analyzed with flow cytometry. Curcumin treatment resulted in a decrease in protein synthesis evidenced by a 50% reduction in fluorescence signal following treatment. Previous reports in literature [188] that curcumin may inhibit mTOR phosphorylation led Wang et al. to determine the effect of curcumin on autophagy. The results of their immunoblotting work confirmed that curcumin induces autophagy, which was signaled by an increased autophagic flux and increased LC3 levels. The inhibitory effect of curcumin on mTOR was also confirmed via decreased phosphorylation of the mTOR substrate S6 [170].

Vitamin D, acting through activation of the vitamin D receptor (VDR), has been determined to induce autophagy and have anti-proliferative activity [189]. A study by our

group demonstrated a novel switch in autophagy function from a cytoprotective to a cytostatic form following vitamin D combination treatment with irradiation in non-small cell lung cancer cells [126]. A recent proteomics study by Tavera-Mendoza et al. also showed autophagy induction by vitamin D in luminal-like breast cancer cells as well as anti-proliferative effects [190]. Tavera-Mendoza et al. used rapid immunoprecipitation mass spectrometry (RIME) to identify proteins associated with DNA-bound VDR. The RIME process followed the procedures reported by Mohammed et al [191]. Briefly, cells are cultured, media exchanged and crosslinked in 1% formaldehyde following which proteins are extracted and the lysate incubated with magnetic beads prebound with antibody for overnight immunoprecipitation. Beads are then washed and digested for LC-MS analysis. The RIME analysis by Tavera-Mendoza et al. revealed a regulatory role of vitamin D-bound VDR for autophagy in breast cancer cells. VDR was shown to downregulate autophagy via constitutive LC3B repression and this repression appears to be abrogated by vitamin D treatment. This abrogation of LC3B repression, together with the upregulation of transcription of autophagy-related genes defines the regulatory role of vitamin D in autophagy in the breast cancer model used [190].

2.2.2 Subcellular Fractionation

Proteomics studies involving subcellular fractionation of various organelles is increasing in popularity and application. These approaches can provide critical information relating to the biological functions of these organelles in the cell and the roles played by specific proteins. While this research strategy is useful by affording a spatiotemporal subset of the proteome and a less complex proteome, it suffers from the cross contamination between subcellular compartments during sample preparation. The subcellular fractionation

methods typically employ differential gradient centrifugation in sucrose or Nycodenz gradients and/or affinity-based immunocapture or enrichment.

2.2.2.1 Autophagosomes

In an effort to provide comprehensive insight as to the types and amounts of proteins associated with the autophagosome, a limited number of studies have isolated and characterized the proteome of the autophagosome [149, 159, 192, 193]. Isolation of autophagosomes have been carried out using Nycodenz gradient centrifugation, GFP-tagged LC3 immunoprecipitation or immunoisolation via GFP antibody purification or a combination of the two. One of the earliest reported methods of autophagosome isolation is the study by Stromhaug et al. and forms the basis of most subsequent proteomics autophagosome isolation studies [194].

Dengjel et al. sought to generate a comprehensive list of various proteins contained in autophagosomes while demonstrating that different inducers of autophagy produce different protein dynamics within the autophagosome [159]. The study employed the protein correlation profile (PCP) method that involves the analysis of peptide contents in the various fractions of a density gradient. The PCP strategy assumes that proteins/peptides from the same organelle would have a largely similar profile across the various fractions [157, 159]. In this work, autophagy was induced via amino acid starvation using Hank's Balanced Salt Solution (HBSS), rapamycin treatment, and concanamycin A (Con A - a lysosomal degradation inhibitor) treatment. Dengjel et al. identified 728 enriched proteins in their autophagosomal fractions, 94 of which were present in the context of all three autophagy stimuli. In contrast, proteomics studies of the autophagosome by Overbye et

al. and Gao et al. did not generate many autophagosome-enriched proteins; 39 and 101 respectively [192, 193]. There was little overlap between the 3 studies which could be due to the fact that the experimental approaches adopted by Overbye et al. and Gao et al. were limited to autophagosomal membrane proteins while the Dengjel et al. study analyzed the entire autophagosome [159]. Also, the Overbye study was conducted in primary hepatocytes isolated from Male Wistar rats and autophagy induction was done via starvation while the Gao study was in HEK293 and HCT116 cells in which autophagy was induced by calcium phosphate precipitate (CPP) [192, 193].

To distinguish between proteins truly associated with autophagosomes and non-specific proteins in the Dengjel et al. study, some cells were treated with Con A during starvation. Autophagosome associated proteins would be expected to be enriched in Con A-treated starved cells (due to the accumulation of autophagosomes) compared to starvation only cells. Dengjel et al. showed less enrichment of known autophagy-related proteins LC3B, p62/SQSTM1, GABARAPL2 in starvation-induced autophagosomes signifying higher depletion of these proteins following stimulation of autophagy compared to rapamycin and Con A treatment.

One of the very significant proteomics studies of autophagy in which autophagosomes were isolated was performed by Mancias et al. and resulted in the identification of nuclear receptor coactivator 4 (NCOA4) as a cargo receptor in iron turnover through autophagy [149]. In this work, autophagosomes were isolated from SILAC-labelled PANC-1 and PA-TU-8988T pancreatic cells as well as MCF7 breast cancer cells using a density gradient. Autophagosome accumulation was increased by treatment with chloroquine, an autophagy inhibitor that acts by raising lysosomal pH and thereby preventing fusion of

autophagosomes and lysosomes. This was a similar strategy to the Dengjel study in which autophagosome accumulation was achieved via Con A treatment. The isolation of autophagosomes enabled the unbiased identification of 94 proteins highly associated with autophagosome formation or maturation and by inference with autophagy.

These autophagosome isolation studies have not only provided insights into the protein composition of the autophagosomal membranes and cargo, but have further provided general insights into the process itself. Dengjel et al. also showed, through their mass spectrometry-based proteomics study, a significant association between autophagy and the proteasome and extended their work to probe the relationship using western blotting and fluorescence microscopy [159]. This research group determined that the levels of various proteasomal proteins decreased upon autophagy stimulation via amino acid starvation as well as rapamycin treatment, and this was reversed when autophagy was inhibited by 3-methyladenine (3-MA). Data analysis of the candidate proteins in the Mancias et al. study revealed NCOA4 as the most consistently enriched protein in all their datasets. NCOA4 was also identified in the Dengjel et al. study but was not highlighted as one of their cluster A proteins [149, 159]. The cluster A proteins consisted of the highly confident autophagosomal candidate proteins believed to not be “contaminated” with proteins from similar organelles such as the endoplasmic reticulum, Golgi apparatus and endolysosomes [159]. Mancias et al. found a novel function of NCOA4, where it is required in ferritin turnover in a process they termed ferritinophagy [149]. The mass spectrometry data from their autophagosomal fraction showed increased expression of both ferritin heavy (FTH1) and light (FTL) chains and affinity purification-mass spectrometry showed

NCOA4 to interact with FTH1 and FTL as well as HERC2 and NEURL4. Importantly, the latter two proteins were not found to be enriched in autophagosomes [149].

Other proteomics studies involving isolation of autophagosomes include the work of Suzuki et al [195] where various cargo proteins were identified in the autophagosomes of *Saccharomyces cerevisiae*. Suzuki et al. optimized autophagosome isolation by monitoring GFP-labeled aminopeptidase 1, which represents a selective cargo in the yeast *Saccharomyces cerevisiae.* It appeared that some of the cargo proteins are delivered to the vacuole, where degradation takes place in yeast, independent of the essential autophagy-related protein Atg11 [195].

2.2.2.2 Lysosomes

Lysosomes are single membrane organelles within the cytoplasm involved in one of two main degradation mechanisms employed by eukaryotic cells [157, 196]. The lysosomes play a major role in autophagy by fusing with autophagosomes and supplying hydrolases for the degradation of the inner membrane and contents of the autolysosome. The hydrolases found in the lysosome function at an optimum acidic pH, which may be inhibited with agents that raise luminal pH. Lysosomes also play a role in the endosomal pathway [196]. Lysosomes have been found to affect many human disease pathologies particularly the lysosomal storage diseases (LSDs) but also other diseases including neurodegenerative disorders such as Alzheimer's disease, the neuronal ceroid lipofuscinoses (NCLs) and some cancers [196, 197]. Understandably, many LSDs have also been shown to have attendant defects in the autophagy pathway [198]. To know the protein content of the lysosome would be to understand the possible range of functions the

organelle controls or is associated with in the cell and human body. Studying the lysosomal proteome under different conditions of stress or stimuli may also help to shed some light on the protein dynamics in response to the various stimuli [197]. Using classical biochemical approaches, lysosomal protein content had been studied but the full complement of the lysosome was not known [199]. However, the combination of mass spectrometry-based proteomics, affinity-based purification techniques, and subcellular fractionation has resulted in an increase in the number of proteins identified in the lysosomes [197]. Kieffer-Jaquinod et al. outline an affinity-based method for the purification of lysosomal proteins that involves the isolation of these soluble proteins on the basis of their mannose-6-phosphate content [200]. This isolation strategy has provided the rationale for many studies that have investigated soluble lysosomal proteins [108, 201–203].

Bagshaw et al. conducted a proteomics study to identify the proteins contained in the lysosomal membrane [196]. Lysosomes were isolated by the Triton WR-1339 approach used by Leighton et al [204] and 215 lysosomal membrane proteins were identified, some of which had not previously been reported as being associated with the lysosomes [196]. Cytochrome P450 enzymes such as CYP2A1, CYP2C13, CYP2D3, and CYP4A3 as well as various ATP synthase subunits including the α, f0 β, F1 complex O, and γ chain subunits were identified in the Bagshaw study [196]. Other lysosomal isolation proteomics studies have contributed to what we now know about the protein composition of lysosomes and their functional relevance, including autophagy. Chapel et al. identified 734 proteins in a study in which rat liver lysosomes were isolated using differential centrifugation in a density gradient [205]. Of the identified proteins in the lysosomal fraction, 207 constituted known and predicted lysosome-associated proteins whereas 527 proteins

were either previously reported to be associated with other cellular organelles or not associated with any organelle. A total of 46 potential lysosomal transport proteins were identified, 12 of which were validated as lysosomal proteins via overexpression in HeLa cells and confirmation of colocalization with lysosomal markers [205]. The authors point out that proteins identified in their study that had not been previously reported could be cargo delivered by autophagy or endocytosis to the lysosomes for degradation. Using iTRAQ labeling and two dimensional peptide separation involving strong cation exchange and reverse phase chromatography, Della Valle et al. also report the MALDI-TOF proteomics identification of high confident lysosomal proteins using an isolation strategy involving the combination of differential centrifugation with sucrose density centrifugation following treatment with Triton-WR1339 [206]. Among their high confident proteins were cathepsin D (CTSD), classical lysosomal acid phosphatases (ACP2, ACP5), and lysosomal associated membrane protein 2 (LAMP2).

The identification of a glycoprotein candidate that may be responsible for classical late-infantile NCL (LINCL) was made possible through a proteomics study [108, 197] using gel electrophoresis and affinity chromatography. In this work, where Sleat et al. compared detergent-soluble extracts of brain autopsies from LINCL patients and controls [108], soluble proteins with a mannose-6-phosphate (M6P) post-translational modification (PTM) were detected using mannose-6-phosphate receptors (MPRs). In LINCL brain samples, a protein that was visibly absent compared to controls was identified using affinity enrichment of M6P modified glycoproteins, gel isolation, and protein sequencing as having a molecular weight of 46-kD. The protein was determined to be resistant to pepstatin degradation and to have considerable sequence similarities with carboxyl

peptidases of both *Pseudomonas* and *Xanthomonas* [108]. The study represented the first report of a pepstatin-insensitive protease in mammals. A study by Vines and Warburton revealed substantial similarities between tripeptidyl peptidase I (TPP-I) and the pepstatin-insensitive protease identified in the Sleat et al. work and suggested the two proteins may be the same [207].

NCL has been found to possess many common features with other mitochondrial disorders in which autophagy of the mitochondria, often termed mitophagy, is defective [208]. Other proteomics studies of the lysosome have resulted in the identification of genes that have been found to be pivotal in the etiology of certain LSDs. Human epididymis-specific protein 1 (HE1), a cholesterol-binding protein was found in a proteomics study by Naureckiene et al. to be contained in lysosomes [201]. This discovery, together with prior knowledge of the involvement of this protein in cholesterol binding, led to the investigation of the possibility of its involvement in LSDs that involve lysosomal cholesterol storage [197, 201]. This resulted in the discovery of mutations in the gene in Niemann-Pick disease type C2 (NP-C2) patients [201]. Kollman et al. employed the M6P immunoisolation strategy in conducting a comprehensive proteomics analysis of lysosomal matrix proteins [209]. Using the knowledge that M6P-deficient mouse embryonic fibroblasts secrete the lysosomal matrix proteins that would have otherwise been targeted to the lysosome, Kollman et al. isolated M6P proteins using affinity purification. The secretome of the fibroblasts was then analyzed which resulted in the identification of 34 known and 4 potential lysosomal matrix proteins [209]. Among the proteins identified were the cathepsins B, D, and Z as well as mammalian ependymin-related protein-2 (MERP-2), Protein CREG, and retinoid-inducible serine carboxypeptidase (RISC).

In a recent proteomics study using NP-C1 as a case study, Tharkeshwar et al. demonstrated a novel method of isolating lysosomes for downstream proteomics applications [210]. The method utilizes a magnetic isolation approach with superparamagnetic iron oxide nanoparticles (SPIONs). These SPIONs consist of an organic or inorganic outer shell covering a magnetic core [210]. These investigators first optimized SPION preparation and found that dimercatosuccinic acid (DMSA)-coated SPIONs were most suited to targeting the late endosomes or lysosomes (LE/LYS). The SPION design and targeting efficiency was validated with proteomics and lipidomics profiling of plasma membranes and isolated lysosomes. The lack of proteins from contaminating organelles and abundance of membrane-enclosed luminal proteins was consistent with the isolation of very homogenous fractions of plasma membranes and LE/LYS. Their study also evaluated proteome and lipidome changes in HeLa cells with and without wild-type NP-C1. The NP-C1 protein is primarily located in the membrane of LE/LYS. NP-C1 and NP-C2 regulate the transport of cholesterol from lysosomes [210, 211]. Tharkeshwar et al. isolated plasma membranes and LE/LYS using their SPIONs and compared the protein as well as lipid content. The analysis showed that the plasma membrane proteome and lipidome remained largely the same with minor changes (6 differentially regulated proteins) whereas the lysosomal fractions showed more variation in protein and lipid levels (53 differentially regulated proteins). This observation was consistent with the view that defects in lipid and protein transport related to lysosomes do not typically affect the composition of the cellular membrane but rather the intracellular proteins. Various autophagy-related proteins including p62/SQSTM1 and the annexins

A3 (ANXA3) and A5 (ANXA5) were identified and quantified in their dataset with ANXA5 being upregulated in LYS of NP-C1 KO [210].

In a study by Gao et al. to determine the effect of bacterial and viral infection on global lysosomal function, the authors isolated lysosomes from murine RAW 264.7 macrophage cells [212]. Changes in the proteome as well as the glycoproteome following infection with *Listeria monocytogenes,* herpes simplex virus type 1 (HSV-1) and vesicular stomatitis virus (VSV) were then assessed [212]. Isolation of lysosomes was achieved via differential centrifugation using density gradients. Without complementing their isolation with any immunocapture techniques, the proteins identified in this study may not be solely representative of the lysosome (>80% lysosomal purity). The identification of proteins annotated as ER, Golgi, and mitochondrial was indicative of the effect of infection on protein turnover resulting in translocation of these proteins to the lysosomes. Gao et al. found that there was generally a decrease in lysosomal enzymes expression following infection except the cathepsins K and L which increased after VSV and *Listeria* infection [212]. They also found, consistent with expectation of lysosomal pH being regulated within very narrow limits, that proton ATPases in the lysosomal remained largely unchanged following infection. The TLRs were identified in the study and HSV-1 infection was observed to cause upregulation of TLR3 and TLR9, which has been shown to induce autophagy in other studies [175, 212].

2.2.2.3 Other Subcellular Fractionation studies

A few other subcellular fractionation proteomics experiments have been conducted that have provided some insights into the process of autophagy. One such study by Yu et al.

examined the impact of aging on the proteome of lipid droplets in a filamentous fungus [165]. Lipid droplets are known to be multifunctional and are involved in many processes in oleaginous microbes including protein storage and autophagy [165, 213, 214]. This study found >400 proteins in the lipid droplets, 62 of which were significantly affected by aging. The authors reported the downregulation of enzymes involved in glutathione metabolism in the aging cells, signaling a malfunction in the antioxidant system. Autophagy has been shown to be involved with cellular signaling in response to reactive oxygen scavenging [215]. Combining findings from a previous study by their group in which mycelia were found to be degraded during aging, the authors then hypothesized that this malfunction resulted in autophagy activation by reason of an autophagy related protein that was upregulated in the aging sample.

Shui et al. isolated phagosomes using latex beads incorporation into mouse macrophages and subsequent sucrose gradient centrifugation and carried out a mass spectrometry-based proteomics analysis of the isolated phagosomes [216]. In this study, they confirmed proteins identified in previous phagocyte isolation studies but also many proteins that had not been identified in phagosomes. Among the proteins identified in their study are Vesicle-associated membrane protein 4 (VAMP4), TLR7, TLR9, and LC3-II [216]. Shui et al. also studied the association between autophagy and phagosomes based, in part, on the presence of LC3-II in the phagosomal isolates. The authors induced autophagy by nutrient starvation and studied the effect of inhibiting autophagosome formation and autophagosome degradation on LC3-II levels in the phagosomes. Inhibition of autophagosome formation using 3-MA reduced LC3-II levels in phagosomes while inhibiting autophagosome degradation using vinblastine resulted in increased LC3-II

levels in phagosomes. The authors therefore hypothesized that LC3-II is transferred from autophagosomes to phagosomes during the process of autophagy. The increased protein expression in phagosomes following inhibition of autophagosome degradation was also evident with VAMP but not with other proteins of the phagosomal isolates including transferrin receptor (TfR), lysosomal-associated membrane protein-1 (LAMP1), and CTSD. This suggests that VAMP4 and LC3-II trafficking to the phagosomes specifically occurs during autophagy.

In a proteomics study involving the purification of the coat protein II (COPII) inner coat of the yeast *Saccharomyces cerevisiae,* Davis et al highlighted the effect of phosphorylation on autophagosome formation [217]. Mass spectrometry-based phosphoproteomics analysis resulted in the identification of 27 phosphorylation sites on Sec24, a COPII cargo adaptor. It was found in this study that phosphorylation of Sec24 at 3 conserved amino acid sites (T324/T325/T328) contributes to the redirection of COPII coated vesicles to autophagosome formation instead of the Golgi during starvation [217]. It was also demonstrated that casein kinase 1 is responsible for the phosphorylation of the sites.

2.3 The Secretome: Proteomics Biomarker Discovery

The secretome represents one of the best opportunities for the application of proteomics as a tool for biomarker discovery in the study of autophagy due to the wealth of information resident in cellular secretomes. The cancer secretome contains various extracellular vesicles (EVs) such as microvesicles (MVs) and exosomes with central roles in tumor development and progression [218–220]. Cell secretions into culture media could represent a surrogate for secretion into blood/plasma, which could potentially provide a

rich resource for the discovery of protein markers of disease [219]. While autophagy has been classically known to be primarily a catabolic mechanism, recent studies have shed light on additional autophagic functions such as cellular protein secretion [221].

2.3.1 Autophagy and Cellular Secretion

Cellular secretion is known to occur either via the endoplasmic reticulum (ER) /Golgi pathway, which represents conventional (classical) secretion, or via alternate pathways that are independent of the ER/Golgi secretory machinery, representing non-conventional secretion. Studies have showed that about 10 – 15% of all synthesized proteins are eventually secreted into the extracellular environment through both classical and non-classical secretory pathways [93, 222]. Proteins secreted via the classical pathway are known to have an N-terminal secretory peptide signal, which makes them recognizable to the secretory apparatus. During protein synthesis, the growing protein sequences are transported to the ER membrane where they are translocated across the membrane [223, 224]. In the lumen of the ER, the signal peptides are cleaved before vesicles coated by COPII fuse with the Golgi apparatus. Proteins to be secreted are stored in vesicles within the Golgi and are directed to the cell membrane where secretion occurs [99].

Non-conventional secretion occurs through various mechanisms and may or may not involve vesicles [99, 221, 225]. Secreted proteins may contain signal peptides but are not transported to the cell surface for secretion through either COPII or the Golgi machinery or do not contain signal peptides nor depend on the ER or Golgi machinery for delivery to the cell surface for secretion [99]. Non-vesicular transport may occur, among others, via

transporters (e.g. MATa, a yeast mating factor) or binding to macromolecules such as lipids (e.g. fibroblast growth factor 2, FGF2) [225].

Autophagy has been found to have a role that interfaces with both conventional and non-conventional secretion. The non-conventional secretion of various proteins that have been shown to be secreted via autophagic vesicles in mammalian cells include high-mobility group box 1 protein (HMGB1), amyloid beta (Aβ), and interleukin 1 beta (IL-1β) [182, 221, 226, 227]. The soluble yeast protein acyl-coenzyme A-binding protein (Acb1) and its homologs in various organisms are also known to utilize the autophagy machinery for secretion [99, 221, 228]. Non-conventional secretion via autophagy is mediated by autophagosomes and also involves the Golgi re-assembly and stacking protein, GRASP which makes it unique from other modes of non-conventional secretion [221, 228]. GRASP, while being a protein that is involved in Golgi organization, plays a role in non-conventional secretion but not classical secretion through the ER/Golgi machinery. The involvement of the autophagosome in non-conventional protein secretion via autophagy coupled with the fact that autophagosomes are believed to possibly originate from certain sections of the ER gives rise to the possible interfacing with conventional secretion [180, 221, 229].

2.3.2 Proteomics of Autophagy-Associated Secretion

Mass spectrometry-based proteomics has contributed to the elucidation of the role of autophagy in secretion. A study by Kraya et al. assessed the role of autophagy in the secretion of proteins from melanoma cell lines, WM793 and 1205Lu, differing in metastatic ability and basal autophagy levels [161]. In the highly autophagic metastatic

1205Lu cell line, levels of IL-1β (interleukin 1, β), CXCL8 (chemokine C-X-C motif ligand 8), LIF (leukemia inhibitory factor), FAM3C (family with sequence similarity 3, member C), and DKK3 (dickkopf WNT signaling pathway inhibitor 3) were found to be elevated. The expression and secretion of these proteins consequently decreased when Atg7 was knocked down in the highly autophagic cells while secretion increased when the low autophagic non-metastatic cells were treated with tat-BECN1, an autophagy inducing peptide. The findings were validated in melanoma patient plasma samples where serum levels of IL-1β, CXCL8, LIF, FAM3C, and DKK3 were found to be upregulated in patients with highly autophagic tumors compared to patients with low autophagic tumors. These results suggest a possible role for autophagy-associated protein secretion in tumor metastasis. The use of proteomics tools, in conjunction with other molecular approaches, to determine the proteins and elucidate the mechanisms involved in these roles is therefore warranted and important.

A proteomics study by Kang et al. determined the effect of oxidative stress on the exosomes, a key component of cellular secretomes, in aqueous humor of patients with Neovascular Age-related Macular Degeneration (AMD). Cells were treated with Paraquat, an agent that induces oxidative stress which is a risk factor in AMD. Oxidative stress via ROS production is widely believed to be one of the major pathways of stress-induced autophagy [215]. The results of the study showed the upregulation of proteins associated with autophagy, including cathepsin D (CTSD) [230]. The cathepsins are known proteases found and activated in the acidic pH of the lysosome. Cathepsins were upregulated both in the exosomes isolated from the secretome of Paraquat-treated human retinal pigment epithelial ARPE-19 cells and the aqueous humor of AMD patients (relative to control cells

and patients respectively). In the Kang et al. study, the number of exosomes in the secretome correlated with oxidative stress levels.

Ohman et al. studied the activation of unconventional protein secretion in human macrophages using 1,3-β-Glucans, curdlan and GBY [227]. They hypothesized that β-Glucans, a principal component of fungal cell walls, constitute a key arm of fungi-induced human cell immunoactivation. Dectin-1 is the primary recognition receptor of β-Glucans in macrophages and dendritic cells [227, 231]. Mouse bone marrow-derived dendritic cells (BMDCs) with and without Dectin-1 were utilized to examine the hypothesis relating to the β-Glucans. Ohman et al. used iTRAQ labeling mass spectrometry to determine differences in protein secretion between control cells (no treatment and LPS treatment as negative controls) and β-Glucan-treated cells. This work determined that β-Glucan treatment significantly increased protein secretion and this was associated with various pathways including immune signaling, chemokine signaling pathway and leukocyte transendothelial migration [227]. The presence of many autophagy-related proteins led the authors to investigate the connection between autophagy and protein secretion in their model. Their findings indicated a strong association between autophagy and the dectin-1-mediated regulation of protein secretion. LC3 conversion was shown to increase upon treatment with curdlan and inhibition of autophagy significantly diminished IL-1β secretion even though IL-1β mRNA levels were not affected. Similarly, total protein secretion decreased with Beclin-1 knockdown (50%). The secretion of CTSD, ANXA1, tubulin, and galectin-3 in dectin-1 activated cells were all shown to decrease upon autophagy inhibition via Beclin-1 knockdown [227].

2.4 Autophagy and Cancer Therapy: How Proteomics Could Help

As discussed above, autophagy has been implicated in the pathogenesis of multiple diseases [154, 126, 232–235]. It is also well-established that autophagy represents a pivotal cellular stress response to DNA-damaging cancer therapy [236]. As a survival mechanism, cancer therapy-induced autophagy has been studied extensively as a resistance mechanism by which tumor cells can thrive under unfavorable conditions [237, 238]. Accordingly, multiple clinical trials have been conducted with the goal of inhibiting autophagy in combination with conventional or targeted cancer therapy (www.clinicaltrials.gov). These studies have overall shown inconsistent outcomes [236, 127]. This is likely due in part to the heterogeneous roles that autophagy can play. As mentioned earlier, autophagy is not always cytoprotective, and we and others have identified different functions of autophagy where it can facilitate cell death rather than survival [125]. In addition, the classical protective function of autophagy can be diminished under certain circumstances and autophagy inhibition may fail to influence tumor cell survival, a form we have termed non-protective autophagy [233]. Collectively, this inconsistency in the functions of autophagy provides a challenge to autophagy modulating therapy which is largely directed to inhibit autophagy induction or completion.

Despite the increased understanding of what dictates the function of autophagy, this field is critically lacking in biomarkers that might predict how autophagy will affect tumor cell response to therapy, which consequently would determine whether pharmacological autophagy inhibition would result in desirable outcomes. It is quite feasible that proteomics could prove to be a rigorous tool for the identification of novel markers that would allow for clinical evaluation of autophagy function in patients [239]. The primary goal is to be able to predict when chemotherapy-induced and/or radiation-induced autophagy

will be cytoprotective, based on specific serum markers secreted by tumor cells that are dependent on autophagy for survival. The compelling need for such markers provides an ideal opportunity for the application of mass spectrometry-based proteomics. To this end, our group has carried out an initial exploratory study of the secretome of p53 wild type (p53wt) and p53-null H1299 (NSCLC) cells as a function of radiation-induced autophagy (manuscript in preparation). This is an effort to identify potential blood based biomarkers that may be useful in distinguishing between NSCLC patients on the basis of response to radiation therapy during cancer treatment.

The study was carried out with the knowledge that about half of all cancers have mutations in p53 which may impact prognosis during treatment [240–242]. We also used the inducible H1299 cell line as a model of NSCLC because our previous studies had shown that in p53wt cells, autophagy was protective whereas autophagy was non-protective in p53-null cells [126, 233]. By exploring the secretome of these cells with differing p53 status using mass spectrometry-based proteomics, we identified differences due to p53 status as well as changes in secretion following ionizing radiation treatment. The proteins, regulated by p53 and/or radiation treatment, included chromogranin B (CHGB), secretogranin 2 (SCG2, glucose phosphate isomerase (GPI), thioredoxin (TXNRD1), protein FAM3C, calnexin (CANX), and eukaryotic translation initiation factor 5A-1 (EIF5A). By this study of secretion differences due to p53 status as a function of radiation treatment in NSCLC, we have demonstrated the potential of mass spectrometry to provide plausible leads in biomarker discovery efforts towards being able to predict whether the response to radiation in the context of autophagy would be beneficial.

2.5 Conclusions

We have discussed the process of autophagy and its signaling machinery from a mass spectrometry-based proteomics perspective in this review. The proteomics studies that have been described and discussed highlight the usefulness of mass spectrometry for global and targeted analysis of the autophagy process and its role on various cellular processes ranging from cell signaling to response to infections and drug action. Improvements in mass spectrometry instrumentation as well as continuous analytical software developments have enhanced the ability to measure qualitatively and quantitatively proteome changes with high accuracy and reproducibility. Autophagy is a potential mechanism of cellular resistance to therapy. However, in the absence of biomarkers that might predict benefit from autophagy modulation as a clinical strategy to complement chemotherapy and/or radiotherapy, proteomics studies that probe the cell and secretome in a largely unbiased approach could provide the avenue for identification and clinical application of such biomarkers.

Table 1 - Summary of reviewed proteomics studies on autophagy

Authors	Reference Number	Type of study/Cell compartment	Cells used	Treatment Conditions	Analytical Approach	No of proteins	Proteins of Interest	Some Key Proteins of Interest
Mancias et al	20	Autophagosomes	PANC-1, PA-TU-8988T, MCF7	Wortmannin, Chloroquine	SILAC-based density gradient centrifugation, LC-MS/MS	>2000	33	**NCOA4**, FTH1, FTL, MAP1LC3B, SQSTM1, CALCOCO1, SLC38A2, SLC7A1
Dengjel et al.	30	Autophagosomes	MCF7-eGFP-LC3, Yeast (Saccharomyces cerevisiae)	Amino acid starvation, rapamycin, concanamycin A	Iodixanol density centrifugation, Protein correlation profile (PCP)-SILAC, LC-MS/MS	7935	94	CAP1, VPS35, EEF1G, LC3, NP, RHEB, GNB2L1
Zhao et al.	31	Whole cell	SH-SY5Y	Bupivacaine	iTRAQ, SCX, LC-MS/MS	4139	241	PIK3CB, PIK3R2
Kraya et al.	32	Secretome	WM793, 1205Lu, WM9, WM1346, WM1361A, WM164, WM1366, A375	tat-BECN1, ATG7 silencing	SDS-PAGE, Label-free LC-MS/MS	599	28	FAM3C, IL1B, CXCL8, LIF, DKK3

Mathew et al.	33	Whole cell	iBMKs	Starvation, ATG5 knockout, Bafilomycin A1	SILAC, SCX, Off-gel fractionation, LC-MS/MS	7184	3181	MCM3, MCM, LIG1, PARP, STAT1, AK4, PARG, PSMC5, PSMC2, SNAP29
Zhong et al.	35	In vivo	Mouse	EVA1A knockout	Label-free LC-MS/MS	5438	28	Nnt, Ugp2, Uqcrq, Plcb3
Yu Y et al.	36	Lipid droplets	*Mortierella alpina* (Filamentous fungus)	Time (aging)	Density gradient fractionation, SDS-PAGE, Label-free LC-MS/MS	>400	62	Histone H4, Histone H2B, Tubulin alpha-1C chain, Actin, GLELO, EC:2.4.1.25-disproportionating enzyme, EC:2.5.1.6-adenosyltransferase, EC:6.3.1.2-ligase, EC:2.5.1.54-synthase
Zhang et al.	38	Whole cell	HeLa	Amino acid starvation	AHA labeling (BONCAT), iTRAQ, LC-MS/MS	711	5	HSPE1, ATP5B, SLC25A3, RACK1/GNB2L1, PNP

Zhang et al.	39	Whole cell	H3255, H1975	Erlotinib	SILAC, TiO_2 Phosphopeptide enrichment, SCX, LC-MS/MS	11207 phosphosite (3086 proteins)	~ 37	ULK1, EGFR, MAPK3, RIPK2, WNK1, STAT5B, ATG16L1, SSH2, PTPN14
Rodolfo et al.	40	Whole cell	A375, CHL-1	Ophiobolin A	2D Gel electrophoresis LC-MS/MS	N/A	24	ALDOA, TPI
Wang et al	41	Whole cell	HCT116	Curcumin	Click chemistry, AHA labeling, iTRAQ, LC-MS/MS	370	212 (197)	PRDX1, TUBB, HS90, GAPDH, FASN
Kang et al.	41	Whole cell	HCT116	EBSS induced starvation	2D Gel electrophoresis, MALDI-TOF	> 1500 GE spots	52	ANXA1, Hsc70, GRP78, PDIA3, ENO1, GST-P1
Zhuo et al.	43	Whole cell	MEFs	Starvation, Cytochalasin D for F-actin depolymerization	Itraq, online 2D LC-MS/MS	1234	114	F-actin network
Bertin et al.	46	Whole cell (In vitro and in vivo)	DHD/K12/RPOb (rat), MCF-7, PC-3, HEK293	CpG-ODN, Rapamycin	2D Gel electrophoresis, LC-MS/MS	N/A	16	ANXA1, PGAM1, DPYSL3, CNN3, TPM1, GAPDH, PHB, eIF4A1, PC, MnSOD, GRP78

Li et al.	48	Whole cell	Mouse brain, NA cells	Rabies viruses (CVS-11, SRV9)	iTRAQ, LC-MS/MS	2285	265	PRDX5, mTOR, SOD1, SOD2, ATP5C1, EIF3E, EIF4B
Patella et al.	49	Whole cell	HUVECs	siATG5, Bafilomycin	SILAC spike-in, LC-MS/MS	7565	2221	CD55, CRIM1, RHOB, ESM1, CTSA, CTSB, CTSZ, PPT1, TPP1
Kim et al.	55	Whole cell	A427, A549, Calu-1, Calu-6,H157	MEK-inhibitors (AZD6244 and MEK162)	ATP-binding proteome labeling and enrichment, LC-MS/MS	1925	24	JAK1, FAK1, MKK3, MKK6, PLK1, ULK1, ULK3, AMPK, AURKA
Tavera-Mendoza et al	62	Whole cell	MCF-7, MDA-MB-231, MDA-MB-453, ZR-75-1, MCF-12A	1,25(OH)2D3	Rapid Immunoprecipitation Mass Spectrometry (RIME)	N/A	N/A	LC3B, p62, HSP90AB1, EGFR, ULK2, CXCR4
Overbye et al	64	Autophagosomes	Male Wistar Rat primary hepatocytes	Starvation, Vinblastine	Differential centrifugation, 2D gel electrophoresis, MALDI-TOF	> 1500 GE spots	39	PEBP, COMT, BHMT, IPP, AMPK

Gao et al.	65	Autophagosomes	HEK-293, HCT116	Starvation, Calcium phosphate precipitate	Immunoisolation, 2D gel electrophoresis, MALD-MS/MS	101	N/A	LC3, ATG5, ATG16, ATG9, SQSTM1, UMP-CMP Kinase, GRP-78, Rab4, Rab5
Suzuki et al.	67	Autophagosomes	Saccharomyces cerevisiae	Starvation	GFP-fuse aminopeptidase fluorescence, Iodixanol gradient centrifugation, LC-MS/MS	40	N/A	prApe1, Ald6, Pyk1, Yef3, Hsc82, Eft1
Bagshaw et al.	68	Lysosomes	Male Sprague-Dawley rat liver lysosomes	Triton-WR1339, Tyloxapol	Differential sucrose density centrifugation, lysosomal subfractionation, sepharose cation exchange, 1D gel electrophoresis,LC-MS/MS	215	N/A	LAMP1, LAMP2, LIMPII, Rab6, Rab7, Rab1A, Rab11B, VAMP8, CYP2A1, CYP2C29, CYP2D2, CYP4A3, APOA-V, APOE, APOB, ST6Gal

Naureckiene et al.	73	Lysosomes	Human brain biopsies, rat liver	NP-C2 disease, Controls (No disease or NP-C1 patients), Triton-WR1339	Mannose-6-phosphate affinity enrichment, 2D SDS-PAGE, differential sucrose density centrifugation, anion exchange chromatography	N/A	1	HE1
Sleat et al.	74	Lysosomes	Human brain biopsies	LINCL disease and Controls (No disease)	Mannose-6-phosphate affinity chromatography purification, 2D gel electrophoresis, Edman degradation	N/A	1	Pepstatin
Chapel et al.	78	Lysosomes	Rat liver lysosomes, HeLa	-	Differential centrifugation, Nycodenz density gradient fractionation, liquid-liquid extraction, SDS-PAGE, LC-MS/MS	734	12	LOH12CR1, MFSD1, PTTG1IP, SLC37A2, SLC38A7, SLC46A3, SLCO2B1, STARD10, TMEM104, TMEM175, TTYH2, TTYH3

Della Valle et al	79	Lysosomes	Male Wistar Rat liver homogenates	Triton-WR1339	Differential centrifugation, isopycnic sucrose density centrifugation, iTRAQ, 2D SCX and RP peptide separation, MALDI-TOF	1273 gene products	N/A	CTSD, ACP2, LAMP2, ACP5
Kollman et al.	82	Lysosomal proteins/Secretome	MEF23-1SV mouse fibroblasts	MPR46 and MPR300 deficiency, Pepstatin A and Leupeptin	Mannose-6-phosphate affinity enrichment, 2D gel electrophoresis, MALD-TOF, MudPIT	38	N/A	CTSD, CTSB, CTSZ, MERP-2, M2B2, CREG, RISC
Tharkeshwar et al.	83	Lysosomes	HeLa	NP-C1 knockout	SPION isolation, LC-MS/MS	~ 2400	53	LIPA, IFI30, IGF2R, GABARAPL2, CALCOCO2, SQSTM1, FOLR1, ACKR3, RHOB, ITGA11, Rab5
Gao et al.	85	Lysosomes	Murine RAW 264.7 macrophage cells	Listeria monocytogenes, HSV-1, VSV infections	Differential centrifugation, density gradient fractionation, TMT labeling, LC-MS/MS	3704	~ 204	Fam120c, Clec4e, Cxcl2, Ccl9, Hspe1, Hspa2, Hmox1, Slc15a3, Cd274

Shui et al.	89	Phagosomes	Mouse macrophages	3-methyladenine (3-MA)	Sucrose gradient centrifugation, 2D SDS-PAGE, LC-MS/MS	546	N/A	EEA1, TfR, CatD, VAMP4, TLR7, TLR9, LC3-II, LAMP1, JAK1
Davis et al.	90	COPII coated vesicles	*Saccharomyces cerevisiae*	Nitrogen starvation, Rapamycin, 0.5% galactose (Sec24 induction)	His tag protein purification, LC-MS/MS	27 phosphosites	3 phosphosites	Sec24 (T324/T325/T328)
Ohman et al.	102	Secretome/Exosome	Human PMBC-derived primary macrophages, Mouse BMDC	LPS, 1,3-β-Glucans (Curdlan, GBY), 3- MA, SykII, Src inhibitor I, Brefeldin A, Dectin-deficiency	Exosome isolation via ultracentrifugation, iTRAQ, SCX, LC-MS/MS	1597	6 pathways	Chemokine signaling pathway, cytokine- cytokine receptor interaction, and MAPK signaling pathways, cytosolic DNA-sensing pathway, Jak-STAT signaling path-way,NOD-like receptor signaling pathways (IL-1β, CTSD, ANXA1, Tubulin, LEG3)

Kang et al.	105	Aqueous humor/Secretome/Exosome	ARPE-19, AMD patients and controls	Paraquat, serum-free	Exosome isolation via precipitation, SDS-PAGE, LC-MS/MS, LC-MRM	1209	6	CTSD, HSPA1, ACTA2, KRT8, KRT14, MYH9

Chapter 3:
Secretome Analysis of Non-Small Cell Lung Cancer in Response to Ionizing Radiation and p53-status

3.1 Lung Cancer Statistics: US and Worldwide

Lung cancer refers to cancer that starts in the tissues or cells of the lungs. It is currently the deadliest type of cancer in both men and women, killing more people than breast, colorectal and prostate cancers combined and accounting for about 1.59 million deaths worldwide in 2012 [243]. About 159,000 estimated deaths are projected in the US in 2014 [244]. In spite of the decreasing trends in the incidence of lung cancer amongst both males and females [245], it continues to be one of the most prevalent cancers both nationwide and around the world.

The majority of lung cancer patients living today were diagnosed within the last 5 years and this is because of the low 5-year relative survival (16.8 – 17.8%) [246]. In 1977, the 5-year relative survival for lung cancer was 12.3% [247]. The 5-year survival depends largely on the stage of diagnosis with early stage NSCLC patients having a survival of >50% compared to <5% in late stage patients [245, 248]. This relatively low survival rate is a result of late diagnosis for majority of cases and/or ineffectiveness of available treatments. Of the new NSCLC patients in the US, ~16% are found to have local or early stage while ~57% are metastatic at diagnosis [245]. This makes early detection of lung cancer a priority

in cancer research in the bid to improve survival. Current standard treatment regimen involves the use of chemotherapy and radiation concurrently [249] but the evidence of relapse, disease recurrence, and/or mortality indicates then that the current treatments are ineffective or at best insufficient. Amongst the possible mechanisms responsible for the resistance and/or relapse is autophagy [161].

3.2 Types of Lung Cancer

Lung cancer is characterized by molecular heterogeneity, comprising of various small populations of cells with distinct features. This heterogeneity may be within a given lung tumor or between one tumor type and another [250].

Lung cancer pathogenesis is associated primarily with smoking even though non-smokers do develop the disease [247, 250]. Smoking accounts for > 80% of lung cancer inside and outside the US [247]. Exposure to different agents including asbestos, arsenic, and radiation from homes and mines (radon gas) as well as air pollution constitute some of the other causative agents in lung cancer [247]. These causative factors may act in synergy to result in disease [247]. There are two main types of lung cancer depending on the type of cells affected in the lungs namely small cell lung cancer (SCLC ~10-15%) and non-small cell lung cancer (NSCLC ~ 85%) [246, 250, 251]. Some sources add a third type, lung carcinoid tumor estimated to be found in ~5% of lung cancer cases. It is uncommon, grows slower than the two major types and constitutes neuroendocrine cells [252, 253].

3.2.1 Small Cell Lung Cancer

SCLC is aggressive and highly metastatic, and estimated to be responsible for ~ 250,000 deaths throughout the world each year. The vast majority of SCLC patients currently

smoke or smoked heavily in the past [251]. With a 5-year survival rate of 7%, SCLC is one of the most aggressive tumors, being designated at one time a “recalcitrant cancer” [251, 254]. The scarcity of tumor tissue, or the lack thereof, for scientific research greatly hindered advancements in the understanding of the disease. Consequently, most SCLC patients are still treated with only the first line platinum-based chemotherapeutic agents such as cisplatin with the only second line treatment option being topotecan [254]. Given that there is loss of the tumor suppressor genes TP53 and RB1 in almost all SCLC cases, it is plausible that relapse and resistance to therapy develops after the responsiveness of tumors to initial treatment [250, 254].

3.2.2 Non-Small Cell Lung Cancer

NSCLC is classified into three types namely squamous cell carcinoma (LUSC), adenocarcinoma (LUAD), and large cell carcinoma [246]. Despite being associated with SCLC, neuroendocrine features have been identified in a “fourth class” of NSCLC cells (~3%) referred to as large cell neuroendocrine tumors [255].

Squamous cell carcinoma and adenocarcinoma constitute the most prevalent subtypes of NSCLC with LUSC making up about 25-30% and LUAD about 40% of all lung cancer cases [246, 250]. Of the two most common types of NSCLC, squamous cell carcinoma has the stronger association with smoking. Most lung cancer patients who have never smoked, mostly women, are usually diagnosed with adenocarcinoma [250]. LUAD is relatively slower growing and may be detected before it metastasizes out of the lungs; it is also normally associated with exposure to second-hand smoke and carcinogens like asbestos and radon gas [246, 250].

Relative to SCLC, NSCLC has seen the most advancement particularly regarding treatment options. Over the past two decades, more sophisticated and individualized treatment options have been developed with some promising results [250]. This has been due to the ready availability of clinical samples for scientific studies owing to the significantly greater number (and percentage) of cases encountered compared to SCLC.

This chapter focuses on NSCLC due to the aforementioned reasons together with the historically poor prognosis following treatment [256] and the 5-year survival rate of ~ 17%.

3.3 NSCLC Treatment Options

NSCLC treatment depends on the stage of the disease at diagnosis. In early disease where the tumor is resectable (based on imaging and biopsies), surgery is the first line of treatment for stage I, II, and IIIA NSCLC [246]. This initial surgery is usually followed by adjuvant chemo- and/or radiotherapy as well as targeted therapy [246, 257].

3.3.1 Chemotherapy

In about 40% of patients where diagnosis first happens at stage IV of the disease, combination cytotoxic chemotherapy constitutes the first line of treatment [246]. This is dependent upon the comorbid conditions, tumor histology, patient age, and performance status (PS). Frequency and types of side effects as well as patient tolerance to side effects of the cytotoxic agents informs the specific combination regimen selected for a patient [246].

When PS ≤ 1, along with palliative care and symptomatic treatment, a platinum in combination with paclitaxel, gemcitabine, docetaxel, vinorelbine, irinotecan, or

pemetrexed is used per the American Society of Clinical Oncology treatment guidelines [246, 258]. Studies have shown that these combinations are similar in effect and no one is superior to the other with median overall survival around 8-10 months. When PS=2, only one chemotherapeutic agent is recommended and this treatment is stopped in the event of intolerable adverse effects, tumor size growth, or stable disease but no decrease in tumor size after four treatment cycles. For patients with PS of 3, supportive care rather than chemotherapy is recommended. Chemotherapy offers little to no benefit to these patients and may result in significant decline in patients' quality of life.

3.3.1.1 Advances in Chemotherapy

There has been significant advancement in the survival stats for NSCLC patients following the advent of personalized medicine that targets certain gene mutations and/or rearrangements. Currently, targeted therapy exists for patients with mutations of the epidermal growth factor receptor (EGFR) (~10-15% of LUAD patients of Caucasian or European descent), and BRAF (1-4% of all NSCLC) genes as well as gene rearrangements involving the anaplastic lymphoma kinase (ALK) gene (~3-7% of usually younger patients) and the ROS1 gene [246, 259]. Other mutations with promising targeted therapy include KRAS and HER2 mutations. EGFR mutations are treated with such agents as erlotinib and gefitinib while crizotinib represents the treatment of choice for patients with previously untreated, advanced ALK mutations in non-squamous NSCLC [246].

The immune system has been shown to have a dual role in cancer development; suppressing growth of some tumor cells while promoting the progression of other tumor

cells able to survive in immunocompetent hosts [260]. In its protective role in cancer, the immune system is able to considerably slow down or stop the growth of tumor cells and prevent metastases [246, 260]. Immunotherapy involving agents that are thought to enhance the immune system's ability to differentiate between cancer cells and normal body cells is one of the advances in cancer treatment. Immunotherapy targets mechanisms through which cancer cells evade the immune system such as checkpoint pathways. For example, the protein PD-L1, a ligand for the inhibitory cytotoxic receptor programmed cell death-1 (PD-1), is a target of the immunotherapy drugs pembrolizumab and nivolumab [246, 259, 261].

3.3.2 Radiation Treatment

Approximately 50% of all cancer patients receive radiation therapy at some point in their disease management [262]. Radiation may be used to target tumors at specific sites in the body and may be used alone, in combination with chemotherapy or as neo-adjuvant therapy to shrink tumors thereby making surgery easier and more likely to succeed. In a retrospective study of patients with stage III NSCLC using the Surveillance, Epidemiology, and End Results (SEER) database, neo-adjuvant radiotherapy was shown to be associated with significant increase in 3-year overall survival compared to other treatment regimen including surgery only, radiation only, or surgery plus postoperative radiotherapy [256]. In patients with advanced lung cancer that is not amenable to surgical resection, or patients that do not respond to surgery or chemotherapy, radiation treatment is used in combination with palliative care [246].

Radiotherapy may be administered with an external radiation machine (external-beam radiation therapy) or internally (internal-beam radiation therapy or brachytherapy) via

radioactive agents, placed in or near the tumor cells inside the body [263]. In stereotactic body radiation therapy (SBRT), tumors are located with precision using imaging techniques and advanced coordinate systems in order to administer precise and concentrated doses of focused radiation to the tumors. This is typically used in early disease where tumors have not metastasized [246].

Radiation treatment involves the use of high-energy beams that result in DNA damage in cancer cells. Radiation is measured in Grays (Gy) and the dose used is dependent on the type, histology, and stage of cancer. Radiotherapy acts by penetrating cells and directly causing clustered DNA damage including double-stranded breaks (DSB) that ultimately results in cell death if the break is not repaired [263, 264]. Radiation also generates single-stranded breaks as well as reactive oxygen species (ROS) that result in the oxidation of proteins and lipids in cells [263]. ROS generated as a by-product of aerobic metabolism induces about 50,000 lesions of DNA damage. This is, in fact, more than the number of lesions induced by ionizing radiation treatment with 2 Gy radiation therapy [264]. However, the major difference between endogenous ROS-induced DNA damage and radiotherapy-induced DNA damage is the production of lethal DSB which has deleterious cytotoxic effects in the body [264]. The lethality of DSB is a consequence of the inherent difficulty of the DNA repair mechanisms to repair such breaks relative to SSBs.

3.3.3 Limitations of Current Treatment

Many advancements have been made in the development of new therapies for lung cancer, particular NSCLC. However, the fact remains that the vast majority of NSCLC diagnosis are often made in late disease when the tumors have grown in size and have

metastasized. Furthermore, in spite of increased disease- or symptom-free periods facilitated by the availability of newer more targeted chemotherapeutic agents, majority of NSCLC patients ultimately relapse or develop resistance to therapy [265]. Tumor cells exploit a number of mechanisms for promoting proliferation and evading cell death by apoptosis amongst which include the functional loss of p53 tumor suppressor activity, and autophagy [266–268].

About 50% of all cancers and ~50-60% of lung cancer cases are estimated to have a loss of function mutation of the TP53 gene [240–242]. NSCLC tumors have been shown to evade treatment-induced cell death via various strategies that include tumor heterogeneity [269], and autophagy [267, 270].

3.3.4 NSCLC Tumor Response to Therapy

Radiation-induced DNA damage induces different responses in cells including but not limited to apoptosis, senescence, and autophagy. When cells are exposed to such external stress as radiotherapy, the initial response is cell survival through mechanisms which include the heat shock response (HSR), unfolded protein response (UPR), the DNA damage response, and the oxidative stress response [271]. However, depending on the level and duration of the stress, these mechanisms may fail and the cell may switch to cell death signaling pathways, which include apoptosis, necrosis, and mitotic catastrophe. Tumor cells seek to disrupt the normal homeostatic balance that exists naturally between cell proliferation and cell death [271]. By decreasing the rate of cell death through the disruption of cell death signaling pathways, tumor cells are able to survive longer and proliferate more.

3.3.4.1 Apoptosis

Apoptosis is a major cellular process and represents one of the most extensively studied mechanisms of cell death. Apoptosis is one of several cellular mechanisms that occur in response to external stress such as ionizing radiation when cell survival mechanisms fail to contain the stress or stop its deleterious effects. Also known as programmed cell death, apoptosis was first coined as a term in 1972 and refers to a highly organized intracellular event [272, 273].

During apoptosis, cells undergo changes that include chromatin condensation, detachment from the surrounding tissue and shrinkage, as well as blebbing of their cell membranes [274]. Two mechanisms of apoptosis namely caspase-dependent and caspase-independent mechanisms have been elucidated. The classical caspase-dependent mechanism may be initiated either by intrinsic or extrinsic factors. The intrinsic or mitochondrial pathway is mediated by activation of BAX and BAK of the Bcl-2 family proteins while the extrinsic pathway is mediated by death receptors (CD95 aka APO1/Fas), Tumor necrosis factor alpha (TNFα) receptors, and TNF related apoptosis inducing ligand (TRAIL) receptors that are stimulated by their respective ligands. When the intrinsic pathway is initiated via the Bcl2 family protein activation, Cytochrome C is released into the cytosol where it binds to apoptotic protease activating factor (Apaf1) and forms a complex [274–276] which activates caspase-9 that sets off the caspase cascade. The intrinsic pathway may however, also initiate the caspase-independent mechanism where it results in the release of apoptosis inducing factor and endonuclease G which results in chromatin condensation [275, 276].

In the extrinsic pathway, binding of the ligands to their receptors results in the attraction of adaptor molecules such as FADD which recognizes a death domain and leads to the activation of caspase-8 which culminates in the activation of the entire caspase cascade [275–277]. A consequence of apoptosis is the proteolytic cleavage of poly(ADP-ribose)polymerase-1 (PARP-1), a DNA repair enzyme involved in transcriptional regulation [278].

In essence, apoptosis is a suicide mechanism where cells that are no longer needed by the organism or that are diseased beyond repair kill themselves. In normal cells, apoptosis represents a homeostatic process which ensures tissue growth and aging, but represents a response mechanism in tumor treatment therapy [271]. Apoptosis may also be a response to DNA double strand breaks resulting from ionizing radiation treatment [279] and is usually the fastest mode of cell death induced in tissues while the other forms such as necrosis and autophagic cell death kick in with inhibition of apoptosis [275, 278]. In his paper, Meyn distinguishes between two different forms of apoptosis following ionizing radiation treatment depending on occurrence before or after post-irradiation mitosis [273]. He refers to the programmed cell death observed before the first mitotic division (4 – 6 hrs. post irradiation) as primary apoptosis whereas he terms as secondary apoptosis, the cell death that occurs after the first mitotic division (>24 hrs. post irradiation). He argues for the relevance of this observation, especially in radiotherapy, because cells that undergo primary apoptosis actually affect clonogenic survival compared with those that undergo secondary apoptosis by which time cells would have divided anyway and so death does not necessarily decrease the live cell number. As a cellular response to radiotherapy, apoptosis may likely be representative of an intrinsic sensitivity or otherwise

in that whereas 5 Gy of IR is usually sufficient to induce apoptosis in sensitive cells, more radiation is needed for secondary apoptosis and is observed much later [273].

Apoptosis may result in decreased tumor clonogenic survival and consequently enhance sensitivity to chemo- and/or radiotherapy. Studies have shown that tumors with high apoptotic indices before treatment tend to respond better and this index has been found to be predictive of patient survival indicating the desirability of apoptosis as a treatment outcome [273, 280, 281]. It is important to note that while apoptosis may be a desirable outcome of treatment, it may also be a mechanism of eliminating susceptible cells so that the tumor can have only resistant cells, which can then mutate to result in 'intractable' tumors.

3.3.4.2 Autophagy and Senescence

Autophagy, as discussed extensively in chapter 2, is an intracellular mechanism in which cells recycle some of their cytoplasmic components, in response to stress such as starvation, chemotherapy or radiation, by degradation with the help of lysosomes for the production of energy [96, 155, 274, 282–285]. The recycling occurs via the fusion of autophagosomes with lysosomes, which release hydrolases to degrade the contents of the autophagosome.

Autophagy was historically thought to serve a protective function; however, recent studies have painted a more complex picture. At least four functional forms of autophagy have been identified in the context of cancer cells; protective, non-protective, cytotoxic, and cytostatic [125]. Within these functional forms, protective autophagy holds the most promise for pharmacological inhibition through increased sensitization to standard of care

chemo-radiation treatment [126]. The identification of these four functional forms of autophagy also provides a plausible explanation for the potential failure of clinical trials exploring the inhibition of autophagy as a cancer treatment enhancement strategy as patients could not be stratified based on the nature of the autophagy exhibited by their tumor [127].

Senescence, on the other hand, is a growth arrest phenomenon whereby cells are alive and have active metabolism but cannot divide [286]. Two major forms of senescence have been identified, replicative senescence in which cells cannot divide anymore owing to reaching a length of telomere referred to as the "Hayflick Limit", which does not permit further cell division and stress-induced premature senescence which does not depend on telomere length [287–290]. When cells reach the Hayflick Limit suggesting that the cells are approaching senescence, a DNA damage response is initiated and cells show foci that indicate the presence of γ-H2AX. The proteins – p53 binding protein 1 (53BP1), nibrin (NBS1), and Mediator of DNA damage checkpoint protein 1 (MDC1) which are DNA damage response proteins also show up and activation of ataxia telangiectasia-mutated (ATM) and ataxia telangiectasia and Rad3-related protein (ATR) is seen [289, 291]. ATM and ATR subsequently induce the activation of checkpoint kinase 1 (CHK1) and CHK, which may then lead to the activation of p53 and several other factors and proteins that are related to and regulate the cell cycle. Together with p53, activation of Rb, a tumor suppressor and its downstream signals including $p16^{INK4A}$ has also been demonstrated in senescence [288, 289]. Silencing $p19^{ARF}$, a direct upstream regulator of p53, or $p16^{INK4A}$ resulted in a consequent prevention of Ras oncogene induced senescence [289, 292].

In various cancer cells, it appears as though autophagy and senescence may be induced concurrently or consecutively. However, reports of the relationship between autophagy and senescence in the literature are inconclusive at best. While some studies have showed autophagy and senescence to be directly related, other studies have essentially established that the two processes are dissociable. Young et al. showed in their work that autophagy was directly related to oncogene induced senescence by showing an increase in LC3-II levels, a marker of the initiation of autophagy, in senescent but not quiescent cells [286]. The association between autophagy and senescence was however, not very strong when autophagy genes ATG5 and ATG7 were silenced. In their silencing studies, the production of IL-6 and IL-8 which are senescence-associated cytokines was comparable in empty vector and ATG5 silenced cells albeit there was slightly more IL-6 on day 4 in the vector. Senescence could therefore be said to have only been delayed by the inhibition of autophagy. This may suggest that autophagy may play a role in the onset of senescence but not the sustenance thereof. Progression of senescence may thus be independent of autophagy. Gewirtz et al. also showed, like in the Young paper, that senescence is reduced and/or delayed, but not entirely blocked, in MCF-7 and HCT-116 tumor cells with pharmacological or genetic inhibition of autophagy [293]. Anna Knizhnik et al. have demonstrated temozolomide-induced senescence that is completely abolished by inhibition of autophagy with 3-methyladenine (3-MA) [294]. Considering however, that 3-MA can have off-target effects, caution has to be exercised in assuming the complete inhibition of senescence was due only to autophagy. Strong evidence of a relationship between autophagy and senescence is also presented in a study of autophagy competent cells which express proteolytic Cyclin E fragment (p18-CycE) where ATG7 was silenced

and which resulted in a consequent decrease in senescence associated markers β-galactosidase and HP1γ [295]. Some studies have pointed to the decrease in chaperone-mediated autophagy in senescent cells [289]. Studies have shown a cytostatic form of autophagy [126] which is functionally similar to senescence in that cells are alive and metabolically active but do not proliferate. Regarding the signaling pathways of both autophagy and senescence, some overlap has been reported. Amongst the overlapping features of both pathways are ROS generation, activation of ATM kinase, TP53 induction, CDKN1A/p21 induction and dephosphorylation of Rb [283, 296]. In a study where Regulator of Cullins-1 (ROC-1) was knocked down, there was the induction of senescence (mediated by p21) and a time dependent increase in autophagy (marked by increased LC3-I to LC3-II conversion) [297]. This points to a strong molecular /signaling association between the two. These pathway signaling overlaps have however, are still undergoing rigorous scientific interrogation and therefore no conclusions can be made yet regarding the interdependence of the two processes.

3.4 TP53 and Autophagy

Autophagy has been considered as a central cellular response for some cancer cell types, including NSCLC cells, to avoid radiation-induced cell death [126, 232, 233]. The TP53 gene has been shown to promote cell survival via autophagy activation following starvation whereas autophagy suppresses the activity of p53 suggesting a negative feedback mechanism [298, 299].

In a study of 188 patients with early stage NSCLC, Ahrendt et al. found that p53 mutations in patients' tumors resulted in poor prognosis and significantly higher deaths [241]. Levine

et al. also showed that the TP53 gene plays a role in the increased cellular secretion of some proteins and the decreased secretion of others particularly via exosomes in NSCLC cell lines including H460 and H1299 [300]. Chenau et al. used the NSCLC cell line as a tool to investigate secreted proteins in which they found the secretion of about 91 proteins to be affected by the p53 status of the cell [301]. The high prevalence of TP53 mutation in cancers and more specifically NSCLC combined with its reported association with treatment resistance arising from protective autophagy makes TP53 an important gene for study.

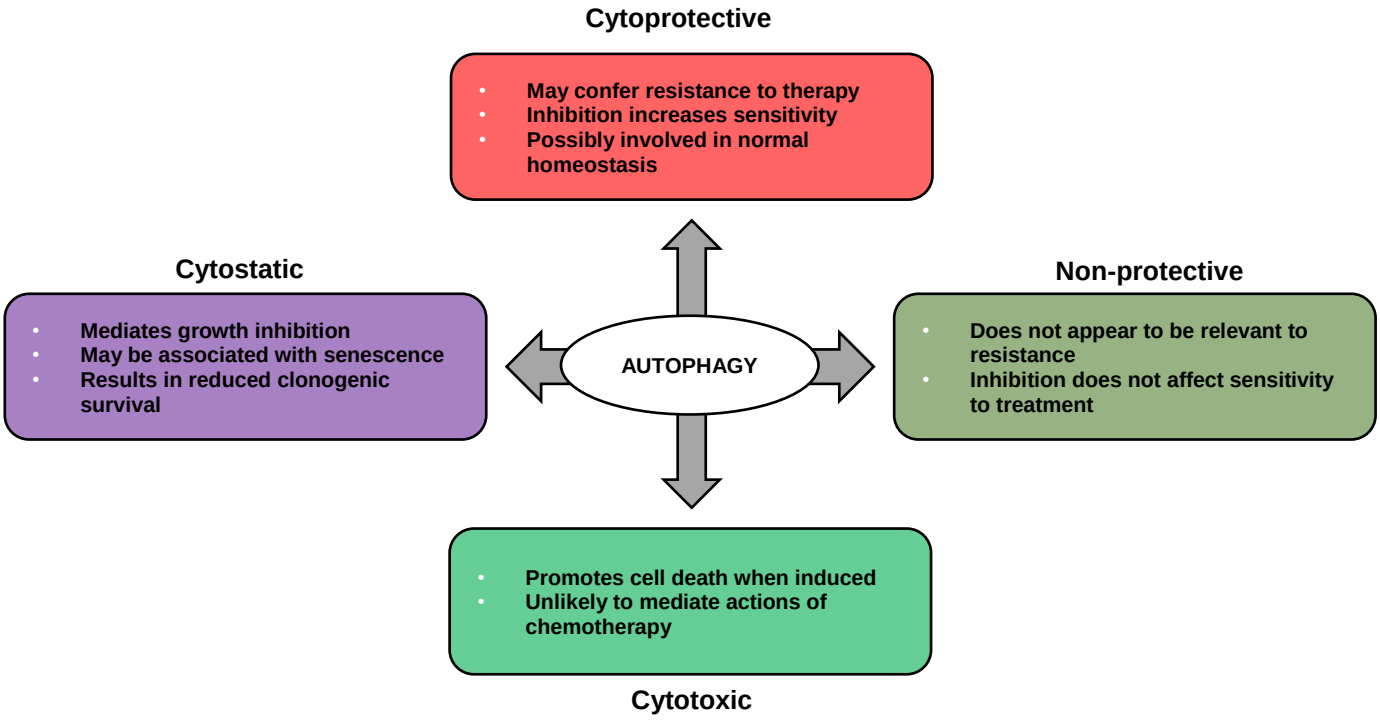


Figure 11 - Functional forms of autophagy

3.5 Justification and Objectives of the Study

The search for biomarkers for the detection of at-risk persons and early disease diagnosis remains one of the major focuses of lung cancer research [97, 247]. Many potential markers of disease have been proposed through various studies but issues of sensitivity and/or specificity and selectivity affect their adaptation in clinics [247]. Blood-based markers are of particular interest due to the limited access to the lung, a visceral organ, except by direct invasive approaches [247]. Working on the hypothesis that some proteins that are eventually secreted into the blood are up- or downregulated due to the development of cancer, proteomics studies aim to detect these proteins early in disease pathogenesis in order to serve as biomarkers [97].

Several reports have shown the value of using cancer cell lines as experimental models for identifying potential protein biomarkers secreted and/or shed into the media (i.e., secretome) [161, 249, 301–304]. High resolution and accurate mass spectrometry (HRAM) LC-MS/MS provides an unbiased method for identifying protein signatures of autophagy that may be useful in differentiating between different responses to radiation in NSCLC.

Various cell lines have shown differences in response to the manipulation of autophagy using both pharmacological and genetic strategies. For instance, cytoprotective autophagy has been demonstrated following ionizing radiation (IR) in tumor cells from a variety of backgrounds, including A549, H460, and CT26 [305], and HTB35 [232] cells. The protective function of radiation-induced autophagy has however, been observed primarily if not exclusively in cancer cells with functional p53, as in H1299 and H460 cells with wild-type p53 [126]. In contrast, H1299 cells that do not express functional p53 demonstrated non-protective autophagy in that the cells were not sensitized to radiation when autophagy was blocked by pharmacologic or genetic approaches [233].

H1299 cells are immortalized non-small cell lung cancer cells derived from lymph nodes. The cells do not produce the p53 protein due to a homozygous partial deletion of the TP53 gene. However, H1299 cells capable of producing p53 via a tetracycline-inducible p53 construct were developed by Constantinos Koumenis resulting in an isogenic cell line with and without p53. Consequently, in the present study, we have used quantitative proteomics to investigate the effect of p53 expression on non-small cell lung cancer response to ionizing radiation, using the isogenic H1299 cell line. The H1299 secretome was analyzed in an effort to identify potential candidate protein biomarkers or therapeutic targets of cytoprotective autophagy.

3.6 Experimental

3.6.1 Materials and Reagents

Chemicals were purchased from Invitrogen, Gibco, Thermo Scientific, Sigma Ultra, Fisher, Promega, Fluka, and Honeywell Burdick & Jackson and used as received unless otherwise stated.

3.6.2 Cell Culture Conditions and Ionizing Radiation Treatment

H1299 TP53-null (p53-null) and TP53-inducible (p53-wt) cells, originally developed by Dr. Constantinos Koumenis,[306] were obtained from Drs. Frank and Suzy Torti (University of Connecticut). H1299 cells were cultured in low glucose Dulbecco's Modified Eagle Media (Invitrogen) supplemented with 10% Fetal Bovine Serum (Thermo Scientific), 1% 100 U/ml Penicillin G sodium/1% 100 µg/ml Streptomycin sulfate (Invitrogen) at 37°C and 5% CO_2. H1299 p53-wt cells were maintained in 1 mg/ml Doxycycline to induce *tp*53 expression. H1299 p53-wt and H1299 p53-null cells were seeded in two sets of 9 × 10

cm dishes at 20,000 cells/mL with 10 mL of media and grown to ~85% confluency over 48 hours. At ~85% confluency, the 10% FBS media the culture media was aspirated from all plates. The plates were washed once with 5 mL of Phosphate Buffered Saline 1X pH 7.4 (Gibco) and then twice with 5 mL serum-free DMEM. A final 5 mL volume of serum-free DMEM was added to all plates. A set of 3× p53-wt and 3× p53-null cells were then immediately treated with ionizing radiation (6Gy) after which all plates were incubated for 12 hours. The serum-free media containing secreted proteins (secretomes) were collected into 15 mL tubes after a 12-hour incubation and stored at -80^{o}C until processing.

3.6.3 Western Blot

Cells were pelleted and washed briefly with cold PBS (Gibco) after which they were lysed using CHAPS buffer containing protease and phosphatase inhibitors (Sigma Aldrich). Total protein concentration of samples was determined using the Bradford assay (Bio-Rad Labs). Equivalent amounts of proteins were then run on an SDS-PAGE gel and blotted onto polyvinylidene difluoride (PVDF) membranes (Bio-Rad Labs). The quality of the transfer was then determined by staining with Ponceau S stain. The blots were then processed and visualized as described by Sharma et al. using antibodies against p53 and GAPDH [126].

3.6.4 Proteomics Sample Preparation

The collected media (secretomes) samples were removed from -80^{o}C, thawed, centrifuged at 2500 rpm for 5 minutes at 8^{o}C, and the supernatant concentrated in 3 kDa MWCO filters (Millipore Amicon Ultra) at 7500 × g for 30 minutes at 8^{o}C. The concentrated secretomes were then transferred into 10 kDa MWCO filters (Millipore Amicon Ultra) and

washed with 3 × 500 µL Tris HCl buffer pH 8.1. Remaining solutions were made up to the same volume and total protein concentration for each secretome determined at 280 nm using a BioTek Synergy H1 plate reader fitted with a Take3 Plate. The secretomes were processed for proteomics analysis using the FASP method [307]. Specifically, each secretome sample in the 10 kDa MWCO filter was reduced using 40 µL of 50 mM Dithiothreitol (Fisher Scientific) by incubation for 45 mins at 56^{o}C, alkylated at room temperature in the dark with 30 µL of 142 mM Iodoacetamide (SigmaUltra), and then rinsed with 300 µL of Tris-HCl pH 8.1 at 14,000 × g for 10 minutes. Digestion was performed with 6 µL of 1 µg/ µL Trypsin Gold (Promega) solution overnight at 37^{o}C and then terminated with 200 µL of 0.1% acetic acid (Fluka Analytical).

3.6.5 LC-MS/MS Analysis

50 µL solutions of 100 ng/µL peptide sample were prepared by dilution with 0.1% Acetic Acid for LC-MS/MS analysis. The LC-MS/MS system consists of an Eksigent nLC 415 (ABSciex) in a trap and elute configuration. The reverse phase trap column (75 µm x 1 cm) and analytical column (75µm x 15cm) were both packed in-house with 5 µm Magic AQ C18, 200Å stationary phase. The nLC system was coupled to a Q-Exactive (Thermo Scientific, San Jose, CA) mass spectrometer equipped with the Nanospray-Flex ionization source fitted with a 10 µm ID emitter tip (New Objective). 2 µL of sample was loaded onto the trap column and desalted at a flow rate of 2.25 µL/min for 5 minutes using mobile phase A. Desalted peptides were then eluted at 300 nL/min with the following gradient: 5% B (0 – 4 minutes), 35% B (95 minutes), 75 % B (105 – 110 minutes), 5% B (115 minutes) and held for 5 minutes until the run finishes at 120 minutes. Mobile phase A consisted of 98% H_2O/2% acetonitrile, 0.1% formic acid and mobile phase B consisted

of 2% H_2O/98% acetonitrile, 0.1% formic acid. The electrospray emitter tip was charged with a voltage of 1.80 kV in positive ion mode and the Q-Exactive inlet temperature and S-lens setting were maintained at 250°C and 62 V, respectively. Full scan (400-1600 m/z) resolution was set at 70,000 FWHM with an AGC target of 3×10^6. MS/MS was set to a resolution of 17,500 with an AGC target of 2×10^4 at 120 ms maximum inject time and selection of the top 12 ions at a 30 second dynamic exclusion. HCD voltage was maintained at 30 NCE throughout.

3.6.6 Data Analysis

Proteomic datasets were processed in MaxQuant (ver. 1.5.2.8) with the Andromeda search algorithm using the Uniprot Human proteome Fasta database (downloaded on 04-04-2016); mass accuracies: MS = 5 ppm, MS/MS = 0.02 Da; fixed modifications: carbamidomethyl (C), variable modifications: acetyl (N-terminus) and methionine oxidation (M), and a false discovery rate (FDR) of 1%.

Protein quantification was done in MaxQuant using the LFQ algorithm[67] and the threshold for quantification set to 2 or more shared peptides. Statistical analyses were carried out in Perseus (ver. 1.5.1.6) and JMP Pro 11 Statistical Software. One-way analysis of variance (ANOVA) tests were carried out followed by t-test pairwise comparisons to determine significant effects using the Benjamini-Hochberg correction. Imputation of missing values was done in Perseus by replacing with random numbers drawn from the lower (left) boundary of a normal distribution. Gene Ontology (GO) annotation enrichment analysis was performed using the Functional Enrichment (FunRich) Analysis Tool, a bioinformatics software tool for functional and network analysis

developed by the Mathivanan lab [308]. FunRich uses hypergeometric tests for significance testing and the Benjamini-Hochberg procedure to decrease false discovery rate.

3.7 Results

3.7.1 Secretome Analysis of Radiation-treated H1299 cells

Inducible H1299 cells, with or without doxycycline treatment to induce stable p53 expression or not (hereon referred to as p53-wt or p53-null respectively), were treated with or without ionizing radiation (+IR or –IR, respectively). Secretomes from the 4 experimental conditions (p53-wt -IR, p53-wt +IR, p53-null -IR, and p53-null +IR) were collected 12-hours post-irradiation and then analyzed by LC-MS/MS. Prior to the secretome analysis, the p53 status of the cells was verified, specifically to ensure that we were using an isogenic paired set of H1299 cells, one of which is p53-inducible with the native cell line being p53-null (**Figure 12**). A total of 364 secreted proteins were identified

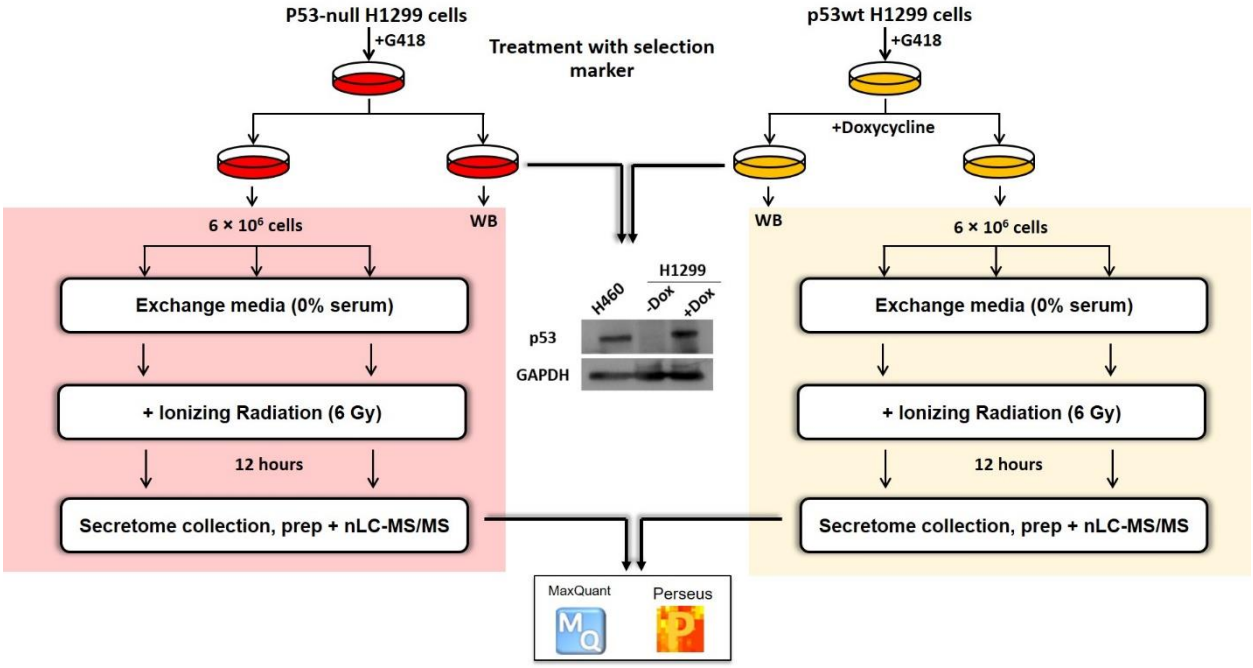


Figure 12 - H1299 Secretome Experimental Design

post-filtering from the 4 experimental secretomes in triplicate (**Figure 13A**). Out of the 364 identified proteins, we considered 163 proteins after filtration based on label-free quantification (LFQ)[67]. Thus, a protein should be quantified with an LFQ value in a minimum of 2 (in at least one condition) out of 12 replicates across all conditions. Out of the 163 quantified proteins, 123 were detected in all 4 groups **(Figure 13B).** Functional enrichment analysis was performed on the 163 quantified using the FunRich functional enrichment analysis tool. Enriched cellular components in the quantified secretome proteins included exosomes (~5.3 fold, p-value = 4.52×10^{-66}), lysosomes (~4.8 fold, p-value = 5.89×10^{-40}), cytoplasm (~2.1 fold, p-value = 6.73×10^{-29}), centrosome (~7.03 fold, p-value = 9.70×10^{-28}), cytoskeleton (~7.3 fold, p-value = 4.51 × 10-66), and mitochondria (~3.61 fold, p-value = 9.35×10^{-16}) as shown in **Figure 13C**. The molecular functions most enriched in this same dataset were related to chaperone activity *(~15.1 fold, p-value = 5.12×10^{-13})* which is one of the more prominent processes in autophagy, and lysosomal function (**Figure 13D)**. Notable secretory proteins (GO Cellular Component - **Figure 13C**) include protein FAM3C, which has been reported in other works as a potential blood based marker of autophagy[103, 161], C-type lectin domain family 11 member A (CLEC11A), Annexin A2 (ANXA2), and the heat shock protein HSP90B1.

3.7.2 Quantitative Analysis of the Secretome

The 163 quantified proteins from the four experimental conditions were analyzed to identify statistically significant changes in secretion. An analysis of variance (ANOVA) for the LFQ values identified 25 proteins (**Table 2)** that showed significant changes ($p<0.05$) in one or more of the experimental conditions tested. The data were transformed into a hierarchically clustered heatmap plotting the log2 LFQ protein levels in each condition

relative to the average log2 LFQs across all four experimental conditions (**Figure 14**). Comparing the protein intensity levels in individual samples to the average intensities

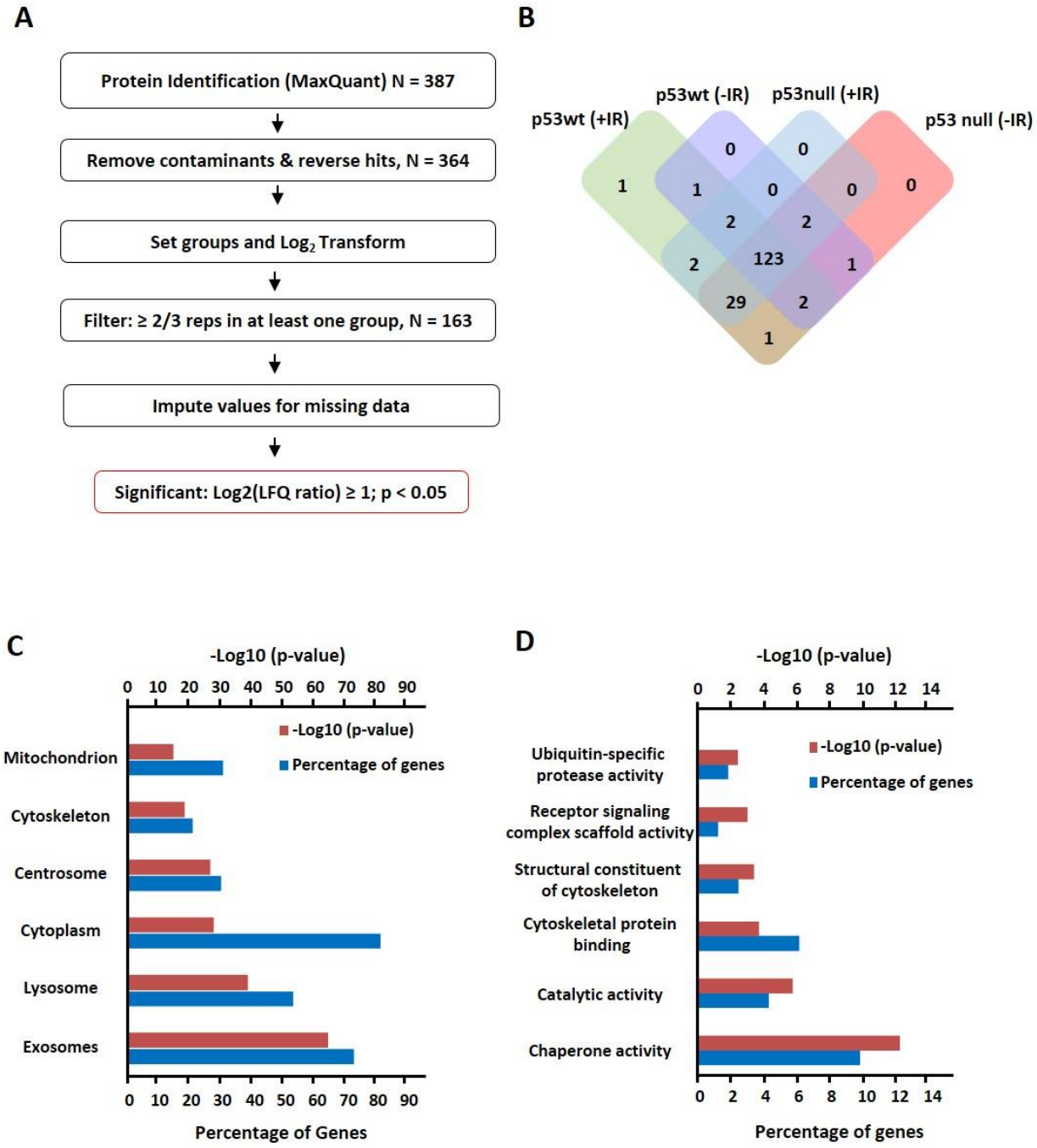


Figure 13 - Preliminary Qualitative Data Analysis of H1299 Secretome

(A) Summary of identified proteins and subsequent filtering Experimental Design of secretome analysis (B) Distribution of LFQ intensities of filtered proteins before imputation in Perseus (C) Gene Ontology Cellular Component analysis of 163 filtered proteins (D) Gene Ontology Molecular Function analysis of 163 filtered proteins

across all samples rather than a baseline intensity value is deliberate. We believe it is essential to define upregulation or downregulation in the context of clinical samples and in using the average intensities, we are able to define these as protein levels above or

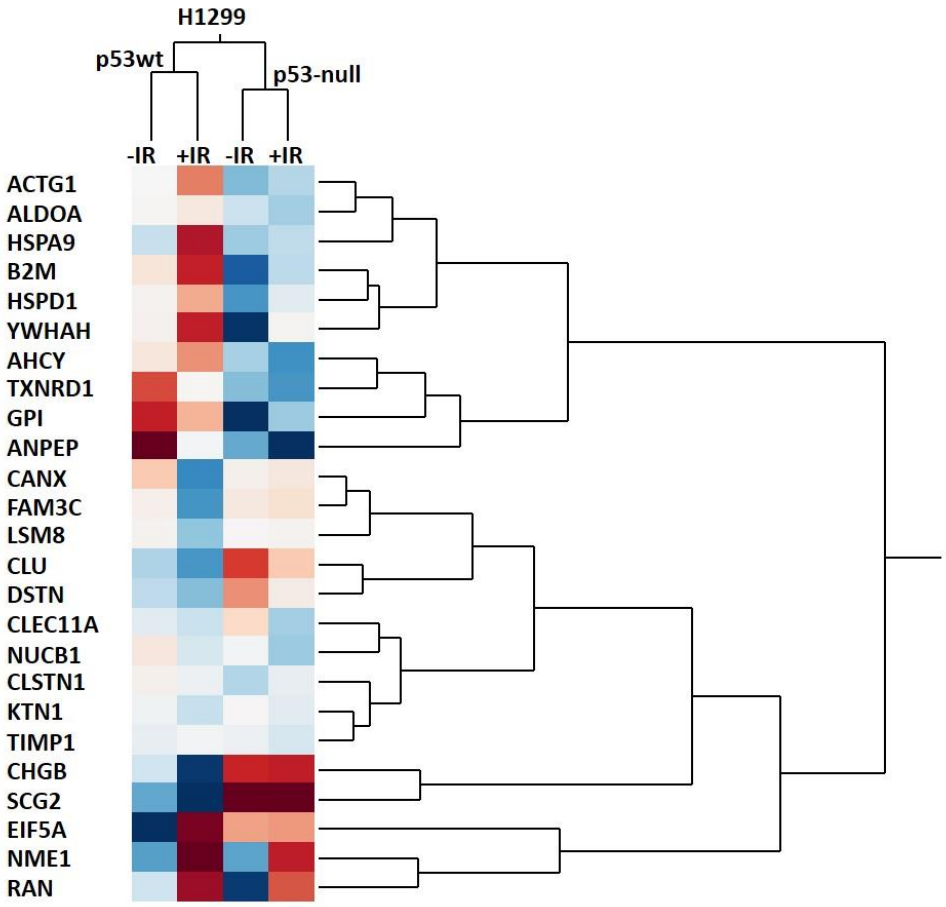


Figure 14 - Hierarchical Clustering of ANOVA significant proteins

Hierarchical clustering and heat map analysis of 25 differentially secreted proteins shows that p53 status drives the major changes in the secretome whereas radiation treatment drives changes within cells of the same p53 status.

below the population mean. The results of the hierarchical clustering indicates that the primary contributing factor to differential protein secretion is p53 expression status. Among the most prominent proteins in this group are Chromogranin B (CHGB), Secretogranin 2 (SCG2), Clusterin (CLU), and Destrin (DSTN) where protein level

differences are dictated by p53 status and irradiation has no significant effect on their secreted levels. Similar effects, but to a lesser extent, are observed for FAM3C, Calnexin (CANX), U6 snRNA-associated Sm-like protein LSm8 (LSM8), C-type lectin domain family 11 member A (CLEC11A), Kinectin (KTN1), Metalloproteinase inhibitor 1 (TIMP1), and Nucleobindin-1 (NUCB1). The levels of the remaining proteins in the secretome are more significantly impacted by irradiation than p53 status. These include Nucleoside diphosphate kinase A (NME1), Eukaryotic translation initiation factor 5A-1 (EIF5A), Aminopeptidase N (ANPEP), Glucose-6-phosphate isomerase (GPI), Thioredoxin reductase 1, cytoplasmic (TXNRD1), Adenosylhomocysteinase (AHCY), GTP-binding nuclear protein Ran (RAN), 14-3-3 protein eta (YWHAH), Beta-2-microglobulin (B2M), Stress-70 protein, mitochondrial HSPA9), Calsyntenin-1 (CLSTN1), Fructose-bisphosphate aldolase A (ALDOA), 60 kDa heat shock protein, mitochondrial (HSPD1), and Actin, cytoplasmic 2 (ACTG1). We also conducted a 2-way ANOVA model to determine the effects of p53 status and IR treatment as well as their interaction on the expression of the proteins in the secretomes.

Table 2 - Proteins that were found to be significantly differentially expressed (using ANOVA) in the secretomes of p53wt and p53-null H1299 cells before and after ionizing radiation treatment.

Gene names	Protein names	Sequence Coverage	-Log ANOVA p value
ACTG1	Actin, cytoplasmic 2	36.9	1.93048
AHCY	Adenosylhomocysteinase	13.4	1.59804
ALDOA	Fructose-bisphosphate aldolase A	26.6	2.73084
ANPEP	Aminopeptidase N	5.8	1.38492
B2M	Beta-2-microglobulin	28.2	2.06836
CANX	Calnexin	3	1.52497
CHGB	Secretogranin-1	5.5	3.66691
CLEC11A	C-type lectin domain family 11 member A	15.8	2.21017
CLSTN1	Calsyntenin-1	9.2	2.11829
CLU	Clusterin	6	1.45451
DSTN	Destrin	14.2	2.47716
EIF5A	Eukaryotic translation initiation factor 5A-1	10.9	1.48163
FAM3C	Protein FAM3C	17.2	1.78124
GPI	Glucose-6-phosphate isomerase	8.6	1.9333
HSPA9	Stress-70 protein, mitochondrial	4.4	1.85985
HSPD1	60 kDa heat shock protein, mitochondrial	11.5	1.79067
KTN1	Kinectin	2.9	3.76623
LSM8	U6 snRNA-associated Sm-like protein LSm8	19.8	1.31747
NME1	Nucleoside diphosphate kinase A	33.6	1.38612
NUCB1	Nucleobindin-1	13.6	1.55122
RAN	GTP-binding nuclear protein Ran	12.9	1.91709
SCG2	Secretogranin-2	7	4.41865
TIMP1	Metalloproteinase inhibitor 1	11.2	1.58608
TXNRD1	Thioredoxin reductase 1, cytoplasmic	4.3	2.14274
YWHAH	14-3-3 protein eta	16.7	1.59071

Of the differentially secreted proteins, GPI and TXNRD1 showed an interaction effect i.e. the effect of p53 status on the expression of GPI and TXNRD1 in the secretome depends on whether or not the cells received IR treatment.

Following the ANOVA analyses, we conducted pairwise comparisons of the secretomes of p53-wt and p53-null H1299 cells before and after ionizing radiation treatment using t-tests. We sought to identify proteins with the best potential to be population-based or individual/personalized biomarkers.

3.7.3 p53 Expression Status Promotes Differential Protein Secretion Before and After Ionizing Radiation Treatment

3.7.3.1 p53wt-IR vs p53-null –IR

15 proteins were found to be differentially secreted between H1299 p53-wt -IR and p53-null -IR cells (**Figure 15A)** and include CHGB, SCG2, DSTN, CLEC11A, CLU, GPI, ANPEP, phosphoglycerate kinase 1 (PGK1), hepatoma-derived growth factor (HDGF), protein disulfide-isomerase A4 (PDIA4), heterogeneous nuclear ribonucleoproteins A2/B1 (HNRNPA2B1), nuclease-sensitive element-binding protein 1 (YBX1), glyceraldehyde-3-phosphate dehydrogenase (GAPDH), aspartate aminotransferase (GOT2), and nuclear migration protein nudC (NUDC).

Among the differentially secreted proteins, CLU, DSTN, PDIA4, YBX1, CLEC11A, HNRNPA2B1, NUDC, and HDGF were overexpressed by ≥ 2-fold in the secretome of p53-null cells compared to p53-wt cells whereas secretion of GAPDH, PGK1, GOT2, ANPEP, and GPI was higher in p53-wt cells by ≥ 2-fold than in p53-null cells. HDGF, also known as HMG-1L2, is a known repressor gene that is involved in DNA binding and cell

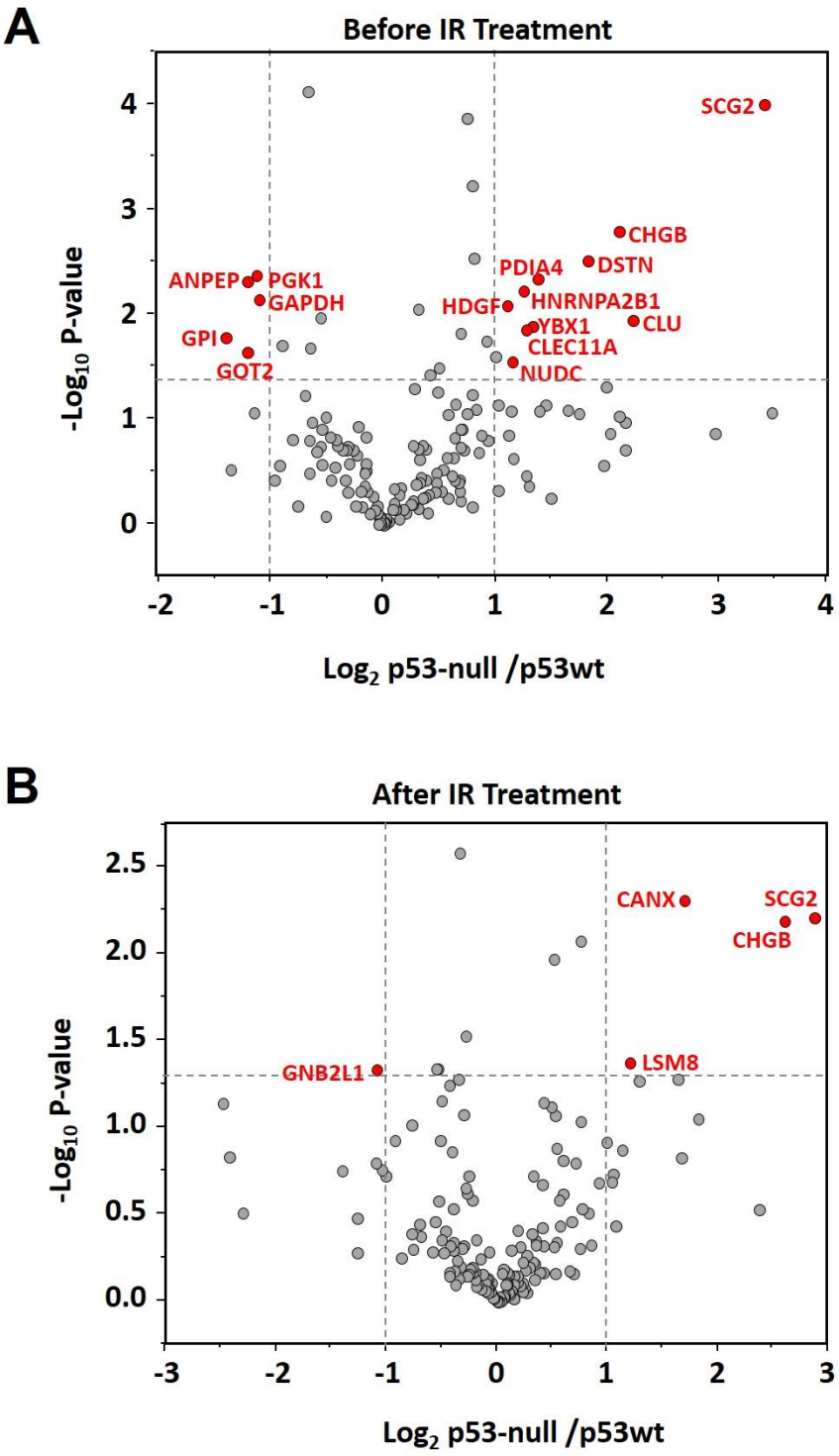


Figure 15 - Volcano Plots showing IR and TP53 effects on H1299 Secretome

Differential protein secretion for H1299 cells (A) un-irradiated and (B) irradiated. Red data points with gene labels indicate proteins with $\log_2$ difference ≥ |1| and p-value < 0.05

growth [309] and has been demonstrated to have the ability to stimulate growth in

fibroblasts and some liver cancer cells [310].

3.7.3.2 p53wt +IR vs p53-null +IR

In the pairwise comparison between p53-null and p53-wt cells treated with IR (**Figure 15B)**, proteins over-secreted in p53-null cells by ≥ 2 fold include LSM8, CANX, CHGB and SCG2 whereas guanine nucleotide-binding protein subunit beta-2-like 1 (GNB2L1) was over-expressed by >2 fold in p53-wt cell secretomes. Calnexin is known to contain a signal peptide (1 – 20) which by manual inference, is responsible for its secretion and is reported to be a molecular chaperone [311, 312]. The significant differences in CHGB and SCG2 secretion between p53-wt and p53-null H1299 cells were observed both before and after ionizing radiation, suggesting a potential role for these proteins as population based biomarkers for patient stratification before treatment particular as a function of autophagy response.

3.7.4 Association between Significant Proteins and Patient Survival

Kaplan-Meier plots for the significant proteins were performed using the online kmplotter by Gyorffy et al [313]. Of the differentially secreted proteins, TXNRD1 and GPI were found to be significantly associated with survival in lung cancer patients **(Figure 16)**.

3.8 Discussion

Approximately 10 – 15% of all proteins are secreted into the extracellular matrix via classical and non-classical pathways [93, 222]. The classical pathways involve the endoplasmic reticulum (ER)/Golgi apparatus. Here a protein undergoing synthesis and having the appropriate signaling sequence, generally at the N-terminus, results in the

transport of the ribosome to the ER membrane where the growing peptide chain is

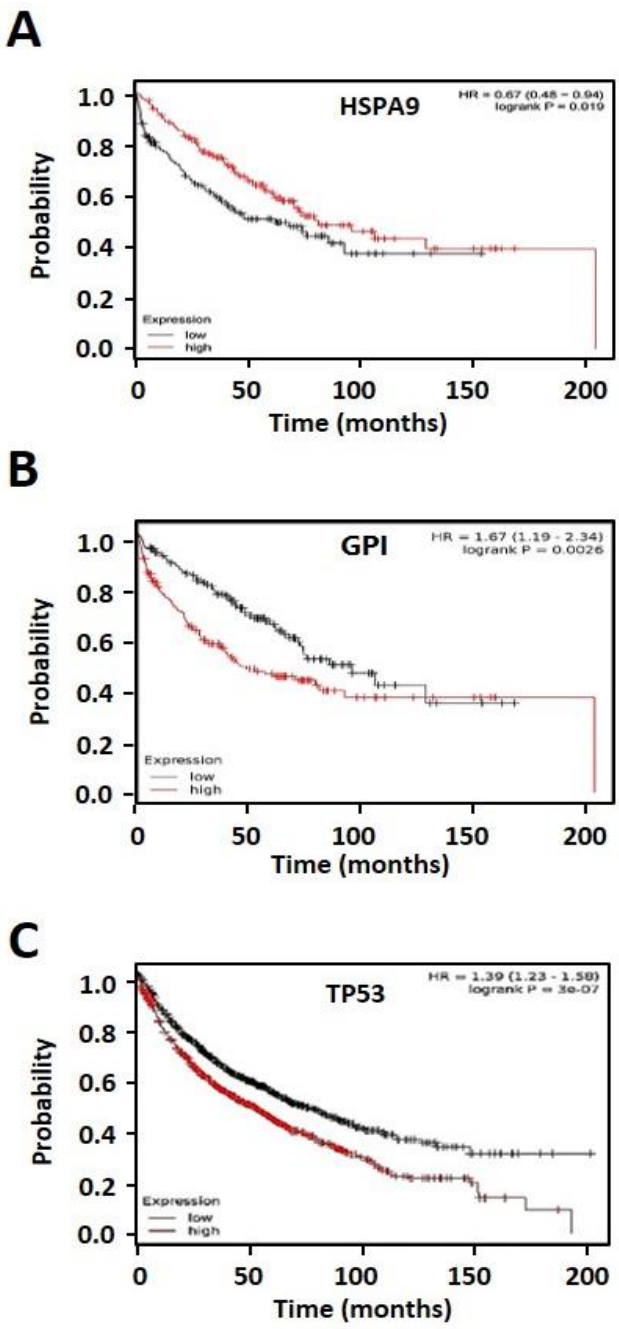


Figure 16 - Kaplan Meier survival plots for selected proteins

Association between GPI levels (A) and HSPA9 levels (B) on probability of survival plot in lung cancer patients with or without radiation but not chemotherapy; n=310. (C) Association between p53 levels and patient survival in lung cancer; n=1926. Analysis conducted using Kaplan-Meier Plotter at kmplot.com Ref: Gyorffy B, Surowiak P, Budczies J, Lanczky A. PLoS One, 2013 Dec 18;8(12):e82241.

translocated across the ER membrane [223]. The newly synthesized proteins are mostly found in tiny vesicles in the ER which fuse with the Golgi apparatus where secretory proteins are eventually sorted into transport vesicles that migrate to the cell membrane and are released via exocytosis into the extracellular environment [224]. The non-classical secretory pathway was proposed and described due to the discovery of secreted proteins that lack the N-terminal secretory signal sequence, are excluded from the ER/Golgi machinery and are not affected by inhibition of the ER/Golgi – dependent secretory mechanism [314]. Both pathways are involved in cells differentially expressing p53 [300, 315] and under different environmental stimuli including ionizing radiation [300].

The secretory pathways associated with autophagy have been investigated in yeast and in cancer cells. The role of autophagy in protein secretion borders on both classical and nonconventional secretion and this has potential downstream ramifications for biomarker discovery [161, 221, 225, 304]. In this study, we quantitatively characterized the H1299 secretome as a function of p53 expression and ionizing radiation to potentially identify proteomic signatures to discriminate between the response to ionizing radiation in the context of protective and non-protective forms of autophagy. As demonstrated in many studies, proteomics affords the ability to view the entire proteome, for internal cellular signatures, and secretome, for surrogate plasma biomarkers [94, 103, 161, 301, 302, 316–318].

Lung cancer has been associated with abnormal protein secretion and evidence shows that interactions between lung cancer cells and their microenvironment contribute to disease progression [302]. Such communication involves the secretion of proteins and exosomes, most of which contain proteins, which play a role in cancer cell – cell interaction and which may either promote tumor progression or serve some anti-tumor

capacities within the body [93, 223]. Autophagy as a cellular mechanism has been shown in various studies to contribute to chemotherapy resistance [270, 319]. This protective form of autophagy appears to be associated with the presence of functional p53 in the case of radiation [126]. During the course of autophagy, protein turnover occurs in response to the applied stress. The findings of our GO Biological Process analysis is consistent with our expectation of cells undergoing stress from ionization radiation and in which autophagy is induced. The biological processes most enriched in our dataset include protein folding *(p-value = 0.0008) (e.g. HSPD1, CANX, and PPIA)*, protein metabolism *(p-value = 1.0 × 10^{-11}) (e.g. CTSZ, FAM3C, and RAN)*, and energy pathways (*p-value = 1.0 × 10^{-4}) (e.g. NME1, GPI, PGD, and PGAM1)*.

3.8.1 Cathepsin D (CTSD)

Our current work has identified a number of proteins reported in different lung cancer studies and that may be closely associated with autophagy. Amongst these proteins are CANX, Bifunctional purine biosynthesis protein PURH (ATIC), caspase 3 (CASP3), cathepsin B (CTSB) and cathepsin D (CTSD). Cathepsin D (CTSD) was only quantified in p53-wt +IR and p53-null -IR cells meaning levels in p53-wt -IR and p53-null +IR were not sufficient for quantitation by the LFQ algorithm. Consistent with reports of cytoprotective autophagy in cells with p53, the e2levated secretion of CTSD in p53-wt +IR cells and its decreased presence in p53-null +IR cells may be an indication of its role in protective autophagy, as reported in a study in which dichloroacetate was used to induce protective autophagy in LoVo cells [320]. Koukourakis et al. have also reported poor prognosis of NSCLC associated with increased CTSD expression [321]. CTSB was identified in our dataset but filtered due to our stringent criteria. It has been associated

with the cleavage of Disabled-2 (DAB2) to promote autophagy in cells treated with Transforming growth factor-β (TGF-β) and when, in that study, CTSB was inhibited cells became apoptotic in the long term [322].

In determining the potential of these differentially secreted proteins to serve as biomarkers to predict benefit from autophagy modulation in combination with chemotherapy or radiation therapy, we reasoned that proteins consistently overexpressed or under-expressed based on p53-status and independent of radiation treatment are more likely to serve as population-based biomarkers before treatment. That is, the secretion of these proteins may contribute to prediction of benefit from treatment and/or help to stratify patients in clinical trials using autophagy inhibition as a tumor sensitization strategy.

We identified CHGB and SCG2 to be consistently overexpressed in p53-null secretomes at very high levels while GIP, and TXNRD1 were consistently overexpressed in the p53-wt secretome. Therefore, CHGB, SCG2, GPI, and TXNRD1 with consistently high levels have the potential to serve as population-based markers. FAM3C, CANX, and LSM8 levels were consistently downregulated in the p53-wt +IR secretome while EIF5A levels were consistently low in p53-wt –IR secretomes; therefore these proteins may serve a role in distinguishing between patients following radiation treatment. Patients whose FAM3C, CANX, and LSM8 levels decrease and those whose EIF5A levels increase following ionizing radiation treatment may be found to have tumors that respond to autophagy inhibition. The secretory profiles of these proteins are shown in **Figure 17**.

3.8.2 Chromogranin B and Secretogranin 2 (CHGB and SCG2)

The chromogranin-secretogranin proteins are known to be associated primarily with neuronal cells[323]. SCG2 has been shown to be involved in angiogenesis and proliferation

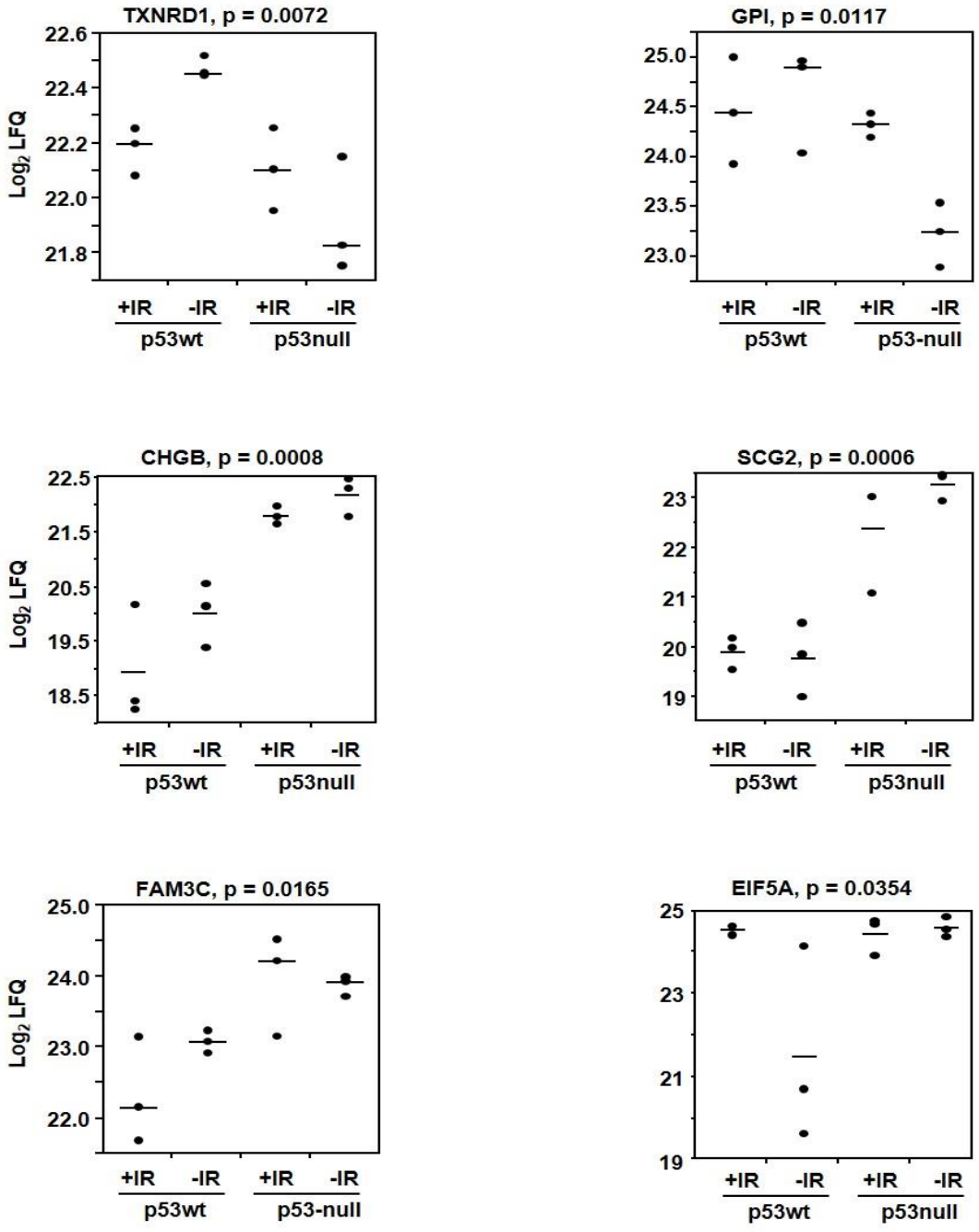


Figure 17 – H1299 secretory profile of selected proteins across the different conditions

Log2 LFQ values were plotted across the four conditions (IR treated and untreated p53-null and p53-wt H1299 for proteins CHGB, EIF5A, FAM3C, GPI, TXNRD1, and SCG2

in endothelial cells through an anti-apoptotic mechanism[324]. SCG2 is cleaved into 2 chains: the 33-amino acid secretoneurin and manserin. Secretoneurin has been shown in in vivo studies to induce neovascularization in cornea and stimulate proliferation of serum-starved human umbilical vein endothelial cells (HUVECs)[324]. With angiogenesis being a prominent feature of tumor progression, it is not unexpected that SCG2 would be overexpressed in the secretomes of H1299 p53-null cells which undergo non-protective autophagy in response to radiation. Secretoneurin has also been shown to contribute to the growth and protection of neuronal cells [325]. This neuroprotection has been observed in primary cortical cell cultures during oxygen or glucose starvation[325]. Glucose starvation is known to induce autophagy. Chromogranin B (CHGB) is also known to be cleaved into the peptides GAWK and CCB which have been found to be elevated in pancreatic islet-cell and bronchial tumors while secretoneurin from SCG2 has been found to be elevated in gastroenteropancreatic neuroendocrine tumors, small cell lung tumors and pheochromocytomas[323, 326, 327]. There are reports of using neuroendocrine differentiation as a prognostic marker of NSCLC disease response to therapy[328, 329]. Given that the chromogranins and secretogranins are closely related to neuroendocrine differentiation, CHGB and SCG2 may have potential prognostic utility in NSCLC.

3.8.3 Glucose-6-phosphate isomerase (GPI)

The differential levels of GPI and TXNRD1 was found to be associated with both p53 status and radiation treatment status. GPI, also known as autocrine motility factor (AMF), was overexpressed in the secretomes of p53-null cells by ~2.8 fold following irradiation. However, in p53-wt cells, levels of GPI in the secretome were not affected by IR treatment, implying that p53 status impacts the response of GPI following radiation as

confirmed by the least squares model including the interaction of p53 status and IR treatment (*p-value=0.0276*). The level of GPI in the p53-null secretome before irradiation was the lowest and significantly different from the other 3 conditions. GPI has different functions in the intracellular and extracellular compartments. GPI has been identified in various cancers including renal cell carcinoma, endometrial cancer, and breast cancer[330, 331]. In the cytoplasm, GPI acts in glycolysis in the conversion of glucose-6-phosphate to fructose-6-phosphate and vice versa. In the extracellular space, it may act to induce secretion of immunoglobulin or facilitate neuronal survival. The differential secretion of GPI between p53-wt and p53-null cells pre-irradiation suggests that it could prove to be useful as a biomarker to predict potential patient benefit from IR and autophagy inhibitor combination treatment before the first dose of ionizing radiation. Using the online KM Plotter[313], GPI was also found to be significantly associated with probability of survival (*log Rank p = 0.0026*) in lung cancer patients who may or may not have been treated with IR but not chemotherapy **(Figure 16)**.

3.8.4 Thioredoxin reductase 1 (TXNRD1)

Studies have shown that TXNRD1, identified in this study to be significantly differentially secreted only in untreated p53-wt cells, has different isoforms that may display different functions. Isoform 4 promotes transcription of estrogen receptors while isoform 5 is involved in interferon beta and retinoic acid induced apoptosis [332, 333]. We determined from this study that the effect of IR on TXNRD1 secretion is dependent on the p53 status *(p-value = 0.0157)*. Like GPI, the significant difference in TXNRD1 secretion between p53-wt cells and p53-null cells before IR treatment suggests that it may be used as a diagnostic marker (before patients undergo radiotherapy) rather than a prognostic marker

given that the levels after radiation are not significantly different. TXNRD1 has been reported to have a prognostic association in breast cancer in a study that found that high levels of TXNRD1 RNA in node-negative breast tumor samples correlated with a higher risk of metastasis [334].

3.8.5 Protein FAM3C

In our study FAM3C, also referred to as interleukin-like EMT inducer (ILEI), was expressed ~1.5 fold higher in p53-null cell secretomes compared to p53-wt cells before radiation whereas there was a ~3.2-fold increase in expression in p53-null cell secretomes compared to p53-wt after radiation treatment. FAM3C is a member of the FAM3 family of proteins and its gene codes for secreted proteins known to play a role in autophagy [161, 335]. A number of studies have also found the gene to be overexpressed in esophageal, colorectal and pancreatic cancer [103, 335, 336]. A study by Ling-Zhi et al. indicates that FAM3C is overexpressed in the exosomes of NSCLC patient plasma compared to healthy patients and is also linked to cell migration in K-ras mutant cells [337]. Its role in cell migration may be related to metastasis and poor prognosis [337]. Overexpression of FAMC3in H1299 p53-null cells, which represent the non-protective form of autophagy, is consistent with previous studies that have linked FAM3C to poor prognosis in colorectal cancer and metastasis in lung cancer cells [336, 338].

3.8.6 Calnexin (CANX)

Together with FAM3C, CANX was shown to be overexpressed principally in the p53-null secretome following radiation treatment. Calnexin (CANX) is a calcium-binding membrane protein of the endoplasmic reticulum [339]. It is thought to be a chaperone in

protein translocatin from the ER to the extracellular membrane [339] and is known to be cleaved during regular growth into 2 fragments in yeasts [340]. This cleavage of calnexin has been found to depend on autophagy in a study that showed CANX processing to be defective following knockdown of autophagy-related genes [340]. In yeasts, autophagy occurs primarily in the vacuoles and CANX has been showed to co-localize with the vacuole following autophagy induction via nitrogen starvation [340]. CANX has also been reported in a recent study as a novel sero-diagnostic marker of lung cancer [341]. Its role as a chaperone protein is consistent with this study in which we have characterized autophagy associated secretion induced by ionizing radiation in the H1299 NSCLC cell line. To our knowledge, this is the first report of differential secretion of calnexin in isogenic cell lines of NSCLC differing only in p53 expression and functional autophagy status. The association of calnexin with autophagy combined with its recent report as a potential sero marker of lung cancer suggests that it could potentially serve as a marker for the autophagic response during chemo- or radiotherapy.

3.8.7 U6 snRNA-associated Sm-like protein LSM8

LSM8, quantified in all conditions except p53-wt -IR and overexpressed in p53-null +IR secretome compared to p53-wt +IR cells, is widely viewed as a component of the spliceosome. In a study of different cell lines including mammary and lung carcinoma cells, the spliceosomal machinery was demonstrated to be associated with autophagy where depletion of key spliceosome genes SNRPE or SNRPD1 resulted in autophagy induction [342]. LSM8 was found to be secreted in similar quantities in p53-null cells before and after radiation; however, secretion was elevated after irradiation in p53-wt cells. This may indicate a role of LSM8 in protective autophagy after IR treatment. LSM8 has been

reported to be directly involved in P-body regulation, its overexpression resulting in a concomittant decrease in P-body structures while depletion of LSM8 results in an increase in P-body structures [343]. LSM8 is believed to interact with the other LSM proteins in the cytoplasm [344, 345]. Persistent higher expression of LSM8 in H1299 p53-null secretomes which represent the model of non-protective autophagy which is unresponsive to modulation may also imply a role in non-protective autophagy when it is upregulated before IR treatment.

3.8.8 Eukaryotic translation initiation factor 5A-1 (EIF5A)

EIF5A is the only protein in humans known to have the amino acid hypusine in its structure [346, 347]. The modification of lysine to hypusine has been shown to be important for cell viability [347]. Two gene encode for two isoforms of the protein; EIF5A1 and EIF5A2 with 84% sequence homology. EIF5A has been shown to be directly or indirectly involved with the p53 signaling pathway in lung cancer cells [348] as well as the regulation of autophagy in drosophila [349]. EIF5A2 has also reported to be important for epithelial-mesenchymal transition in NSCLC and overexpressed in stage 1 NSCLC patients with poor prognosis [350, 351]. EIF5A has been found to promote the metastatic phenotype of pancreatic cancer through modulatory activity on various proteins [352]. In our study, EIF5A is virtually absent prior to irradiation in p53-wt cells whereas it is increased by more than 3-fold following irradiation. Conversely, EIF5A levels in the p53-null cells do not change in response to irradiation. The marked upregulation of EIF5A in p53-wt cell secretomes following radiation treatment suggests a role in the cellular response to radiation in p53-wt cells and makes EIF5A a potential predictor of functional autophagy type in response to radiation.

A comparison of our data with the protein list from the study by Daub et al [353] where 23 NSCLC cell lines were profiled reveals overlap between proteins identified in this study and their study (**Figure 18**). We also found a number of proteins that may prove to be

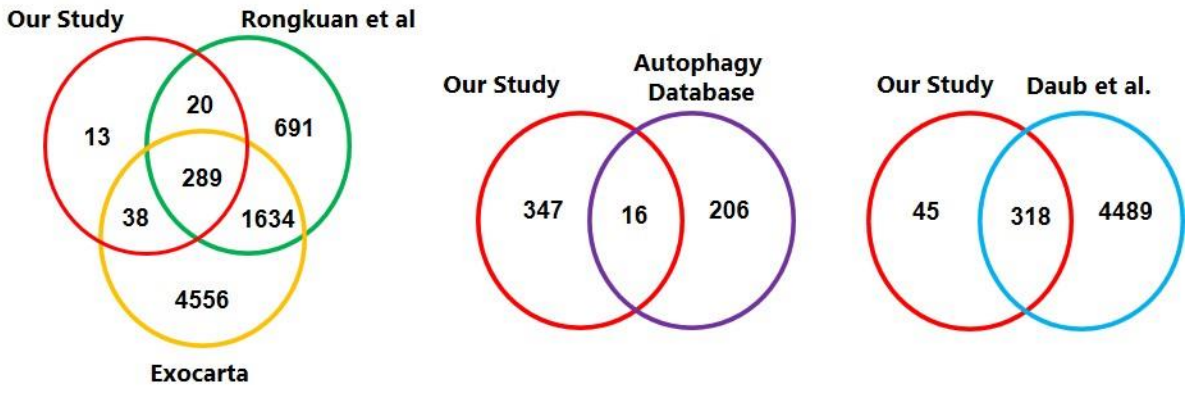


Figure 18 - Comparative Venn Diagrams

Comparison of proteins identified in our study to the Exocarta database and a recently published study by Rongkuan et al. to the autophagy protein database and to a study of 23 NSCLC cell lines by Daub et al.

peculiar to the H1299 cell line that were not identified in the Daub et al. study including H2AFV, HIST1H2BK, TMA7, PKM, SUMO2 and RHOBTB3. Similar comparisons of our data with the proteins in the exocarta database as well as the protein list obtained by Hu et al. who used TMT-labeled samples to examine the secretome of the H1993 metastatic NSCLC cell line shows substantial overlap between the identified proteins (**Figure 18**) [302]. The overalp between these findings supports the validity of the label-free approach we used in our study.

3.9 Conclusions

The current study has identified, among many proteins, CHGB, SCG2, GPI, TXNRD1, FAM3C, CANX, and EIF5A as potential diagnostic and prognostic blood based biomarkers in NSCLC. While FAM3C and CANX have been reported as being

overexpressed in some cancers including NSCLC, for many of the proteins identified in our study, this is the first report of overexpression in a NSCLC (H1299) secretome in the context of autophagy and ionizing radiation treatment. We determined that CHGB, SCG2, GPI, and TXNRD1 have potential as population-based biomarkers (for patient stratification) due to their consistent overexpression as a function of p53 status and not radiation treatment whereas FAM3C, CANX, and EIF5A hold potential as personalized biomarkers to predict patient response following ionizing radiation treatment due to their differential secretion in radiation treated versus untreated cells. On the basis of this study, plans are in place to conduct larger studies to optimize the sensitivity and specificity of these proteins in NSCLC and the validity of using these proteins as stratification tools to determine which patients may benefit from autophagy modulation in combination with chemo- and/or radiotherapy.

It is necessary to provide a critical caveat to these findings. Although our stated goal in this work was to identify proteins that could potentially be utilized as biomarkers to distinguish between the cytoprotective and non-protective forms/functions of autophagy, we fully recognize that the observed profile of proteins in the secretome may be related primarily to the p53 status of the tumor cell models. Identifying potential linkages to autophagy will require further studies where the levels of these proteins are genetically modulated to determine their impact on autophagic function.

Chapter 4:
Glycoproteomics Characterization of the SILAC-Labeled HepG2 Secretome as a Platform for the Generation of Stable Isotope Labeled Plasma Proteins

4.1 Introduction

The development and implementation of untargeted (i.e., global) quantitative proteomics strategies has transformed the study of complex biological systems. With regard to clinical biospecimen analysis, label free (i.e., spectral counts, LFQ) and isobaric tagging (i.e., iTRAQ and TMT) have played a predominant role in comparative un-targeted proteomics studies. The successful implementation of both approaches requires minimal sample preparation variation including total protein determination, consistent digestion efficiency, and control of chemical modifications (e.g. alkylation, oxidation, isobaric tagging, etc.). Although it is possible to minimize these sources of variation, it can become increasingly difficult scaling up to larger numbers of samples. Finally, as the number of samples increases and the desire to venture outside of traditional peptide-level quantification (e.g., complex carbohydrates), so too does the difficulty in maintaining quantitative precision across samples. Furthermore, the recognition that large amounts of biologically important information in the form of PTMs is not captured.

Plasma remains an important clinical biospecimen for comparative proteomics analysis but it is one of the most complex proteomes to quantitatively study with $>10^{10}$ dynamic

range of concentration, disproportionate levels of high abundant proteins such as albumin and immunoglobulins, heterogeneous protein complexes (e.g. lipoproteins), and protein glycoforms [82, 130]. Although many comparative plasma proteomics studies have focused on identifying low abundant proteins using extensive depletion and fractionation, many of the major plasma proteins hold important biological functions and potential diagnostic value. In fact, many high- and mid-level plasma proteins have been identified as candidate biomarkers for cancer [354]. For instance, some of these proteins (e.g., CA-125, APOA1, B2M, TF, and ALB used in the OVA-1 test) have been used for FDA-approved diagnostics and prognostics applications. An added dimension of complexity and opportunity lies in plasma protein PTMs. Glycosylation is very common for PTM, particularly for major plasma proteins, and dysregulated forms of glycoproteins are believed to be potentially important. Despite their importance, there are limited quantitative approaches to address this potentially important aspect of comparative plasma proteomics measurements.

The extension of the SILAC quantitative proteomics strategy to biological samples termed Super-SILAC has improved the precision and accuracy of LC-MS/MS analysis of biospecimen [80]. We have showed in a proof-of-principle study that this approach makes possible the quantitative analysis of changes in the plasma proteome [82]. In that study, the HepG2 cell, shown in the literature to secrete many of the major plasma proteins, was used [82, 355]. The use of the HepG2 cell line secretome for the generation of stable isotope labeled (SIL) intact proteins is however, not limited in applicability to only plasma analyses. There are ongoing studies in our lab in which we show extended applicability of the SIL labeled secretome to enzyme digestion kinetics, biospecimen degradation, and

interaction studies. These studies are respectively in pursuit of sequences that influence protein digestion and peptide formation rates, factors such as collection strategies affecting biospecimen integrity, and novel binding motifs of heparin and heparan sulfate. Glycosylation is one of the most important PTMs in cellular protein production [107, 356, 357]. The importance of glycosylation in normal cell function as well as in disease is well established in the literature [358–360]. Glycosylation introduces carbohydrate (or sugar) moieties onto certain amino acids in the protein structure. Glycans may be N-linked or O-linked referring to the atom of the amino acid to which the glycan is linked to the protein. To use the HepG2 cell line as the platform for the generation of intact SIL proteins to be used as standards for mass spectrometry analyses of plasma and other biospecimens, it is essential to characterize the complement of secreted proteins and to determine similarities and differences compared to plasma especially regarding PTMs such as glycosylation. The results of our previous study [82] suggested potential differential glycosylation between plasma and the HepG2 secretome as was the case with alpha-1-glycoprotein 2 (ORM2). In this study, we present an in-depth glycoproteomics characterization of the HepG2 secretome.

4.2 Experimental

4.2.1 Reagents

SILAC DMEM (arginine- and lysine-free high glucose Dulbecco's Modified Eagle Media) was purchased from Thermo Scientific. $^{13}C_6$ L-arginine-HCl (>98%) and $^{13}C_6$ L-lysine-2HCl (>98%) were purchased from Cambridge Isotopes Laboratories. Dialyzed Fetal Bovine Serum was purchased from Gibco. Penicillin, streptomycin, and amphotericin B were obtained as a 100X solution from Invitrogen. LC-MS grade acetonitrile (ACN) and

water were obtained from Burdick and Jackson. LC-MS grade formic acid, Iodoacetamide (IAA), and Dithiothreitol (DTT) were purchased from Sigma-Aldrich. Reagent grade Tris base and 12 M hydrochloric acid were obtained from Fisher Scientific, and trypsin gold was obtained from Promega.

4.2.2 Cell Culture Conditions

HepG2 cells were obtained from the American Type Culture Collection (ATCC, HG-8065, Manassas, VA), and cultured in high glucose SILAC DMEM supplemented with 10% dialyzed Fetal Bovine Serum (FBS), 50 mg $^{13}C_6$ L-arginine-HCl (>98%), 50 mg $^{13}C_6$ L-lysine-2HCl (>98%), and 1% 100 U/ml Penicillin G sodium/1% 100 µg/ml Streptomycin sulfate at 37°C and 5% CO_2 for nine population doublings. Cells were then transferred into freeze media and stored in liquid nitrogen until needed. Heavy SILAC-labeled HepG2 cells were seeded in three sets of 3 × 10 cm dishes at ~200,000 cells in 10 mL of media and grown to ~85% confluency over 48 hours. At ~85% confluency, the 10% dialyzed FBS culture media was aspirated from all plates, cells washed with 2 × 5 mL of Phosphate Buffered Saline 1X pH 7.4 (Gibco) and then once with 5 mL serum-free DMEM [361]. A final 10 mL volume of serum-free SILAC heavy DMEM was added to the 9 plates. All plates were then incubated at 37°C. The serum-free media containing secreted proteins (secretomes) were collected into 3 × 15 mL tubes after 24-, 48-, and 72-hour incubation times and stored at -80°C. Cells from the plates were then harvested, pelleted by centrifugation at 1500 rpm for 5 minutes at 8°C, washed with PBS, and stored at -80°C until use.

4.2.3 Annexin V/Propidium Iodide staining

Quantification of apoptotic cells via flow cytometry was performed per the manufacturer's instructions (Annexin V - FITC apoptosis detection kit, BD Biosciences, 556547). Flow cytometry analysis was performed using BD FACSCanto II and BD FACSDiva software at the Virginia Commonwealth University Flow Cytometry Core Facility.

4.2.4 Proteomics Sample Preparation

Secretome Only

The collected media (secretomes) samples were removed from -80^{o}C, thawed, and centrifuged at 2500 rpm for 5 minutes at 8^{o}C to remove cellular debris. The supernatant was then collected and concentrated in 3 kDa MWCO filters at 6900 × g for 30 minutes at 8^{o}C in 5 mL aliquots. Centrifugation was repeated until there was ~500 µL of concentrated secretome in the filter. To remove the phenol red color, 3 × 500 µL of Tris HCl buffer pH 8.1 was added to each secretome and centrifuged at 6900 × g for 30 minutes at 8^{o}C. The concentrated secretomes were then transferred into 10 kDa MWCO filters and washed with 2 × 400 µL of Protease Inhibitor (PI) in Tris HCl pH 8.1 buffer by centrifugation at 15,000 × g for 10 minutes. Remaining secretome in the filters were then made up to 300 µL with PI in Tris HCl pH 8.1 buffer and transferred into 1.5 mL centrifuge tubes. 300 µL of PI in Tris HCl pH 8.1 buffer was pipetted into each filter, vortexed briefly, and transferred into the same corresponding 1.5 mL centrifuge tube. The biological replicates of each time point were then pooled to give 3 samples: 24h, 48h, and 72h. Total protein concentration was determined at 280 nm using a BioTek Synergy H1 fitted with a Take3 Plate. 6 × 100 µL aliquots of each sample was then pipetted into 6 new 10 kDa MWCO filters secretomes and were processed for proteomics analysis using the FASP

method [307]. For each time point, samples in three filters were subjected to PNGase F digestion preceding trypsin digestion while samples in remaining three filters served as controls in which no PNGase F digestion was carried out before trypsin digestion. Samples were processed concurrently. Briefly, each filter was centrifuged at 15,000 × g for 10 minutes. Samples in the 10 kDa MWCO filter were then reduced with 50 µL of 50 mM DTT (to give a final concentration of 25 mM DTT) by incubation for 45 mins at 56^{o}C. Reduced samples were then alkylated with 100 µL of IAA to a final concentration of 40 mM IAA at room temperature in the dark. Reduced and alkylated samples were centrifuged at 15,000 × g for 10 minutes and subsequently rinsed with 350 µL of Tris-HCl pH 8.1. PNGase F digestion was carried out based on protein concentration with 5 µL, 6 µL, and 7 µL PNGase F for 24, 48, and 72 h samples respectively. Samples were incubated at 37^{o}C overnight. Control samples were incubated without PNGase F. Trypsin digestion was performed with 5.75 µL, 6.91 µL, and 8.03 µL of 1 µg/ µL Trypsin Gold (Promega) solution corresponding to 50:1 protein to enzyme ratio. Samples were incubated overnight at 37^{o}C and then digestion terminated with 200 µL of 0.1% acetic acid (Fluka Analytical).

Secretome - Plasma Combinations

Previously concentrated secretome samples and frozen human plasma samples were removed from -80^{o}C, and thawed at room temperature. Total protein concentration was determined at 280 nm using a BioTek Synergy H1 fitted with a Take3 Plate. 35.6 µL aliquots of 24 h, 48 h, and 72 h secretomes were then combined with 22.7 µL (150 µg) of plasma in different centrifuge tubes. Different weight-to-weight combinations (1:1, 1:10, and 10:1) of 72 h secretome and plasma were also prepared. Each secretome-plasma

combination sample was reduced with 50 mM DTT (to give a final concentration of 25 mM DTT) by incubation for 45 mins at 56^{o}C. Reduced samples were then alkylated with 80 mM IAA to a final concentration of 40 mM IAA at room temperature for 30 minutes in the dark. Reduced and alkylated samples were centrifuged at 15,000 × g for 10 minutes and subsequently rinsed with 350 µL of Tris-HCl pH 8.1. Trypsin digestion was performed with 5 µL, of 1 µg/ µL LC-MS/MS Trypsin (Pierce) solution. Samples were incubated overnight at 37^{o}C and then digestion terminated with 200 µL of 50 mM acetic acid (Fluka Analytical).

Peptides were then desalted using C18 StageTips (Thermo) according to the instructions in the manufacturer kit. Briefly, 24 × 10 µg capacity StageTips were washed with 20 µL of 80% acetonitrile in 5% formic acid by centrifugation at 500 × g for 5 minutes. The StageTips were then equilibrated with 20 µL of 5% formic acid. 15 µL of each peptide sample was loaded onto a StageTip and centrifuged at 1000 × g for 5 minutes. StageTips were then washed with 20 µL of 5% formic acid to remove salts. Peptides were then manually eluted with 2 × 20 µL of 80% acetonitrile in 5% formic acid using the CombiTips provided with the Thermo StageTips kit. Eluted peptides were then dried in the speedvac and reconstituted in 20 µL of 2% acetonitrile in 0.1% formic acid. The reconstituted peptide samples were then transferred into LC vials and analyzed by nanoLC-MS/MS.

Lectin Enrichment and PNGase F Digestion

Two 1 g solutions of a 1:1 w/w mixture of the 72 h secretome and plasma were prepared in two different tubes for lectin enrichment. The secretome-plasma mixtures were diluted with the 5X binding buffer from the Thermo glycoprotein kit in a 4:1 protein buffer ratio.

100 µL each of WGA and Con A lectin slurries were pipetted into separate columns, labeled, and centrifuged at 1000 × g for 1 minute. Unless otherwise stated, centrifugation of lectin columns was done at 1000 × g for 1 minute. The columns were then rinsed three times with 1X binding buffer via centrifugation. The diluted protein samples were then transferred into the labeled lectin columns and incubated at room temperature for 60 mins with mild shaking on a vortex. The columns were then centrifuged and the flow-through (FT) collected and transferred into centrifuge tubes labeled WGA FT or Con A FT. 2 × 200 µL of 1X binding buffer was added to the columns, centrifuged and the eluates added to the FT tubes. The columns were then incubated with 200 µL of 1X binding buffer for 15 mins at room temperature. Columns were centrifuged followed by a wash with 200 µL of 1X binding buffer. Eluates were combined in the FT tubes. 200 µL of glycoprotein elution buffer was then added to each column and incubated at room temperature for 20 mins. Columns were subsequently centrifuged and transferred into tubes labeled WGA binders or Con A binders. Incubation with the elution buffer was repeated for 20 mins following which columns were centrifuged and the glycoprotein eluates transferred into their corresponding labeled centrifuge tubes. Flow-through proteins and the eluted lectin-enriched glycoproteins were then reduced, alkylated, and digested with PNGase F overnight followed by an overnight digestion with trypsin.

4.2.5 Gel Electrophoresis

A 1D gel electrophoresis of concentrated secretome samples (24h, 48h, and 72 h) was run as previously described [82]. 20 µg aliquots of secretomes were pipetted and made up to 20 µL with HPLC grade water. The samples were denatured and reduced by mixing with 20 µL of 50 mM DTT in 2X Laemmli buffer and incubating at 100^{o}C for 5 minutes.

The resulting 40 µL of each sample was then loaded onto a 10% Criterion Tris-HCl precast gel (Bio-Rad) and run at 200V for 1 hr. The gel was washed with distilled water for 3 × 10 minutes followed by Coomassie staining on an orbital shaker for 30 minutes. The stain was then washed with distilled water for 30 minutes. The stained gels were then visualized with a Bio-Rad gel imager and stored at 8^{o}C.

4.2.6 LC-MS/MS Method

75 µL solutions of peptide samples were pipetted into LC vials for nLC-MS/MS analysis. The LC-MS/MS system consists of an Eksigent nLC 415 (ABSciex) in a vented column, trap and elute configuration. The reverse phase trap column (75 µm x 5 cm) and analytical column (75µm x 15cm) were both packed in-house with Magic AQ C18 3µm and 200Å material. The nLC system was coupled to a Q-Exactive (Thermo Scientific, San Jose, CA) mass spectrometer equipped with the Nanospray-Flex ionization source fitted with a 10 µm ID emitter PicoFrit tip (New Objective). 2 µL of peptide sample was loaded onto the trap column and desalted at a flow rate of 2.25 µL/min for 5 minutes using mobile phase A (98% H_2O/2% acetonitrile, 0.1% formic acid). Desalted peptides were then separated on the C18 self-pack column at 300 nL/min with increasing mobile phase B (2% H_2O/98% acetonitrile, 0.1% formic acid) using the following gradient: 5% B (0 – 4 minutes), 35% B (95 minutes), 75 % B (105 – 110 minutes), 5% B (115 minutes) and held for 5 minutes until the run finishes at 120 minutes. The electrospray emitter tip was charged with a voltage of 1.80 kV in positive ion mode and the Q-Exactive inlet temperature and S-lens setting were maintained at 250°C and 62 V, respectively. Full scan (400-1600 m/z) resolution was set at 70,000 FWHM with an AGC target of 3×10^6. MS/MS was set to a resolution of 17,500 with an AGC target of 2×10^4 at 120 ms maximum inject time and

selection of the top 12 ions at a 30 second dynamic exclusion. HCD voltage was maintained at 30 NCE throughout.

4.2.7 LC-MS/MS Protein Identification, Quantification, and Data Analysis

Proteomic raw files were processed in the MaxQuant computational platform (ver. 1.5.8.3) with the Andromeda search algorithm using the Uniprot Human proteome database (common contaminants were added to the search). Up to two missed cleavages, mass accuracies: MS = 5 ppm, MS/MS = 0.02 Da; fixed modifications: carbamidomethyl (C), variable modifications: acetyl (N-terminus) and methionine oxidation (M), and a false discovery rate (FDR) of 1% were included in the search parameters. Quantification of proteins was done in MaxQuant using the LFQ algorithm requiring at least 2 shared peptides[67]. Statistical analyses were carried out in Perseus (ver. 1.5.8.8) and JMP Pro 13 Statistical Software. One-way analysis of variance (ANOVA) tests were carried out followed by t-test pairwise comparisons to determine significant effects using the Benjamini-Hochberg correction. Imputation of missing values was done in Perseus by replacing with random numbers drawn from the lower (left) boundary of a normal distribution. Gene Ontology (GO) annotation enrichment analysis were performed using the David Bioinformatics Resource (ver 6.8) [362, 363]. For the GO enrichment analysis in David, a term is said to be enriched if the proportion of proteins associated with the GO term in the data set is significantly higher (using a modified Fisher Exact test) than the proportion of proteins associated with the term in the human proteome. In the Perseus statistical analysis, protein identification lists were filtered by removing contaminants and reverse database sequence hits.

4.3 Results & Discussions

The use of the SIL HepG2 secretome as a global internal standard for plasma versus cell lysates offers some critical advantages for LC-MS/MS comparative studies. The majority

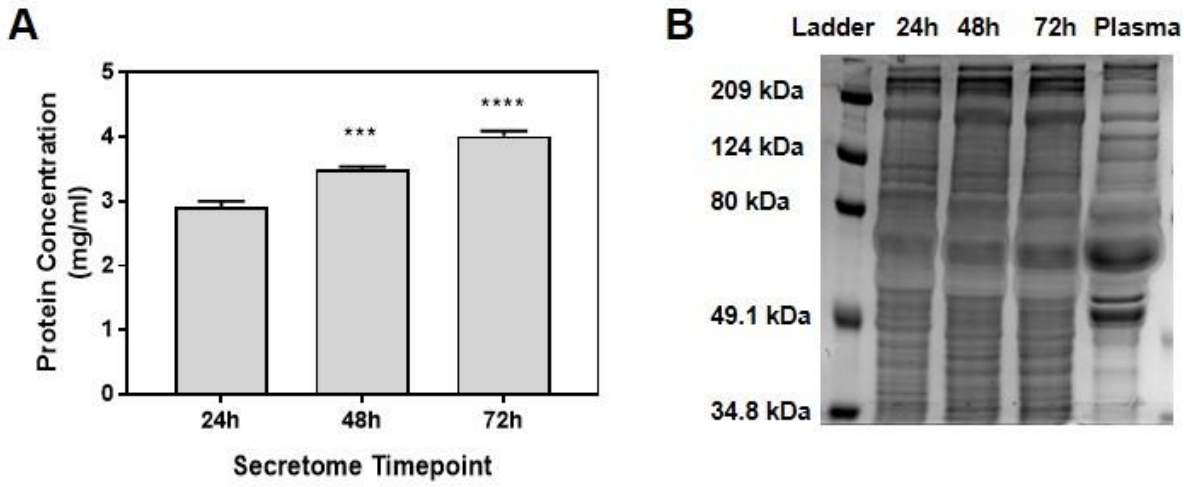


Figure 19 - Protein Content of the HepG2 Secretome over 72 hours

A) Total protein concentration determined by UV absorbance at 280nm on a plate reader. Total protein levels increased from 2.88 mg/mL (± 0.11) at 24 hours to 3.46 mg/mL (± 0.08) at 48 hours and 4.02 mg/mL (± 0.08) at 72 hours. Statistical significance was determined using t-tests. B) Coomassie stained 1D-Gel image of HepG2 secretomes at 24, 48, 72 hours relative to human plasma.

of plasma proteins are produced by liver hepatocytes. The HepG2 cell line is derived from liver adenocarcinoma and has been used as a model for liver hepatocyte function including plasma protein synthesis. Primary hepatocytes were not used due to concerns with the ability to continue cell propagation after stable isotope labeling that occurs in about nine doublings (over 2 – 3 passages). To use the HepG2 cell line to produce SIL plasma proteins, serum-free conditions are necessary to avoid the issue of high protein background from fetal bovine serum [361]. However, there are concerns regarding cell survival as well as the contribution, from lysed apoptotic cells, of proteins that would otherwise not have been secreted into the media to the total secretome [361]. In other

words, sustained incubation of HepG2 cells under serum free conditions may stress the cells and potentially reduce the overall number of viable cells. Thus, our initial efforts were focused on exploring the protein production of HepG2 cells under serum-free conditions

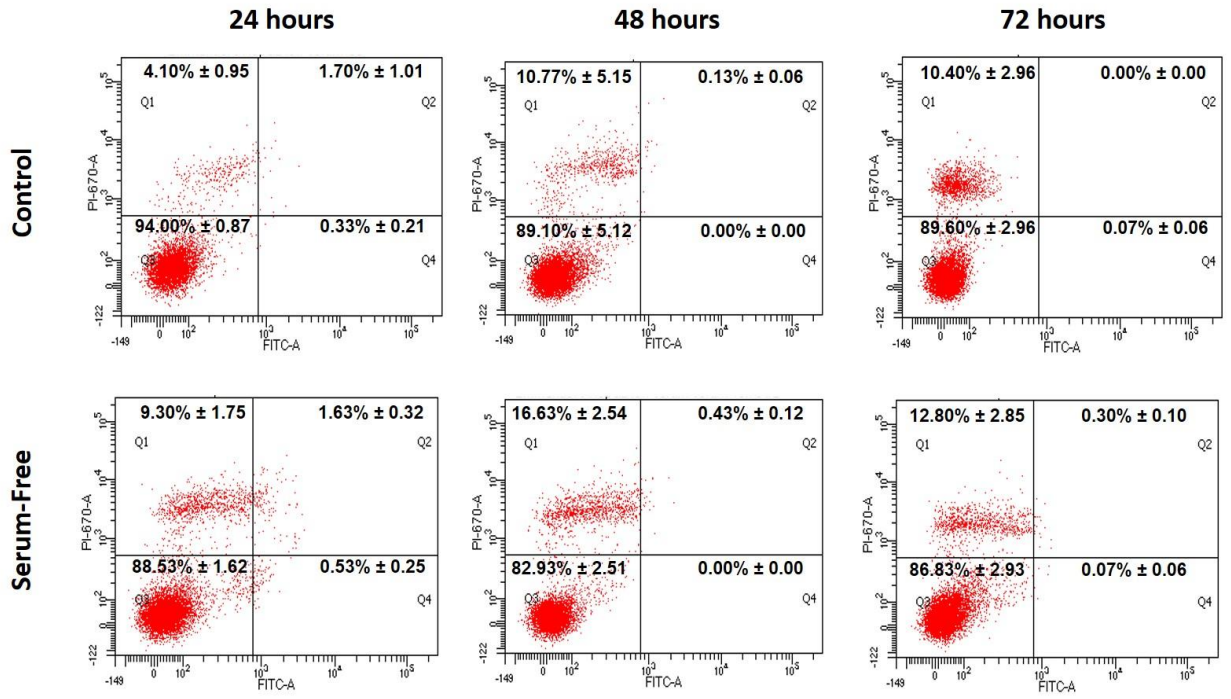


Figure 20 - Annexin V/PI Apoptosis Assay

Flow cytometry of Annexin V/PI stained cells shows no significant induction of apoptosis in HepG2 cells following serum starvation (Serum-free) compared to serum fed (Control) cells at 24, 48, and 72 hours following initiation of serum-free conditions

over an extended period while concurrently monitoring cell viability.

To address these issues, we measured total protein production of HepG2 cells at 24h, 48h, and 72h under serum-free conditions while concurrently measuring cell viability under the same conditions. HepG2 cells were cultured in triplicate in 6-well plates in high glucose SILAC DMEM supplemented with 10% dialyzed FBS, $^{13}C_6$ L-arginine-HCl, $^{13}C_6$ L-lysine-2HCl, and 1% 100 U/ml Penicillin G sodium/1% 100 µg/ml Streptomycin sulfate

at 37°C and 5% CO_2. At ~ 85% confluence (after 48 hrs), the media was replaced with serum-free SILAC DMEM-high glucose media following which conditioned media (secretomes) were collected and total protein concentration measured at 280 nm on the

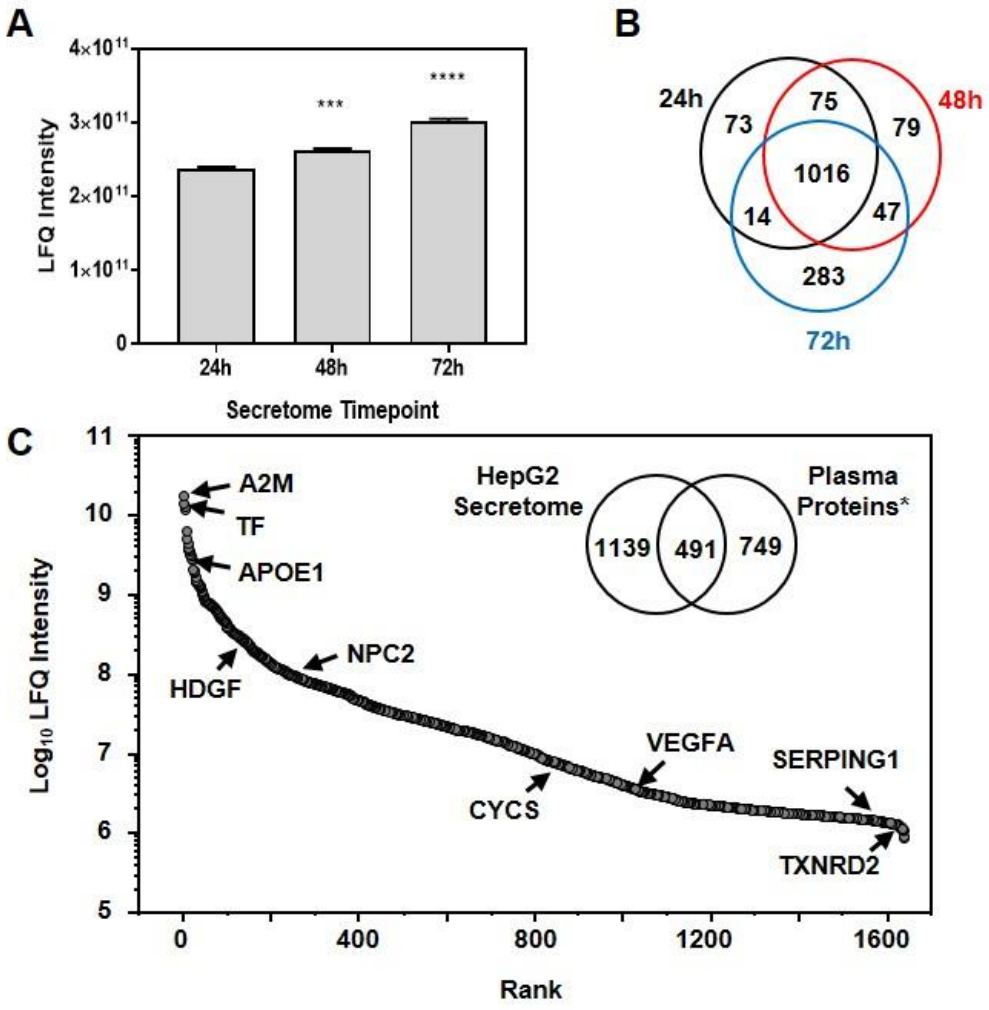


Figure 21 - Protein Distribution in the HepG2 Secretome

Quantitative proteomics results showing the A) total LFQ intensities at 24h, 48h, and 72h and B) distribution of individual quantified proteins. C) Rank ordered LFQ abundance levels of plasma proteins (*Nanjappa *et. al.* 2014) and cancer biomarkers (Polanski and Anderson 2007) quantified in this study.

BioTek Synergy Hybrid Take3 Plate Reader. The results reported in **Figure 19A** showed a sustained and statistically significant increase with time. Total protein levels increased from 2.88 mg/mL (± 0.11) at 24 hours to 3.46 mg/mL (± 0.08) at 48 hours and 4.02 mg/mL

(± 0.08) at 72 hours. The distribution of secreted proteins was also assessed by 1D gel electrophoresis (**Figure 19B**). Furthermore, all time points showed similar bands to human plasma including the prominent albumin band at ~ 66 kDa. Concurrent with the secretome analysis in **Figure 19**, we determined the effect of serum starvation up to 72 hours on HepG2 cells using the BD Biosciences apoptosis kit. The extent of apoptosis was evaluated at 24, 48, and 72 hours according to the manufacturer's directions. The results revealed very low levels of apoptosis in both control cells and cells exposed to serum starvation **(Figure 20)**. There were very few cells in early or late apoptosis with most samples in both control and serum-starved cells showing mostly no cells in late apoptosis.

4.3.1 Quantitative Analysis of the SILAC HepG2 Secretome

Pooled concentrated secretome samples were then reduced, alkylated, and digested for nLC-MS/MS analyses. This increasing protein concentration correlated directly with total LFQ intensity levels which increased from 2.38E11 (± 3.20E9) at 24 hours to 2.63E11 (± 2.78E9) at 48 hours and then to 3.03E11 (± 2.55E9) at 72 hours in the PNGase F-treated secretome samples (**Figure 21A**). As shown in Supplementary Figure 1, increased protein secretion over time was not a result of increased protein from lysed apoptotic cells with increasing incubation time in serum-free media.

The nLC-MS/MS analysis of the HepG2 secretome revealed the identification of 1635 unique proteins. Using LFQ[67] quantification to filter the proteins, 1229 were identified to be quantified in 2 or more replicates of at least one time point. Further quantitative analysis showed that 1187, 1225, and 1374 proteins were quantified at least once in the

24h, 48h, and 72h secretomes respectively. A summary of the distribution, including overlap, of quantified proteins across the three time points is illustrated in **Figure 21B**. 1016 proteins were quantified at least once in all time points while the 72h secretome had the most unique proteins quantified (283). Bioinformatics functional enrichment analysis

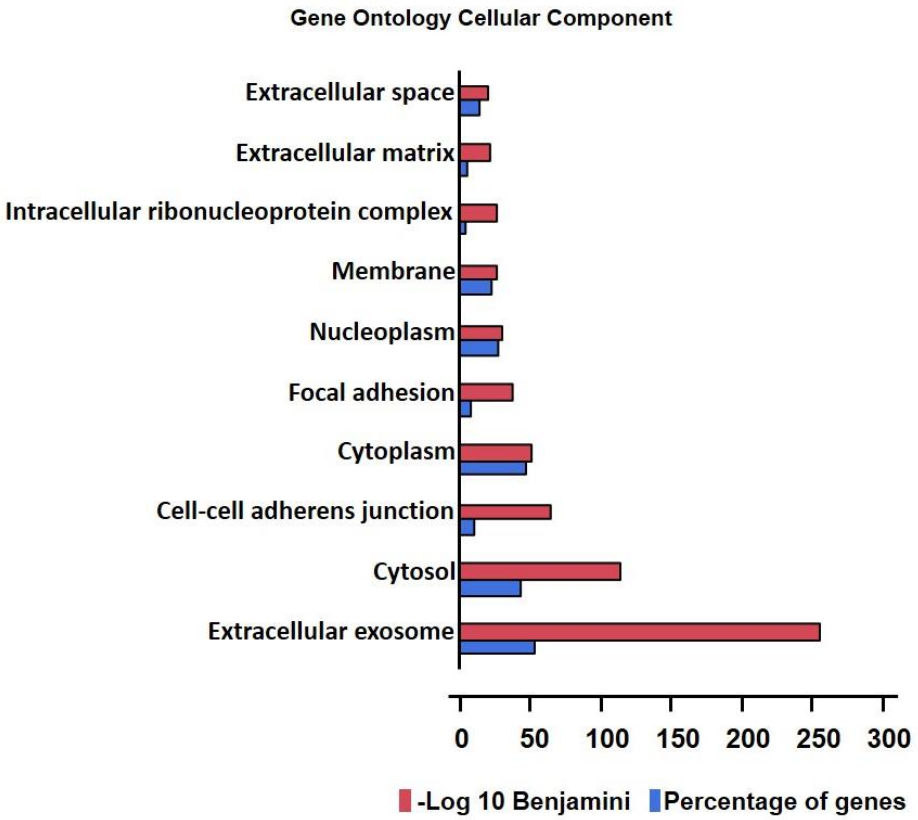


Figure 22 - Gene Ontology Cellular Component (GOCC) Analysis

Bioinformatics functional analysis of 1229 proteins quantified in ≥ 2 replicates of one or more groups (secretome only analysis) revealed the Gene Ontology Cellular Component (GOCC) most enriched in the HepG2 secretome include extracellular exosome, cytosol, cell-cell adherens junction, focal adhesion, membrane, and intracellular ribonucleoprotein complex.

was performed on the 1229 proteins using David 6.8. The results revealed the enrichment of cellular components including extracellular exosomes (~3.57 fold, p-value = 3.13×10^{-257}), cytosol (~2.48 fold, p-value = 7.02×10^{-115}), cell-cell adherens junction (~5.87 fold, p-value = 1.69×10^{-65}), focal adhesion (~4.16 fold, p-value = 1.92×10^{-38}), membrane

(~1.90 fold, p-value = 1.27 × 10^{-27}), and intracellular ribonucleoprotein complex (~5.98 fold, p-value = 1.75 × 10^{-27}) as shown in **Figure 22**.

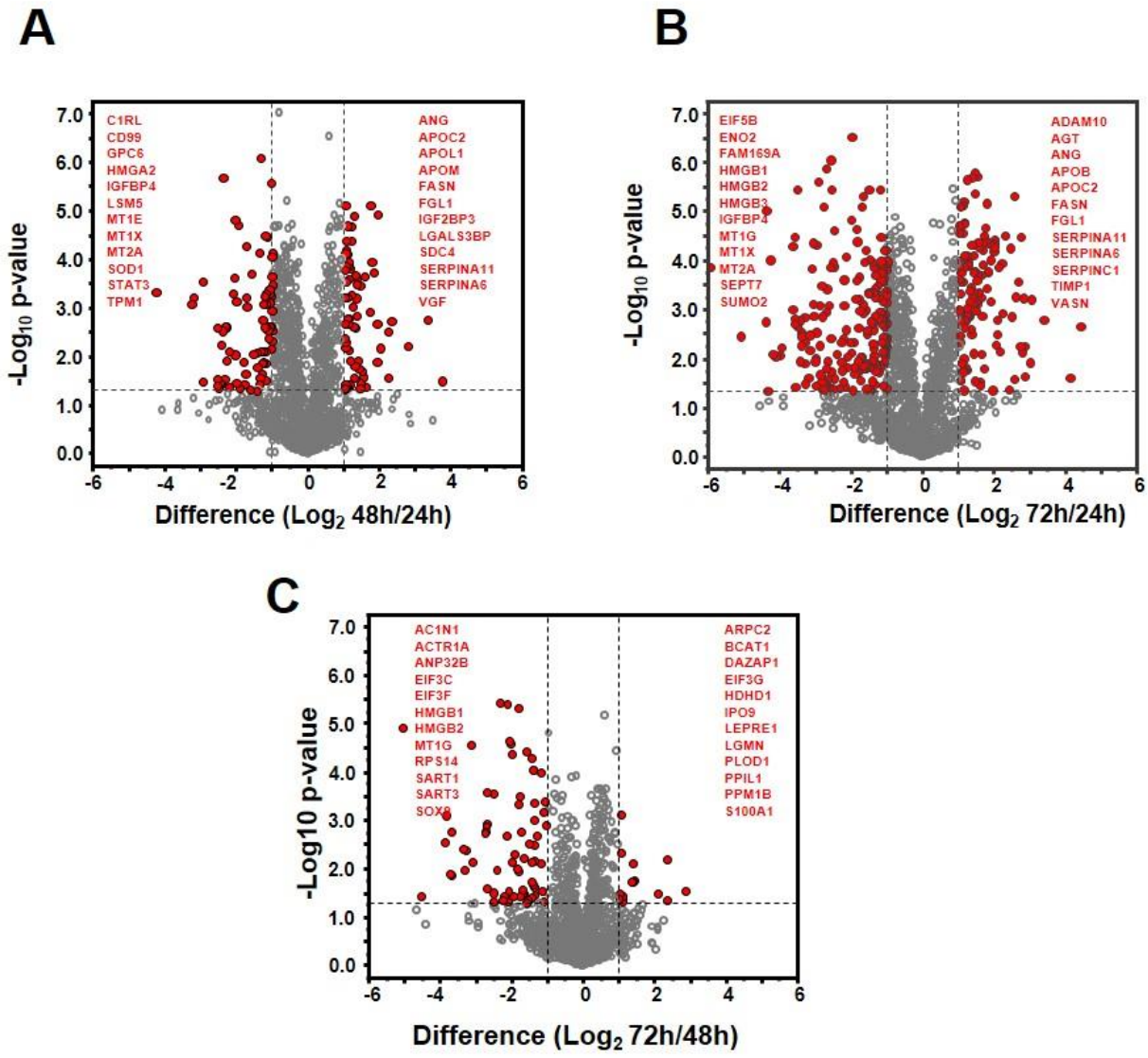


Figure 23 - Differential protein secretion over 72 hours in HepG2 cells.

Volcano plots showing significantly different proteins (shown in red circles) between the 48 hour secretomes compared to the 24 hour secretome (A), the 72 hour secretome compared to the 24 hour secretome (B), and the 72 hour secretome compared to the 48 hour secretome (C).

The HepG2 hepatocarcinoma cell line has been demonstrated to secrete many proteins, including the major plasma proteins at levels comparable to plasma, with function in different physiological processes[82, 364]. Many of these plasma proteins (e.g. ORM1, TF,

FGA, A2M, and APOA2) serve as biomarkers of disease e.g. cancer [365]. Among the major plasma proteins of interest are apolipoproteins, protease inhibitors, complement factors, and coagulation factors. A comparison of the 1635 identified proteins in the HepG2 secretome with the plasma protein database curated from the Nanjapa et al manuscript [365] and downloaded from http://plasmaproteomedatabase.org revealed an overlap of 491 non-redundant proteins. Included in the 491 proteins are many apolipoproteins including APOA1, APOA2, APOA4, APOA5, APOB, APOC1, APOC3, APOE, APOH, APOL, and APOM, as well as many serine protease inhibitors (SERPINs)

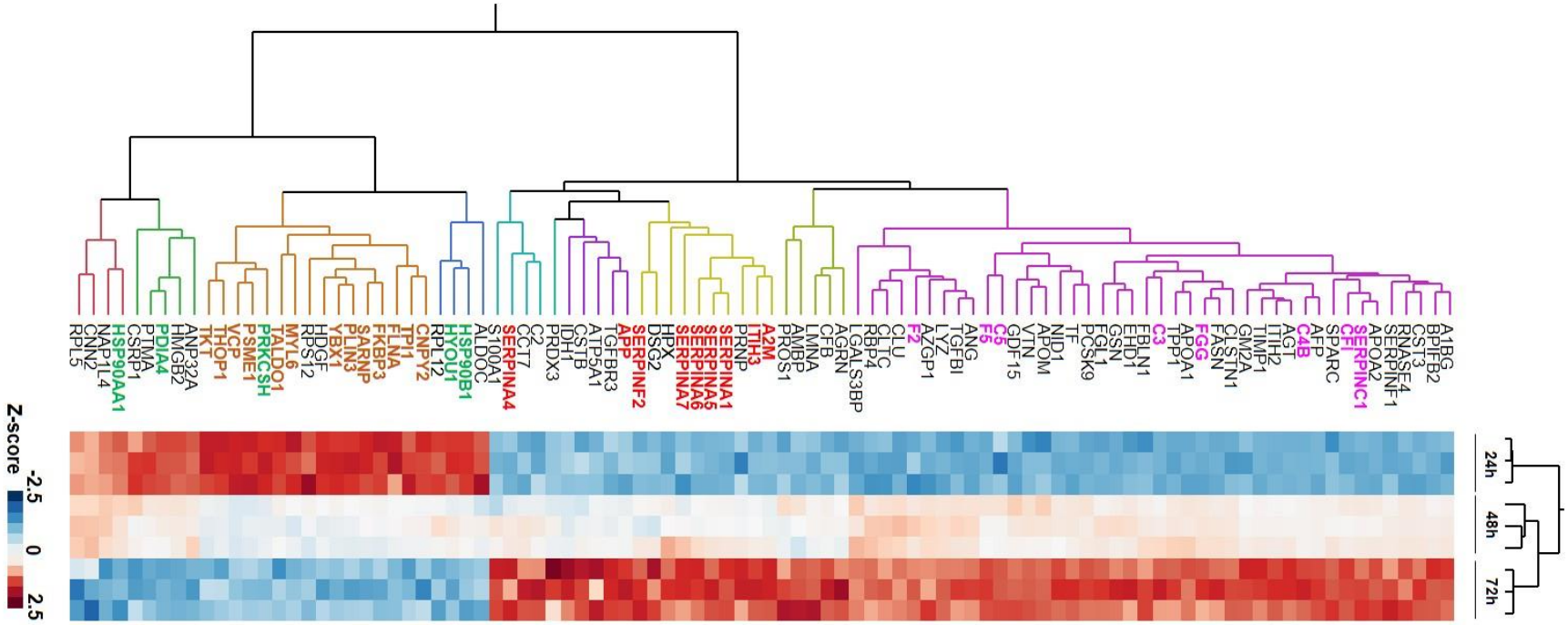


Figure 24 – Hierarchical Clustering of Secreted Proteins in HepG2 Cells

Heatmap of z-score normalized HepG2 protein secretion over 72 hours showing increasing and decreasing levels of different plasma proteins over time. Proteins corresponding to various GO terms clustered together and are colored as follows: purple – associated with complement and coagulation cascade, red – associated with serine-type endopeptidase inhibitor activity, brown – associated with protein binding, and green – associated with protein processing in the ER.

such as alpha-1-antitrypsin (SERPINA1), antithrombin III (SERPINC1), and heparin cofactor 2 (SERPIND1). Apolipoproteins are the major players in the plasma lipid transport system and their quantitative measurements serve important roles in cardiovascular disease diagnosis and prognosis. A plot of the rank ordered total LFQ intensity (computed by an algebraic sum of the LFQ intensity across all technical replicates of each sample) of the 491 plasma proteins showed quantification over six orders of magnitude **(Figure 21C)**.

Statistical analysis using t-tests were performed on the HepG2 secretome dataset to determine changes in the HepG2 protein secretion dynamics. Significant differences were found mainly to be between the 24 h secretome and 48 h or 72 h secretomes. Few differences in protein secretion were observed between the 48 h and 72 h secretomes. 112 proteins were found to be significantly different between 24 h and 48 h **(Figure 23A)**, while 279 proteins differed significantly between the secretomes at 24 h and 72 h **(Figure 23B)**, using an α value of 0.0167 for 3 pairwise comparisons and ≥ 2-fold difference. There were 44 significantly different proteins between the secretomes at 48 h and 72 h **(Figure 23C)**.

Additionally, we were interested in proteins with consistently increasing or decreasing secretory levels in the secretome from 24 – 72 hours. A z-score hierarchical cluster of 96 plasma proteins found to be significantly decreasing or increasing over the 72-hour period is shown in **Figure 24**. Of these, 67 showed increasing levels, while 29 showed decreasing levels in the HepG2 secretome.

4.3.2 N-Glycosylation of Plasma Proteins

To carry out N-glycosylation analysis of HepG2 secreted proteins, the concentrated secretome was treated with PNGase F prior to trypsin digestion. Following PNGase F treatment to cleave N-glycans from asparagine residues and generation of tryptic digests, the deglycosylated peptides were analyzed by LC-MS/MS and raw data processed.

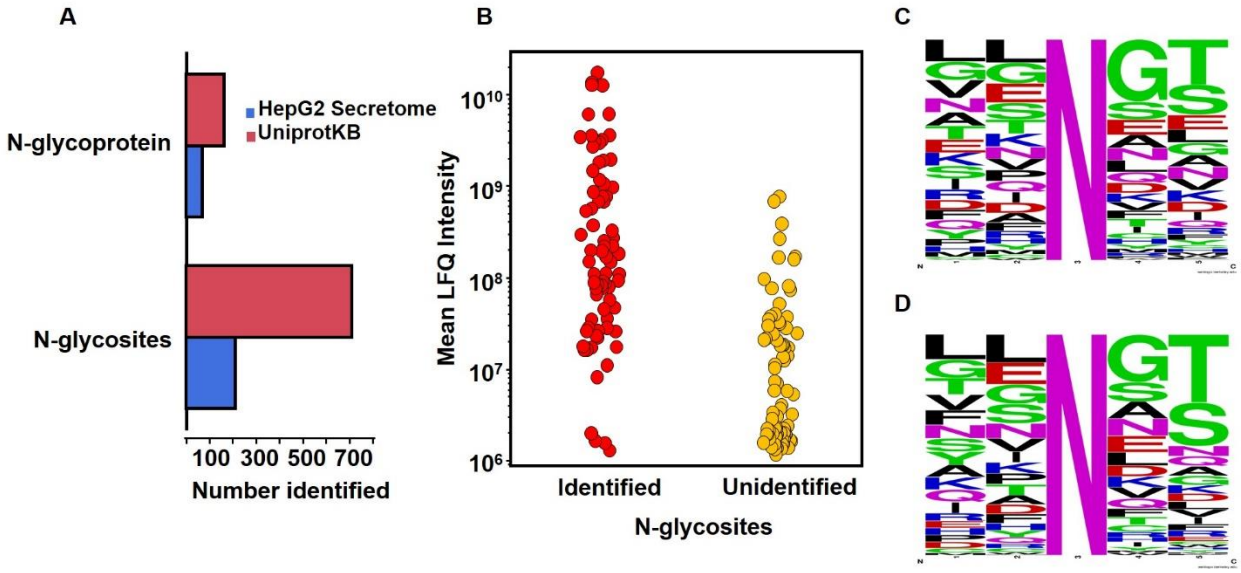


Figure 25 - Plasma N-glycoproteins in the HepG2 Secretome

A) Bar graph showing the distribution of Uniprot-verified plasma N-glycoproteins and the corresponding number of N-glycosites found in the HepG2 Secretome. **B)** Scatter plot of the 163 plasma glycoproteins with identified and unidentified glycosites in the PNGase F-treated HepG2 secretome. Glycoproteins with identified sites of glycosylation had a greater mean LFQ value than those with unidentified glycosites. **C)** N-glycosite analysis of 293 proteins corresponding to 621 glycosites identified in the PNGase F-treated HepG2 secretome showed 34% consensus with the N-X-T/S motif **D)** N-glycan motif analysis of the 76 plasma glycoproteins with 283 identified glycosites showed 143 glycosites (~50%) consistent with the N-X-T/S motif.

Manual data analysis via a cross reference of the 491 identified plasma proteins against the Uniprot database indicated 163 of them were N-glycosylated and represented 707 unique N-glycosites. Among the 163 Uniprot plasma N-glycoproteins, 76 were identified with deglycosylated peptides in the PNGase F-treated HepG2 secretome **(Figure 25A)**. Further analysis revealed that plasma N-glycoproteins in the HepG2 dataset with

identified N-glycosites had a higher mean LFQ intensity across all time points compared to N-glycoproteins with unidentified glycosites **(Figure 25B)**.

An illustration of the most highly represented sequence motifs in the glycoproteins identified in the HepG2 secretome data is displayed in **Figures 25C-D**. PNGase F treatment revealed the presence of 621 sites of N-glycosylation in the HepG2 secretome, ~ 34% of which were consistent with the conserved N-X-T/S motif as shown in **Figure 25C**. However, for the 76 plasma N-glycoproteins, a greater proportion of the 283

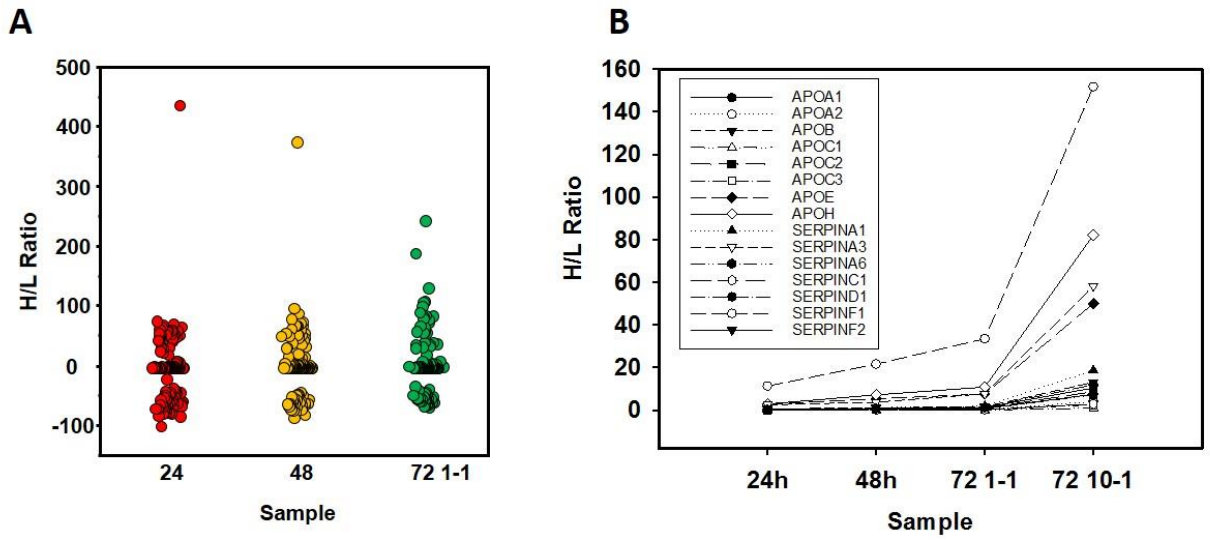


Figure 26 - Plasma Protein Secretion Changes over 72 hours in HepG2 Cells

A) Median heavy:light (H/L) ratios for 143 proteins quantified in 2 replicates of at least one sample as a function of time. B) Extracted median H/L ratios of apolipoproteins and SERPINs identified and quantified in the HepG2 secretome

identified N-glycosites (143 representing ~ 50%) were identified as valid N-glycosites **(Figure 25D)**.

4.3.3 SIL HepG2 secretome protein levels versus plasma

This present study involves the characterization of the stable isotope labeled (SIL) HepG2 secretome with downstream applications including plasma analysis. Consequently, we

spiked the SILAC heavy labeled HepG2 secretome into an ethylenediaminetetraaceticacid (EDTA) plasma sample. Here, three SILAC secretome:plasma samples were prepared by spiking an equal volume (35.60 µL) of each secretome sample (24 h – 2.976 µg/mL, 48 h – 3.596 µg/mL, and 72 h – 4.213 µg/mL) into 150 µg of plasma (6.607 µg/mL) in three different 1.5 mL centrifuge tubes. The SILAC mixtures were subsequently processed and analyzed via nLC-MS/MS before and after PNGase F treatment, with and without lectin enrichment of N-glycoproteins.

For the samples analyzed without PNGase treatment, an additional 10:1 heavy to light mixture (300 µg:30 µg) was prepared by combining 71.21 µL of the 72 h HepG2 secretome with 4.54 µL of plasma, resulting in twelve sample replicates representing four sample groups (24 h, 48 h, 72 h 1:1, and 72 h 10:1). This was done to assess linearity of the nLC-MS/MS method. In the plasma spike-in study 1405 proteins in total were identified and of these 143 proteins were quantified with ≥ 2 valid H/L ratios in at least one group **(Figure 26A)**. Consistent with the total protein concentration at 280 nm, LFQ intensity and 1D gel results of the secretome only analysis, heavy:light (H/L) ratios of identified proteins in the spike-in study revealed increasing protein ratios from 24 to 72 hours. **Figure 26B** shows that for apolipoproteins and protease inhibitors (quantified with valid H/L ratios in all replicates of all samples), secretion was generally trending upwards. H/L ratios were seen to increase by approximately ten-fold from the 72h 1:1 sample to 72h 10:1 sample. However, this was not observed for all proteins due to the presence of proteins for which valid H/L ratios were obtained for the 72h 1:1 sample but not for the 72h 10:1 sample.

In a proof of concept study [82] showing the potential for using the SILAC-labeled HepG2 secretome as an internal standard for comparative LC-MS/MS analysis of plasma, the potential impact of glycosylation on the quantitative value of this approach was demonstrated. In that study, carried out without enzymatic deglycosylation, the heavy isotope labeled peptide QNQCFYNSSYLNVQR from alpha-1-acid glycoprotein 2 (ORM2) was identified in the HepG2 secretome sample whereas the light-labeled peptide from plasma was not identified. We predicted that this might have been a result of differential

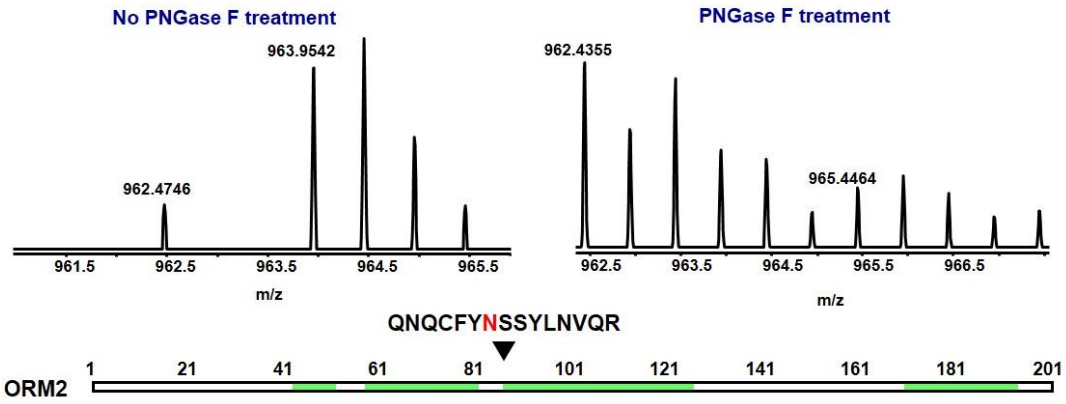


Figure 27 - Peptides from alpha-1-acid-glycoprotein 2 (ORM2)

Heavy and Light isotopes of the peptide QNQCFYNSSYLNVQR ORM2. Heavy isotope was identified with or without PNGase F treatment of the SILAC combined sample whereas the light isotope from plasma was only identified with PNGase F treatment.

glycosylation at the 93 position of the protein between the HepG2 secretome and plasma. Thus, we incorporated PNGase F cleavage of N-glycans into the workflow for the present study. Manual data analysis and interpretation of the secretome:plasma mixture without PNGase F treatment revealed the same observation wherein the QNQCFYNSSYLNVQR was identified in the HepG2 secretome but not plasma **(Figure 27)**. However, both the heavy and light isotopes of the peptide were detected and quantified in the PNGase F-treated mixture **(Figure 27)**. That the peptide was not identified in plasma without

PNGase F treatment is suggestive of differential glycosylation between the HepG2 and plasma forms of the protein. Additionally, this finding is indicative of the effect of glycosylation on the quantitation of certain peptides.

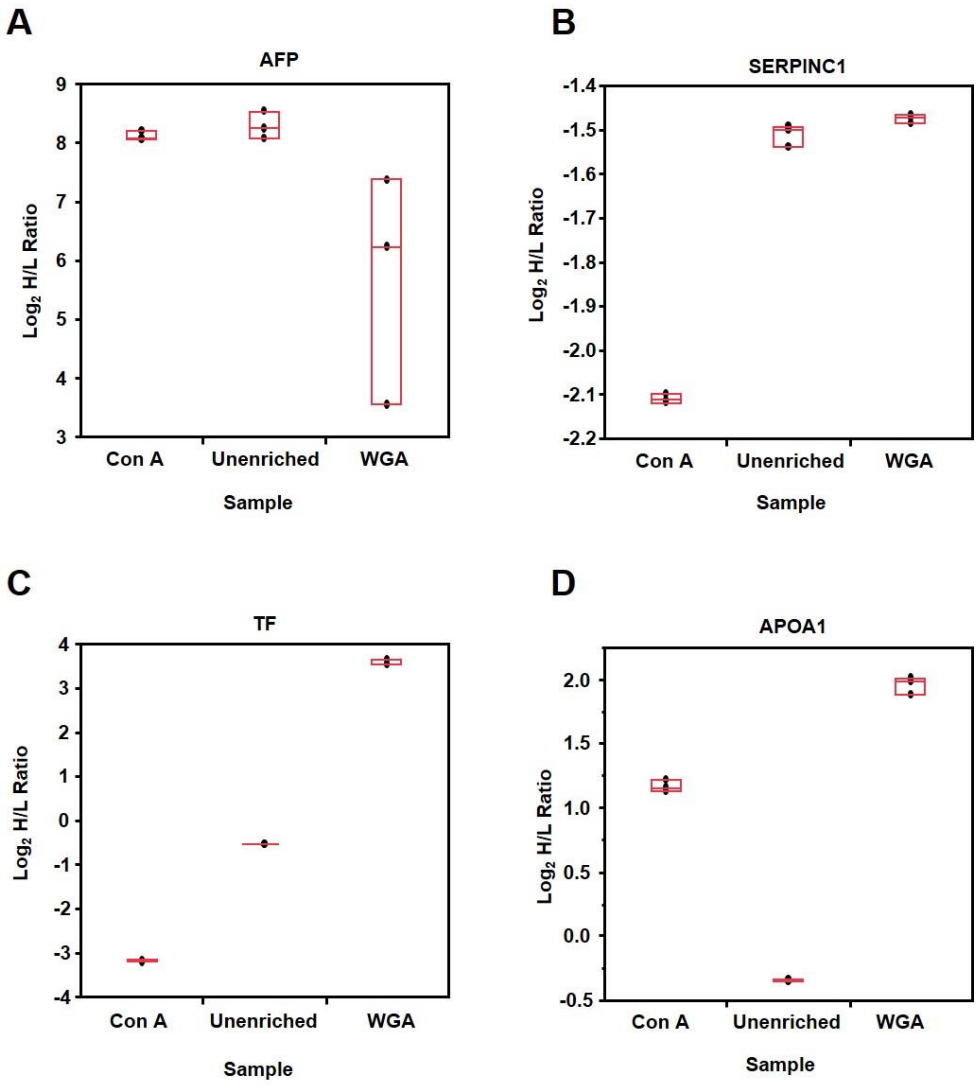


Figure 28 - The effect of lectin enrichment on N-glycoprotein levels between proteins in the HepG2 secretome and plasma

Proteins that were quantified with valid H/L ratios in ≥ 2 replicates of at least one sample group were extracted and their Log_2 H/L ratios computed; missing values were imputed. The graphs were then plotted to illustrate the changes in protein levels due to enrichment using lectins with different specificities (Con A and WGA). Graphs are shown for **A)** Alpha Fetoprotein (AFP), **B)** Anti-thrombin III (SERPINC1), **C)** Serotransferrin (TF), and **D)** Apolipoprotein A1 (APOA1).

4.3.4 Differential glycosylation of SIL HepG2 secretome versus plasma glycoproteome

Complex glycopeptides and/or glycan identification can be improved using glycoprotein enrichment strategies including lectin affinity capture[80, 81, 366] and solid phase extraction[107]. Several lectins exist with widely differing binding affinities for different glycan moieties. Wheat germ agglutinin (WGA) and Concanavalin A (Con A) are two of the most widely lectins for glycoprotein/glycopeptide enrichment. Despite the general knowledge in glycomics that there is some overlap between WGA and Con A binding preferences, the degree of overlap is generally not known. It is widely accepted however, that WGA binds preferentially to glycoproteins with N-acetylglucosamine residues and sialic acid residues (e.g. N-acetylneuraminic acid) whereas Con A binds to mannose[81, 123, 367]. Consequently, we pursued lectin enrichment of the 72 h secretome:plasma SILAC sample and explored, where present, glycan differences in major plasma proteins including apolipoproteins and SERPINS using WGA and Con A. 894 proteins were identified in the lectin enrichment study with 119 proteins quantified with valid H/L ratios in at least one of the groups. Among the 119 proteins are many apolipoproteins, complement factors, and serine protease inhibitors identified with sequence coverages as high as 70%. H/L protein ratios were transformed to Log_2 values following which missing values were computed. We then determined median Log_2 H/L ratios for each protein in each group. While protein glycosylation plays a vital role in various molecular processes in the body, the microheterogeneity of glycans found on a single site of glycosylation in proteins poses challenges [368].

It is important to note that when the same types of glycan structures are present at a given glycosite of both the heavy (HepG2 secretome) and light (plasma) isotopes of a given protein, the H/L ratios determined for the Con A or WGA lectin-enriched samples would be expected to largely be similar, if not equal. Furthermore, this ratio will not be different from the unenriched sample. However, when a given glycosites harbors different kinds of glycans in the secretome compared with the plasma, the H/L protein ratio determined for the unenriched sample will differ from the ratios determined for either Con A- or WGA-enriched samples. For example, AFP shows similar ratios for the unenriched sample and the Con A-enriched sample but a significantly different ratio for the WGA-enriched sample **(Figure 28A)** signifying that the heavy- and light-labeled isotopes of AFP have similar types of mannose-containing glycans but different sialic acid or N-acetylglucosamine residues. While AFP has two possible sites of N-glycosylation, only one known site is reported in Uniprot. Although controversial, serum AFP levels have been used as a marker for hepatocellular carcinoma (HCC) and nonseminomatous germ cell tumors (NSGCT) [369, 370]. Johnson et al. reported, in 1999, the isolation and structural elucidation of eleven glycan structures from HCC and NSGCT patients including seven N-linked glycans with different levels of sialylation, fucosylation, and galactosylation[369]. The authors stated that the glycans present on AFP may be useful for diagnosis as the glycan structures may be related to the type of tumor present in the patient. A study by Kim et al.[370] then showed that the glycosylated (or deglycosylated) form of AFP performed better as a diagnostic tool for HCC than the un-glycosylated form of the protein. Ajdukiewicz et al.[371] also reported abnormal glycosylation of AFP in some HCC making it plausible that the glycosylated heavy-labeled AFP in the secretome differs from that

present in the plasma sample due to the different origin of the samples. Moreover, this highlights the value of the SILAC HepG2 secretome as an internal standard for the analysis and detection of differential glycosylation in patient plasma.

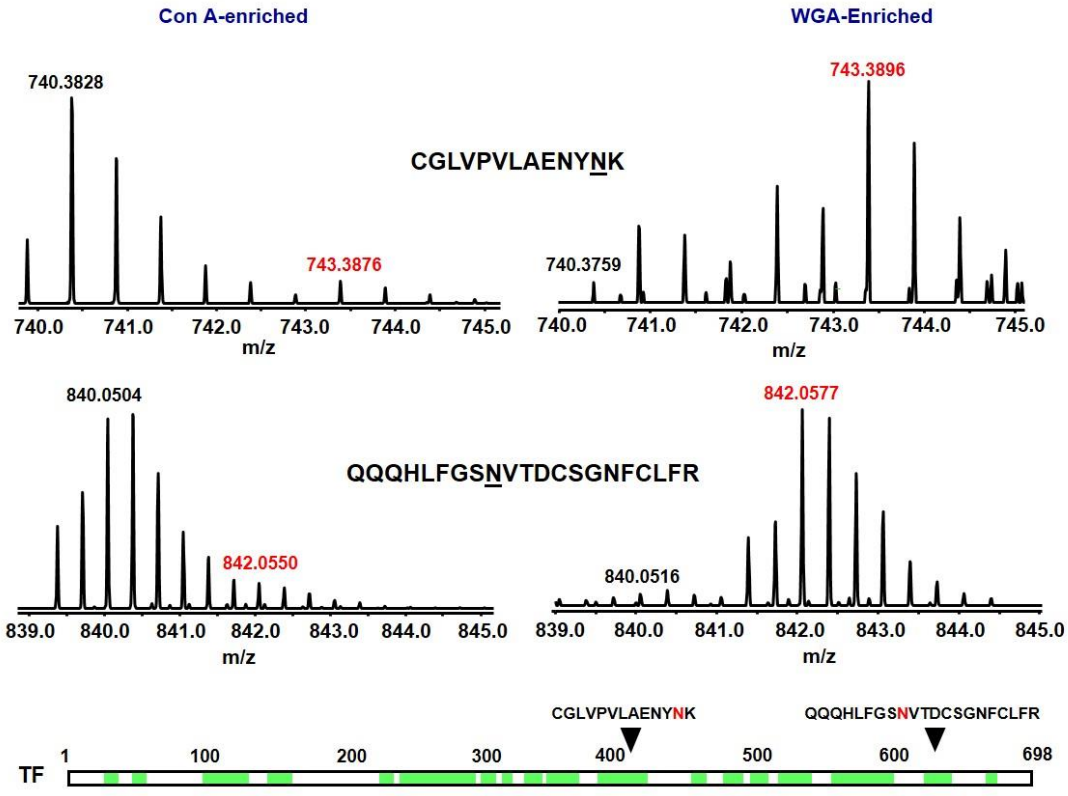


Figure 29 - Deglycosylated Peptides from serotransferrin (TF)

Heavy and Light isotopes of two peptides from serotransferrin (TF), CGLVPVLAENYNK and QQQHLFGSNVTDCSGNFCLFR showing distinct differences in the amounts present in the Con A-enriched and WGA-enriched samples. Heavy peptides are shown in red font and light peptides in black.

Again, SERPINC1 showed similar ratios for the unenriched and WGA-enriched samples but a different for the Con A-enriched sample as illustrated in **Figure 28B** suggesting similar sialic acid residues but different mannose residues between the HepG2 secreted protein and its plasma isotope. As a necessary factor for the inhibitory effect of heparin on thrombin and factor Xa in the clotting cascade, the discovery of potential differences

in SERPINC1 glycosylation between the HepG2 secretome and human plasma is very relevant. Moreover, it is known that differences between the alpha and beta forms of SERPINC1 regarding the glycan composition is responsible for the approximately ten-fold difference in heparin binding affinity[372]. We also observed differing H/L ratio for

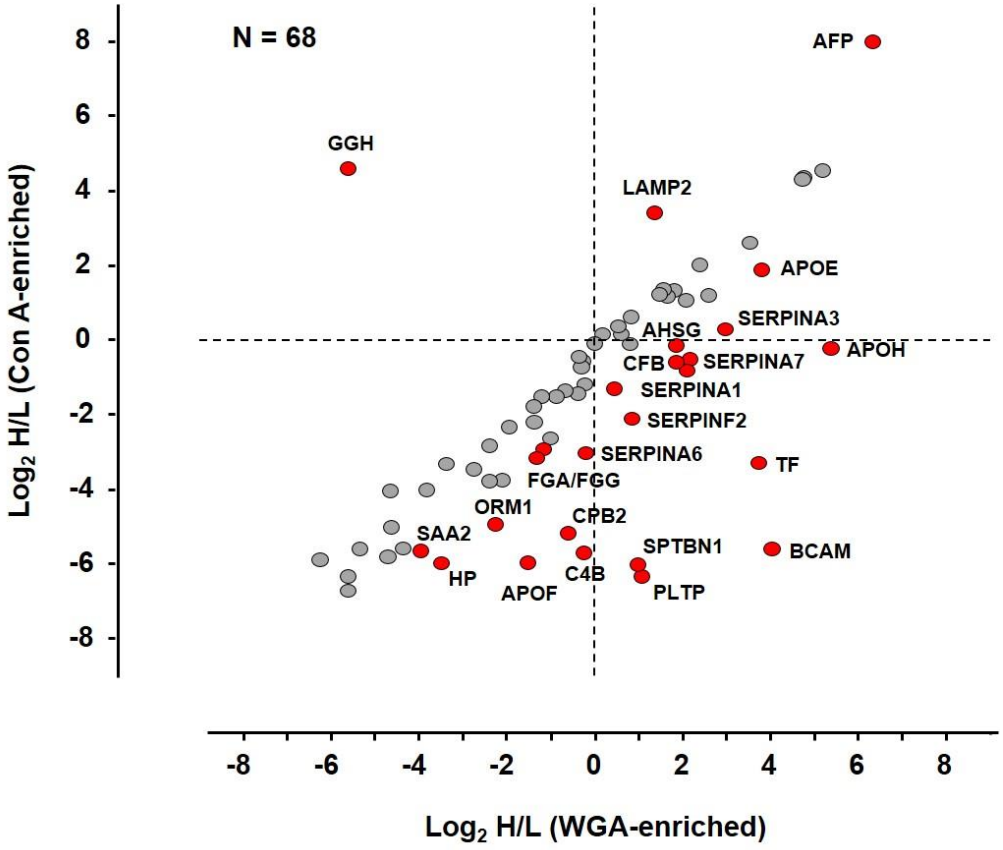


Figure 30 - Differential enrichment of 68 SIL HepG2 glycoproteins relative to human plasma

Glycoproteins enriched by more than ±1.5 are indicated in red. SIL glycoproteins were enriched primarily with WGA lectin affinity relative to ConA lectin affinity.

various proteins including serotransferrin (TF), and APOA1 between the Con A- and WGA-enriched samples compared to samples that were not enriched **(Figures 28C-D)**. N-glycan analysis of TF by Fu et al. resulted in the identification of trisialylated carbohydrate species [373]. The higher H/L ratio of TF following WGA enrichment

compared to the unenriched sample suggests an overexpression of sialylated glycan residues in the HepG2 secretome relative to plasma. It is still largely unknown, for most major proteins, how glycosylation differs between healthy and disease patients. Consequently, the interaction of under- or overexpression of certain glycans with patient disease remains unclear [374]. Using our SILAC HepG2 secretome as an internal standard approach for comparative plasma studies, these interactions can be delineated to provide some insight into these very pertinent questions. The results of the quantitative analysis showed that of the two known N-glycopeptides from TF that were identified and quantified in the unenriched PNGase F-treated SILAC sample, CGLVPVLAENYNK was quantified only in the Con A-enriched sample and QQQHLFGSNVTDCSGNFCLFR only in the WGA-enriched sample. However, further manual data interpretation revealed that both peptides were present in both lectin-enriched samples albeit the light peptide was always more abundant in the Con A-enriched sample while the heavy peptide was always more abundant in the WGA-enriched sample **(Figure 29)**. In addition, the lack of quantitative information on CGLVPVLAENYNK in the WGA-enriched sample is attributable to the low quality of the spectra compared to the spectra seen in the Con A-enriched sample.

Finally, in **Figure 30**, we evaluated the general relationship between all proteins with sufficient quantitative information in the Con A and WGA-enriched samples. Except for a few proteins, the results indicate that there is a general bias towards WGA enrichment in the HepG2 secretome compared to plasma. This observation is consistent with previous reports in the literature that tumor cells present with hypersialylation [375, 376].

4.4 Conclusion

The extension of the applications of the SILAC labeled proteome to the analysis of biospecimen has been demonstrated in different studies. We provided, in a previous proof of principle study, evidence to show that these applications can be extended to plasma, the most abundant biospecimen in clinical settings. Apolipoproteins (e.g. APOA1, and APOE) and serine protease inhibitors (e.g. SERPINA1, SERPINA3, and SERPINA5) serve as biomarkers in various diseases in the clinic. In this present study, we have identified and quantified many of these usually difficult to measure apolipoproteins and SERPINs, demonstrating the utility of the SILAC HepG2 secretome as a spike-in internal standard for the analysis of clinical specimen such as plasma. Moreover, in comparison to our previously published proof-of-principle study, we have improved upon the total number of proteins quantified in the secretome:plasma sample from 62 to 143. Furthermore, we have successfully characterized the temporal SILAC-labeled HepG2 secretome particularly as it relates to similarities and differences in glycosylation compared to human plasma. In so doing, we have further demonstrated the applicability and value of mass spectrometry-based proteomics, using the SILAC-labeled HepG2 secretome as an internal standard, in comparative analysis of clinical samples. We have showed differences in the glycosylation of proteins e.g. AFP, SERPINC1, TF, and APOA1 between the HepG2 secretome and plasma. Lastly, our results suggest an overrepresentation of sialylated glycans in the HepG2 secretome relative to plasma. Ongoing studies in our lab aim to demonstrate further applications of this approach and the potential to use the SILAC-labeled library of proteins generated from the HepG2 secretome in identifying different glycan-binding ligands in plasma.

Chapter 5: Overall Conclusions

Mass spectrometry-based proteomics has contributed enormously to research efforts that have led to many important discoveries that have resulted in improved understanding of cellular function and disease states over the past two decades [377]. With the current focus on the bridging the gap between clinical practice and basic science research, research efforts are directed at findings that can be easily transferred to the clinic to guide both clinical practice – early disease detection and diagnosis – as well as patient management [378, 379]. At the forefront of the clinical translation research paradigm is the search for novel, faster, and simpler ways of carrying out point-of-care tests. For this reason, MS-based proteomics leads the search for biomarkers of disease aimed towards helping in early detection, diagnosis, and prognosis of disease and treatment. In addition, current MS-based efforts are aimed at the extension of traditional sample analysis to more complex and detailed analysis of biomolecular species including complex carbohydrates often found bound to proteins and sometimes lipids.

Plasma represents the primary biospecimen for disease diagnosis and prognosis in the clinic and is therefore the sample of choice in the search for clinical biomarkers [93, 130, 380]. Initial research efforts however, benefit from looking into secreted proteins (secretome)

in cell culture. This approach is due, in part, to the complexity of plasma with protein levels spanning more than 10 orders of magnitude, as well as the analytical challenges involved in its LC-MS/MS analysis [82, 130]. As part of the normal homeostatic mechanism of the body and/or in response to different types of injury, cells secrete various proteins and other macromolecules into the extracellular environment. These secreted proteins may be found in the interstitial fluid, blood, cerebrospinal fluid (CSF), and bronchoalveolar lavage (BAL) fluid [130, 381]. Cell secretions contain many proteins with described function in cell-cell signaling and communication, immunity, and coagulation [93].

The cell culture secretome represents a surrogate for blood useful for early discovery and development research [93, 96] due to a combination of decreased complexity, more control of the experimental sample, and the relatively lower heterogeneity of samples compared to plasma [382, 383]. The secretome represents a great resource for these studies due also to the depth of information contained therein.

5.1 H1299 NSCLC Study

As has been demonstrated using MS-based proteomics, the secretome holds great experimental value as a model for biomarker discovery and development. In Chapter 3, the suitability of the secretome of the isogenic H1299 NSCLC cell line in biomarker discovery involving radiation-induced autophagy was demonstrated by the discovery of candidate biomarkers with potential use for diagnostic and prognostic purposes. Having been established, in previous studies in the Gewirtz lab, that cytoprotective and non-cytoprotective autophagy are induced in response to ionizing radiation treatment in p53-wt and p53-null H1299 cells respectively, this study aimed to explore the differences in

protein expression and secretion between p53-wt and p53-null cells before and after ionizing radiation treatment. The study was to delineate the value of different proteins as candidates for distinguishing between patients for whom autophagy modulation in addition to radiotherapy would or would not be beneficial. Initial data analysis revealed that the H1299 secretome was enriched for proteins associated with exosomes (~5.3 fold, p-value = 4.52×10^{-66}) and chaperone activity (~15.1 fold, p-value = 5.12×10^{-13}). An analysis of variance was conducted to determine protein differences across samples and yielded 25 proteins. Hierarchical clustering subsequently revealed that differential protein secretion was primarily dependent on p53 status. We then compared the protein levels in each sample to the mean protein level across all samples as a means of expressing protein secretion per sample/condition as a function of a population value. Levels of CHGB, SCG2, GPI, TXNRD1, FAM3C, CANX, and EIF5A were determined to be differ in the secretome based on radiation treatment and/or p53 status.

Overall, we demonstrated differences in protein secretion and discovered the above-mentioned candidate biomarkers that may serve a diagnostic purpose in early disease detection or prognostic role following radiotherapy.

5.1.1 Future Studies

While we have successfully developed MS-based proteomics methods for global protein analysis in the H1299 for biomarker discovery, it is necessary to recognize that more work may be done to build on the results of these studies.

First, the study may be replicated in different NSCLC cell lines to determine if the protein secretion response observed in our study is generalizable or limited to only H1299

NSCLC cells. Following the identification of candidate biomarkers, it is essential to screen and validate the panel of proteins. Consequently, the H1299 radiation-induced autophagy biomarker study will benefit from further studies involving immunoassay (Western blot or ELISA) analysis of the proteins, both in the secretome and in cell lysates, to verify the increased or decreased protein levels observed in our MS-based proteomics study. Furthermore, it would be appropriate to explore the effect of pharmacological and genetic inhibition of autophagy on the levels of these proteins before and after radiation treatment. Indeed, if the secretion of the candidate proteins is altered significantly in autophagy-competent versus autophagy incompetent cells, that would offer more support to their potential roles as biomarkers of radiation response in these H1299 cells. Finally, studies may be conducted to validate the presence and differential levels of these candidate proteins in the plasma of NSCLC patients before and after radiotherapy.

5.2 HepG2 Study

In Chapter 4, secreted proteins from HepG2 hepatocellular carcinoma cells were characterized as part of a larger effort in the generation of a library of intact SIL internal standards for comparative proteomics. HepG2 cells were cultured in serum-free conditions for 24h, 48h, and 72h and the conditioned media collected and analyzed via nLC-MS/MS. Protein secretion was generally observed to increase from 2.88 mg/mL (± 0.11) at 24 hours to 3.46 mg/mL (± 0.08) at 48 hours and 4.02 mg/mL (± 0.08) at 72 hours. Expectedly, the increasing protein secretion over time translated into increasing protein intensities (LFQ) from 2.38E11 (± 3.20E9) at 24 hours to 2.63E11 (± 2.78E9) at 48 hours and then to 3.03E11 (± 2.55E9) at 72 hours. Among the Gene Ontology (GO) terms found

to be highly enriched in the HepG2 secretome are extracellular exosomes (~3.57 fold, p-value = 3.13×10^{-257}), and membrane (~1.90 fold, p-value = 1.27×10^{-27}).

A total of 1635 proteins were identified in the HepG2 secretome out of which 491 were determined to be plasma proteins through comparative analysis with the plasma proteins database curated by Nanjappa et al [365]. Despite known difficulties in LC-MS/MS analysis of apolipoproteins, the study showed the identification and quantification of many apolipoproteins including APOA1, APOA2, APOB, APOC1, and APOM. Serine protease inhibitors including SERPINA1, SERPINC1, and SERPIND1 were also among the major plasma proteins identified and quantified with high sequence coverage. 67 proteins were determined to increase consistently with time whereas 29 proteins decreased consistently over 72 hours. A Uniprot cross reference of the 491 plasma proteins revealed that 163 proteins were N-glycosylated at 707 unique N-X-S/T sites. Of the Uniprot N-glycoproteins, 76 were identified with 143 N-glycosites in the PNGase F-treated HepG2 secretome.

A plasma spike in study was conducted, with and without lectin (Con A and WGA) enrichment of glycoproteins, in comparison of protein glycosylation between the SIL HepG2 secreted proteins and plasma. Differential glycosylation of proteins including AFP, SERPINC1, TF, and APOA1 was observed between plasma and the HepG2 secretome using the difference in levels of the Con A-enriched samples and WGA-enriched samples.

Taken together, this study validated the HepG2 cell line as a good source for the generation of a library of intact SILAC labeled plasma proteins. Moreover, it demonstrated the extended applicability of the SILAC labeling strategy to plasma and other bodily fluids.

5.2.1 Future Studies

In bottom-up MS-based glycomics, three essential pieces of information are necessary for a complete understanding and elucidation of glycoprotein structure. These are the sites of glycosylation, the intact glycopeptide sequences, and the types and numbers of different glycan species occupying the determined glycosites. Currently however, one of the major bottlenecks in glycomics is the successful elucidation of the glycans occupying specific glycosites on a protein on a global scale. Issues of the number of possible glycan structures combined with the paucity of reliable software necessary to automate the structural elucidation of glycan species contribute to this bottleneck.

Future studies may focus on the identification of glycopeptide sequences from the HepG2 secretome alone and in combination with plasma. Furthermore, it will be necessary to determine the structures of the glycans shown in the lectin enrichment study to differ between plasma and the HepG2 secretome. Information obtained from determining intact glycopeptide sequences may then be utilized in the determination of glycan structures; the glycopeptide weight minus the weight of the deglycosylated peptide gives the weight of the glycan(s) that inhabit the N-glycosite(s) on the peptide sequence. This offers a good starting point for beginning to elucidate the structure of the glycans. Through glycan structure elucidation, it will be possible to conclusively determine the differences in sialylation between plasma and the HepG2 secretome.

In conclusion, the results of the studies outlined in this dissertation support the use of the cancer cell secretome for biomarker discovery as demonstrated in the H1299 NSCLC study as well as the extension of the SILAC proteome to comparative plasma studies. Furthermore, the HepG2 study offers a foundation for building on the glycomics of the

HepG2 secretome as a platform for the generation of plasma proteins capable of use as internal standards.

Bibliography

References

[1] Girolamo F Di.; Lante I.; Muraca M.; Putignani L. The Role of Mass Spectrometry in the ‘ Omics ’ Era. *Curr Org Chem* 2013; 17: 2891–2905.

[2] Siuzdak G. An introduction to mass spectrometry ionization: An excerpt from The Expanding Role of Mass Spectrometry in Biotechnology, 2nd ed.; MCC Press: San Diego, 2005. *J Assoc Lab Autom* 2004; 9: 50–63.

[3] Crutchfield CA.; Clarke W. High resolution accurate mass (HRAM) mass spectrometry. In: *Mass Spectrometry for the Clinical Laboratory*. Elsevier, pp. 247–259.

[4] Perry RH.; Cooks RG.; Noll RJ. Orbitrap Mass Spectrometry: Instrumentation, Ion Motion And Applications. *Mass Spectrom Rev* 2008; 27: 661–699.

[5] Khatri K.; Klein JA.; Haserick JR.; Leon DR.; Costello CE.; McComb ME.; et al. Microfluidic Capillary Electrophoresis-Mass Spectrometry for Analysis of Monosaccharides, Oligosaccharides, and Glycopeptides. *Anal Chem* 2017; 89: 6645–6655.

[6] Rodrıguez-Suarez E.; Whetton AD. The Application of Quantification Techniques

in Proteomics for Biomedical Research. *Mass Spectrom Rev* 2013; 32: 1–26.

[7] Aebersold R.; Mann M. Mass spectrometry-based proteomics. *Nature* 2003; 422: 198–207.

[8] Sung H-J.; Cho J-Y. Biomarkers for the lung cancer diagnosis and their advances in proteomics. *BMB Rep* 2008; 41: 615–625.

[9] Pandey A.; Mann M. Proteomics to Study Genes and Genomes. *Nature* 2000; 405: 837–846.

[10] O'Farrell PH. High Resolution Two-Dimensional Electrophoresis of Proteins. *J Biol Chem* 1975; 250: 4007–4021.

[11] Hao Z.; Zhang Y.; Eliuk S.; Blethrow J.; Horn D.; Zabrouskov V.; et al. A Quadrupole-Orbitrap Hybrid Mass Spectrometer Offers Highest Benchtop Performance for In-Depth Analysis of Complex Proteomes. *Thermo Scientific Application Notes 552*. Epub ahead of print 2012. DOI: AN63515_E 04/12S.

[12] Bittremieux W.; Tabb DL.; Impens F.; Staes A.; Timmerman E.; Martens L.; et al. Quality control in mass spectrometry-based proteomics. *Mass Spectrom Rev* 2017; 1–15.

[13] El-Aneed A.; Cohen A.; Banoub J. Mass spectrometry, review of the basics: Electrospray, MALDI, and commonly used mass analyzers. *Appl Spectrosc Rev* 2009; 44: 210–230.

[14] Dronsfield A. Mass spectrometry - the early days. *Education in Chemistry*https://eic.rsc.org/feature/mass-spectrometry-the-early-

days/2020189.article (2010, accessed 24 January 2018).

[15] Borman S.; Russell H.; Siuzdak G. A Mass Spec Timeline. *Today's Chem Work* 2003; 47–49.

[16] Griffiths J. A Brief History of Mass Spectrometry. *Anal Chem* 2008; 80: 5678–5683.

[17] Finehout EJ.; Lee KH. An introduction to mass spectrometry applications in biological research. *Biochem Mol Biol Educ* 2004; 32: 93–100.

[18] Pitt JJ. Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry. *Clin Biochem Rev* 2009; 30: 19–34.

[19] Zhang Y.; Fonslow BR.; Shan B.; Baek M-C.; Yates JR. Protein analysis by shotgun/bottom-up proteomics. *Chem Rev* 2013; 113: 2343–94.

[20] Wilm M. Principles of Electrospray Ionization. *Mol Cell Proteomics* 2011; 10: M111.009407.

[21] Wilm MS.; Mann M. Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last? *Int J Mass Spectrom Ion Process* 1994; 136: 167–180.

[22] Wilm M.; Mann M. Analytical properties of the nanoelectrospray ion source. *Anal Chem* 1996; 68: 1–8.

[23] The Royal Swedish Academy of Sciences. Press Release: The Nobel Prize in Chemistry 2002. *Nobel Media AB*https://www.nobelprize.org/nobel_prizes/chemistry/laureates/2002/press.html (2002, accessed 6 December 2017).

[24] Fenn JB. Electrospray ionization mass spectrometry: How it all began. *J Biomol Tech* 2002; 13: 101–118.

[25] Dole M.; Mack LL.; Hines RL.; Mobley RC.; Ferguson LD.; Alice MB. Molecular Beams of Macroions. *J Chem Phys* 1968; 49: 2240–2249.

[26] Konermann L.; Ahadi E.; Rodriguez AD.; Vahidi S. Unraveling the mechanism of electrospray ionization. *Anal Chem* 2013; 85: 2–9.

[27] Thomson BA.; Iribarne J V. Field induced ion evaporation from liquid surfaces at atmospheric pressure. *J Chem Phys* 1979; 71: 4451–4463.

[28] Kebarle P. A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry. *J Mass Spectrom* 2000; 35: 804–817.

[29] Ho CS.; Lam CWK.; Chan MHM.; Cheung RCK.; Law LK.; Lit LCW.; et al. Electrospray Ionisation Mass Spectrometry : Principles and Clinical Applications. *Clin Biochem Rev* 2003; 24: 3–12.

[30] Cañas Montalvo B.; López-Ferrer D.; Ramos-Fernández A.; Camafeita E.; Calvo E. Mass spectrometry technologies for proteomics. *Briefings Funct Genomics Proteomics* 2006; 4: 295–320.

[31] Karas M.; Hillenkamp F. Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10 000 Daltons. *Anal Chem* 1988; 60: 2299–2301.

[32] Karas M.; Krüger R. Ion formation in MALDI: The cluster ionisation mechanism. *Chemical Reviews* 2003; 103: 427–439.

[33] Tanaka K.; Waki H.; Ido Y.; Akita S.; Yoshida Y.; Yoshida T.; et al. Protein and

polymer analyses up to *m/z* 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* 1988; 2: 151–153.

[34] Michalski A.; Damoc E.; Hauschild J-P.; Lange O.; Wieghaus A.; Makarov A.; et al. Mass Spectrometry-based Proteomics Using Q Exactive, a High-performance Benchtop Quadrupole Orbitrap Mass Spectrometer. *Mol Cell Proteomics* 2011; 10: M111.011015.

[35] Yates JR.; Gilchrist A.; Howell KE.; Bergeron JJM. Proteomics of organelles and large cellular structures. *Nat Rev Mol Cell Biol* 2005; 6: 702–14.

[36] Mann M.; Kelleher NL. Precision proteomics: the case for high resolution and high mass accuracy. *Proc Natl Acad Sci U S A* 2008; 105: 18132–8.

[37] Grayson M. History Posters: 1st Annual ASMS Conference. In: *Annual Meeting of the American Society for Mass Spectrometry*. 1953.

[38] March RE. An Introduction to Quadrupole Ion Trap Mass Spectrometry. *J Mass Spectrom* 1997; 32: 351–369.

[39] Marshall AG.; Hendrickson CL.; Jackson GS. Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom Rev* 1998; 17: 1–35.

[40] Hipple JA.; Sommer H.; Thomas HA. A precise method of determining the faraday by magnetic resonance. *Phys Rev* 1949; 76: 1877–1878.

[41] Makarov A. Electrostatic axially harmonic orital trapping a high performance technique of mass analysis. *Anal Chem* 2000; 72: 1156–1162.

[42] Hardman M.; Makarov AA. Interfacing the orbitrap mass analyzer to an

electrospray ion source. *Anal Chem* 2003; 75: 1699–1705.

[43] Catherman AD.; Skinner OS.; Kelleher NL. Top Down proteomics: Facts and perspectives. *Biochem Biophys Res Commun* 2014; 445: 683–693.

[44] ThermoFisher Scientific. In-Depth Characterization of Structurally Diverse Targets. 2017; 1–2.

[45] Gallien S.; Duriez E.; Crone C.; Kellmann M.; Moehring T.; Domon B. Targeted proteomic quantification on quadrupole-orbitrap mass spectrometer. *Mol Cell proteomics* 2012; 11: 1709–23.

[46] Hawkridge AM. Practical Considerations and Current Limitations in Quantitative Mass Spectrometry-based Proteomics. In: Eyers CE, Gaskell SJ (eds) *Quantitative Proteomics*. Cambridge: The Royal Society of Chemistry, 2014, pp. 3–25.

[47] Hsieh EJ.; Bereman MS.; Durand S.; Valaskovic GA.; MacCoss MJ. Effects of Column and Gradient Lengths on Peak Capacity and Peptide Identification in nanoflow LC-MS/MS of Complex Proteomic Samples. *J Am Soc Mass Spectrom* 2013; 24: 148–153.

[48] Ettre LS. M . S . Tswett and the Invention of Chromatography. *LGGC Eur* 2003; 2–7.

[49] Horvath CG.; Preiss BA.; Lipsky SR. Fast Liquid Chromatography: An Investigation of Operating Parameters and the Separation of Nucleotides on Pellicular Ion Exchangers. *Anal Chem* 1967; 39: 1422–1428.

[50] Gama MR.; Collins CH.; Bottoli CBG. Nano-liquid chromatography in pharmaceutical and biomedical research. *J Chromatogr Sci* 2013; 51: 694–703.

[51] Chervet JP.; Ursem M.; Salzmann JP. Instrumental requirements for nanoscale liquid chromatography. *Anal Chem* 1996; 68: 1507–1512.

[52] Swart R.; Rieux L.; Sneekes E-J. Nano LC: Principles, Evolution, and State-of-the-Art of the Technique. *LCGC North Am* 2011; 29: 926–934.

[53] Liu Q.; Cobb JS.; Johnson JL.; Wang Q.; Agar JN. Performance Comparisons of Nano-LC Systems , Electrospray Sources and LC – MS-MS Platforms. 2017; 120–127.

[54] Meek JL.; Rossetti ZL. Factors affecting retention and resolution of peptides in high-performance liquid chromatography. *J Chromatogr A* 1981; 211: 15–28.

[55] Resing KA.; Ahn NG. Proteomics strategies for protein identification. *FEBS Lett* 2005; 579: 885–889.

[56] Gevaert K.; Vanderkerckhove J. Protein identification methods in proteomics. *Electrophoresis* 2000; 21: 1145–1154.

[57] Edman P. Method for determination of the amino acid sequence in peptides. *Acta Chemica Scandinavica* 1950; 4: 283–293.

[58] Schaab C.; Geiger T.; Stoehr G.; Cox J.; Mann M. Analysis of High Accuracy , Quantitative Proteomics Data in the MaxQB Database. *Mol Cell Proteomics* 2012; 11: 1–10.

[59] Sadygov RG.; Cociorva D.; Yates JR. Large-scale database searching using

tandem mass spectra: looking up the answer in the back of the book. *Nat Methods* 2004; 1: 195–202.

[60] Zhou W.; Liotta LA.; Petricon EF. The Spectra Count Label-free Quantitation in Cancer Proteomics. *Cancer Genomics Proteomics* 2012; 9: 135–142.

[61] Martens L. Bioinformatics Challenges in Mass Spectrometry-Driven Proteomics. In: Gevaert K, Vandekerckhove J (eds) *Gel-Free Proteomics: Methods and Protocols*. New York: Humana Press, 2011, pp. 359–371.

[62] Perkins DN.; Pappin DJC.; Creasy DM.; Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 1999; 20: 3551–3567.

[63] Bern M.; Goldberg D.; McDonald WH.; Yates JR. Automatic Quality Assessment of Peptide Tandem Mass Spectra. *Bioinformatics* 2004; 20: i49–i54.

[64] Matrix Science. Mascot database search: MS/MS Results Interpretationhttp://www.matrixscience.com/help/interpretation_help.html (accessed 3 November 2015).

[65] Neuhauser N.; Michalski A.; Scheltema RA.; Olsen J V.; Mann M. Andromeda : A Peptide Search Engine Integrated into the MaxQuant Environment. 2011; 1794–1805.

[66] Nakamura T.; Oda Y. Mass spectrometry-based quantitative proteomics. *Biotechnol Genet Eng Rev* 2007; 24: 147–164.

[67] Hein MY.; Luber CA.; Paron I.; Nagaraj N.; Mann M. Accurate Proteome-wide

Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction , Termed MaxLFQ. *Mol Cell Proteomics* 2014; 13: 2513–2526.

[68] Ruse C.; Chen EI. Strategies for large-scale quantitative proteomics analyses. *Proteomics Genomics Res* 2014; 1: 4–14.

[69] Altelaar AFM.; Munoz J.; Heck AJR. Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat Rev Genet* 2012; 14: 35–48.

[70] Bantscheff M.; Schirle M.; Sweetman G.; Rick J.; Kuster B. Quantitative mass spectrometry in proteomics: A critical review. *Anal Bioanal Chem* 2007; 389: 1017–1031.

[71] Hawkridge AM.; Wysocky RB.; Petitte JN.; Anderson KE.; Mozdziak PE.; Fletcher OJ.; et al. Measuring the intra-individual variability of the plasma proteome in the chicken model of spontaneous ovarian adenocarcinoma. *Anal Bioanal Chem* 2010; 398: 737–49.

[72] Zhu W.; Smith JW.; Huang CM. Mass spectrometry-based label-free quantitative proteomics. *J Biomed Biotechnol*; 2010. Epub ahead of print 2010. DOI: 10.1155/2010/840518.

[73] Li Z.; Adams RM.; Chourey K.; Hurst GB.; Hettich RL.; Pan C. Systematic Comparison of Label-Free, Metabolic Labeling, and Isobaric Chemical Labeling for Quantitative Proteomics on LTQ Orbitrap Velos. *J Proteome Res* 2012; 11: 1582–1590.

[74] Caruso MB.; Trugilho MRO.; Higa LM.; Teixeira-Ferreira AS.; Perales J.; Da

Poian AT.; et al. Proteomic analysis of the secretome of HepG2 cells indicates differential proteolytic processing after infection with dengue virus. *J Proteomics* 2016; 151: 106–113.

[75] Sury MD.; Chen J-X.; Selbach M. The SILAC fly allows for accurate protein quantification in vivo. *Mol Cell proteomics* 2010; 9: 2173–83.

[76] Krijgsveld J.; Ketting RF.; Mahmoudi T.; Johansen J.; Artal-Sanz M.; Verrijzer CP.; et al. Metabolic labeling of C. elegans and D. melanogaster for quantitative proteomics. *Nat Biotechnol* 2003; 21: 927–931.

[77] Gouw JW.; Tops BBJ.; Krijgsveld J. Metabolic Labeling of Model Organisms Using Heavy Nitrogen (15N). In: Gevaert K, Vandekerckhove J (eds) *Gel-Free Proteomics: Methods and Protocols*. New York: Humana Press, 2011, pp. 29–42.

[78] Ong S-E.; Blagoev B.; Kratchmarova I.; Kristensen DB.; Steen H.; Pandey A.; et al. Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Mol Cell Proteomics* 2002; 1: 376–386.

[79] Ong SE.; Mann M. Stable isotope labeling by amino acids in cell culture for quantitative proteomics. In: Sechi S (ed) *Quantitative Proteomics by Mass Spectrometry. Methods in Molecular Biology*. Totowa, NJ: Humana Press, pp. 37–52.

[80] Boersema PJ.; Geiger T.; Wisniewski JR.; Mann M. Quantification of the N-glycosylated Secretome by Super-SILAC During Breast Cancer Progression and in Human Blood Samples. *Mol Cell Proteomics* 2013; 12: 158–171.

[81] Deeb SJ.; Cox J.; Schmidt-Supprian M.; Mann M. N-linked glycosylation enrichment for in-depth cell surface proteomics of diffuse large B-cell lymphoma subtypes. *Mol Cell Proteomics* 2014; 13: 240–51.

[82] Mangrum JB.; Martin EJ.; Brophy DF.; Hawkridge AM. Intact stable isotope labeled plasma proteins from the SILAC-labeled HepG2 secretome. *Proteomics* 2015; 15: 3104–3115.

[83] Aragão-Leoneti V.; Campo VL.; Gomes AS.; Field R a.; Carvalho I. Application of copper(I)-catalysed azide/alkyne cycloaddition (CuAAC) ‘click chemistry’ in carbohydrate drug and neoglycopolymer synthesis. *Tetrahedron* 2010; 66: 9475–9492.

[84] Liang L.; Astruc D. The copper(I)-catalyzed alkyne-azide cycloaddition (CuAAC) ‘click’ reaction and its applications. An overview. *Coord Chem Rev* 2011; 255: 2933–2945.

[85] Zhang G.; Neubert TA. Use of Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC) for Phosphotyrosine Protein Identification and Quantitation. *Methods Mol Biol* 2009; 527: 79–90.

[86] Kramer G.; Sprenger RR.; Back J.; Dekker HL.; Nessen MA.; van Maarseveen JH.; et al. Identification and quantitation of newly synthesized proteins in Escherichia coli by enrichment of azidohomoalanine-labeled peptides with diagonal chromatography. *Mol Cell Proteomics* 2009; 8: 1599–611.

[87] Dieterich DC.; Hodas JJL.; Gouzer G.; Shadrin IY.; Ngo JT.; Triller A.; et al. In situ visualization and dynamics of newly synthesized proteins in rat hippocampal

neurons. *Nat Neurosci* 2011; 13: 897–905.

[88] Thompson A.; Schäfer J.; Kuhn K.; Kienle S.; Schwarz J.; Schmidt G.; et al. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* 2003; 75: 1895–904.

[89] Rauniyar N.; Yates JR. Isobaric Labeling-Based Relative Quantification in Shotgun Proteomics. *J Proteome Res* 2014; 13: 5293–5309.

[90] ThermoFisher Scientific. Quantitative Proteomicshttps://www.thermofisher.com/us/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/quantitative-proteomics.html (2015, accessed 6 November 2015).

[91] Werner T.; Sweetman G.; Savitski MF.; Mathieson T.; Bantscheff M.; Savitski MM. Ion coalescence of neutron encoded TMT 10-plex reporter ions. *Anal Chem* 2014; 86: 3594–601.

[92] Chong PK.; Gan CS.; Pham TK.; Wright PC. Isobaric tags for relative and absolute quantitation (iTRAQ) reproducibility: Implication of multiple injections. *J Proteome Res* 2006; 5: 1232–40.

[93] Pavlou MP.; Diamandis EP. The cancer cell secretome: A good source for discovering biomarkers? *J Proteomics* 2010; 73: 1896–1906.

[94] Faça VM.; Palma CS.; Albuquerque D.; Canchaya GNS.; Grassi ML.; Epifânio VL.; et al. The secretome analysis by high-throughput proteomics and multiple

reaction monitoring (MRM). In: Martins-de-Souza D (ed) *Shotgun Proteomics: Methods and Protocols*. Humana, pp. 323–35.

[95] Ibrahim SA.; Kulshrestha A.; Katara GK.; Amin MA.; Beaman KD. Cancer derived peptide of vacuolar ATPase 'a2' isoform promotes neutrophil migration by autocrine secretion of IL-8. *Sci Rep* 2016; 6: 36865.

[96] Cudjoe EK.; Saleh T.; Hawkridge AM.; Gewirtz DA. Proteomics Insights into Autophagy. *Proteomics*; 17. Epub ahead of print 2017. DOI: 10.1002/pmic.201700022.

[97] Massion PP.; Caprioli RM. Proteomic strategies for the characterization and the early detection of lung cancer. *J Thorac Oncol* 2006; 1: 1027–1039.

[98] Rabouille C. Pathways of Unconventional Protein Secretion. *Trends Cell Biol* 2017; 27: 230–240.

[99] Abrahamsen H.; Stenmark H. Protein Secretion: Unconventional Exit by Exophagy. *Curr Biol* 2010; 20: R415–R418.

[100] Benham A. Protein secretion and the endoplasmic reticulum. *Cold Spring Harb Perspect Biol*; 4. Epub ahead of print 2012. DOI: 10.1101/cshperspect.a012872.

[101] Ng F.; Tang BL. Unconventional Protein Secretion in Animal Cells. In: Pompa A, Marchis F De (eds) *Unconventional Protein Secretion: Methods and Protocols*. New York: Humana Press, pp. 31–46.

[102] Méndez O.; Villanueva J. Challenges and opportunities for cell line secretomes in cancer proteomics. *Proteomics - Clin Appl* 2015; 9: 348–357.

[103] Grønborg M.; Kristiansen TZ.; Iwahori A.; Chang R.; Reddy R.; Sato N.; et al. Biomarker discovery from pancreatic cancer secretome using a differential proteomic approach. *Mol Cell Proteomics* 2006; 5: 157–71.

[104] Taniguchi N.; Korekane H. Branched N-glycans and their implications for cell adhesion, signaling and clinical applications for cancer biomarkers and in therapeutics. *BMB Rep* 2011; 44: 772–781.

[105] Lemjabbar-Alaoui H.; McKinney A.; Yang YW.; Tran VM.; Phillips JJ. *Glycosylation alterations in lung and brain cancer*. 1st ed. Elsevier Inc. Epub ahead of print 2015. DOI: 10.1016/bs.acr.2014.11.007.

[106] Clerc F.; Reiding KR.; Jansen BC.; Kammeijer GSM.; Bondt A.; Wuhrer M. Human plasma protein N-glycosylation. *Glycoconj J* 2016; 33: 309–343.

[107] Sun S.; Shah P.; Eshghi ST.; Yang W.; Trikannad N.; Yang S.; et al. Comprehensive analysis of protein glycosylation by solid-phase extraction of N-linked glycans and glycosite-containing peptides. *Nat Biotechnol* 2016; 34: 84–88.

[108] Sleat DE.; Donnelly RJ.; Lackland H.; Liu C-GG.; Sohar I.; Pullarkat RK.; et al. Association of mutations in a lysosomal protein with classical late-infantile neuronal ceroid lipofuscinosis. *Science* 1997; 277: 1802–5.

[109] Loziuk PL.; Hecht ES.; Muddiman DC. N-linked glycosite profiling and use of Skyline as a platform for characterization and relative quantification of glycans in differentiating xylem of Populus trichocarpa. *Anal Bioanal Chem* 2016; 409: 487–497.

[110] Ryle AJ.; Davie S.; Gould BJ.; Yudkin JS. A study of the effect of diet on glycosylated haemoglobin and albumin levels and glucose tolerance in normal subjects. *Diabet Med* 1990; 7: 865–70.

[111] Aizpurua-Olaizola O.; Sastre Toraño J.; Falcon-Perez JM.; Williams C.; Reichardt N.; Boons GJ. Mass spectrometry for glycan biomarker discovery. *TrAC - Trends Anal Chem* 2018; 100: 7–14.

[112] Gornik O.; Wagner J.; Pučić M.; Knežević A.; Redžić I.; Lauc G. Stability of N-glycan profiles in human plasma. *Glycobiology* 2009; 19: 1547–1553.

[113] Huhn C.; Selman MHJ.; Ruhaak LR.; Deelder AM.; Wuhrer M. IgG glycosylation analysis. *Proteomics* 2009; 9: 882–913.

[114] Varki A.; Sharon N. Historical Background and Overview. In: Varki A, Cummings R, Esko J, et al. (eds) *Essentials of Glycobiology*. New York: Cold Spring Harbor Laboratory Press, 2009. Epub ahead of print 2009. DOI: 10.1101/glycobiology.3e.001.

[115] Harvey DJ. Derivatization of carbohydrates for analysis by chromatography; electrophoresis and mass spectrometry. *J Chromatogr B Anal Technol Biomed Life Sci* 2011; 879: 1196–1225.

[116] Brockhausen I.; Schachter H.; Stanley P. O-GalNAc Glycans. In: Varki A, Cummings R, Esko J, et al. (eds) *Essentials of Glycobiology*. New York: Cold Spring Harbor Laboratory Presshttp://www.ncbi.nlm.nih.gov/pubmed/20301232 (2009, accessed 13 February 2018).

[117] Stanley P.; Schachter H.; Taniguchi N. N Glycans. In: Varki A, Cummings R, Esko J, et al. (eds) *Essentials of Glycobiology*. New York: Cold Spring Harbor Laboratory Press, 2009.

[118] Zhu R.; Zacharias L.; Wooding KM.; Peng W.; Mechref Y. CHAPTER 7: Glycoprotein Enrichment Analytical Techniques: Advantages and Disadvantages. *Methods Enzymol* 2017; 585: 397–429.

[119] Walker SH.; Lilley LM.; Enamorado MF.; Comins DL.; Muddiman DC. Hydrophobic derivatization of N-linked glycans for increased ion abundance in electrospray ionization mass spectrometry. *J Am Soc Mass Spectrom* 2011; 22: 1309–1317.

[120] Qin X.; Chen Q.; Sun C.; Wang C.; Peng Q.; Xie L.; et al. High-throughput screening of tumor metastatic-related differential glycoprotein in hepatocellular carcinoma by iTRAQ combines lectin-related techniques. *Med Oncol*; 30. Epub ahead of print 2013. DOI: 10.1007/s12032-012-0420-8.

[121] Hongsachart P.; Huang-Liu R.; Sinchaikul S.; Pan FM.; Phutrakul S.; Chuang YM.; et al. Glycoproteomic analysis of WGA-bound glycoprotein biomarkers in sera from patients with lung adenocarcinoma. *Electrophoresis* 2009; 30: 1206–1220.

[122] Wang Y.; Ao X.; Vuong H.; Konanur M.; Miller FR.; Goodison S.; et al. Membrane glycoproteins associated with breast tumor cell progression identified by a lectin affinity approach. *J Proteome Res* 2008; 7: 4313–4325.

[123] Monsigny M.; Roche A -C.; Sene C.; Maget-Dana R.; Delmotte F. Sugar-Lectin

Interactions: How Does Wheat-Germ Agglutinin Bind Sialoglycoconjugates? *Eur J Biochem* 1980; 104: 147–153.

[124] Gewirtz DA. An autophagic switch in the response of tumor cells to radiation and chemotherapy. *Biochem Pharmacol* 2014; 90: 208–11.

[125] Gewirtz DA. The Four Faces of Autophagy: Implications for Cancer Therapy. *Cancer Res* 2014; 74: 647–651.

[126] Sharma K.; Goehe RW.; Di X.; Hicks MA.; Torti S V.; Torti FM.; et al. A novel cytostatic form of autophagy in sensitization of non-small cell lung cancer cells to radiation by vitamin D and the vitamin D analog, EB 1089. *Autophagy* 2015; 10: 2346–2361.

[127] Poklepovic A.; Gewirtz DA. Outcome of early clinical trials of the combination of hydroxychloroquine with chemotherapy in cancer. *Autophagy* 2014; 10: 1478–1480.

[128] Geiger T.; Cox J.; Ostasiewicz P.; Wisniewski JR.; Mann M. Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nat Methods* 2010; 7: 383–385.

[129] Tebbe A.; Klammer M.; Sighart S.; Schaab C.; Daub H. Systematic evaluation of label-free and super-SILAC quantification for proteome expression analysis. *Rapid Commun Mass Spectrom* 2015; 29: 795–801.

[130] Rifai N.; Gillette MA.; Carr SA. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat Biotechnol* 2006; 24: 971–83.

[131] American Association for Cancer Research (AACR). Nobel Prize Honors

Autophagy Discovery. *Cancer Discov* 2016; 6: 1298–1299.

[132] Mizushima N. Autophagy: Process and function. *Genes Dev* 2007; 21: 2861–2873.

[133] Cuervo AM. Autophagy: In sickness and in health. *Trends Cell Biol* 2004; 14: 70–77.

[134] Mizushima N.; Levine B.; Cuervo AM.; Klionsky DJ. Autophagy fights disease through cellular self-digestion. *Nature* 2008; 451: 1069–75.

[135] Klionsky DJ.; Ohsumi Y. Vacuolar Import of Proteins and Organelles from the Cytoplasm. *Annu Rev Cell Dev Biol* 1999; 15: 1–32.

[136] Yang Z.; Klionsky DJ. Eaten alive: a history of macroautophagy. *Nat Cell Biol* 2010; 12: 814–22.

[137] Li WW.; Li J.; Bao JK. Microautophagy: Lesser-known self-eating. *Cell Mol Life Sci* 2012; 69: 1125–1136.

[138] Cuervo AM.; Wong E. Chaperone-mediated autophagy: roles in disease and aging. *Cell Res* 2014; 24: 92–104.

[139] Boya P.; Gonzalez-Polo, Rosa-Ana Casares N.; Perfettini J-L.; Dessen P.; Larochette N.; Métivier D.; et al. Inhibition of Macroautophagy Triggers Apoptosis Inhibition of Macroautophagy Triggers Apoptosis. *Mol Cell Biol* 2005; 25: 1025–1040.

[140] Susmita Kaushik.; Cuervo AM. Chaperone-mediated autophagy: a unique way to enter the lysosome world. *Trends Cell Biol* 2012; 22: 407–417.

[141] Janku F.; McConkey DJ.; Hong DS.; Kurzrock R. Autophagy as a target for anticancer therapy. *Nat Rev Clin Oncol* 2011; 8: 528–539.

[142] Alotaibi M.; Sharma K.; Saleh T.; Povirk LF.; Hendrickson EA.; Gewirtz DA. Radiosensitization by PARP Inhibition in DNA Repair Proficient and Deficient Tumor Cells: Proliferative Recovery in Senescent Cells. *Radiat Res* 2016; 185: 229–245.

[143] Saleh T.; Cuttino L.; Gewirtz DA. Autophagy is not uniformly cytoprotective: a personalized medicine approach for autophagy inhibition as a therapeutic strategy in non-small cell lung cancer. *Biochimica et Biophysica Acta - General Subjects* 2016; 1860: 2130–2136.

[144] Dou Q.; Chen HN.; Wang K.; Yuan K.; Lei Y.; Li K.; et al. Ivermectin induces cytostatic autophagy by blocking the PAK1/Akt Axis in breast cancer. *Cancer Res* 2016; 76: 4457–4469.

[145] Sharma K.; Le N.; Alotaibi M.; Gewirtz DA. Cytotoxic autophagy in cancer therapy. *Int J Mol Sci* 2014; 15: 10034–51.

[146] Wong PM.; Puente C.; Ganley IG.; Jiang X. The ULK1 complex sensing nutrient signals for autophagy activation. *Autophagy* 2013; 9: 124–137.

[147] Pyo J-O.; Nah J.; Jung Y-K. Molecules and their functions in autophagy. *Exp Mol Med* 2012; 44: 73.

[148] He C.; Klionsky DJ. Regulation Mechanisms and Signaling Pathways of Autophagy. *Annu Rev Genet* 2010; 43: 67–93.

[149] Mancias JD.; Wang X.; Gygi SP.; Harper JW.; Kimmelman AC. Quantitative proteomics identifies NCOA4 as the cargo receptor mediating ferritinophagy. *Nature* 2014; 509: 105–9.

[150] Sarkar S.; Perlstein EO.; Imarisio S.; Pineau S.; Cordenier A.; Maglathlin RL.; et al. Small molecules enhance autophagy and reduce toxicity in Huntington's disease models. *Nat Chem Biol* 2009; 3: 331–338.

[151] Mercer CA.; Kaliappan A.; Dennis PB. A novel, human Atg13 binding protein, Atg101, interacts with ULK1 and is essential for macroautophagy. *Autophagy* 2009; 5: 649–62.

[152] Hosokawa N.; Sasaki T.; Iemura S.; Natsume T.; Hara T.; Mizushima N. Atg101, a novel mammalian autophagy protein interacting with Atg13. *Autophagy* 2009; 5: 973–9.

[153] Russell RC.; Tian Y.; Yuan H.; Park HW.; Chang Y-Y.; Kim J.; et al. ULK1 induces autophagy by phosphorylating Beclin-1 and activating VPS34 lipid kinase. *Nat Cell Biol* 2013; 15: 741–50.

[154] Yang ZJ.; Chee CE.; Huang S.; Sinicrope FA. The Role of Autophagy in Cancer: Therapeutic Implications. *Mol Cancer Ther* 2011; 10: 1533–1541.

[155] Tooze SA.; Abada A.; Elazar Z. Endocytosis and autophagy: exploitation or cooperation? *Cold Spring Harb Perspect Biol* 2014; 6: 1–15.

[156] Nishida Y.; Arakawa S.; Fujitani K.; Yamaguchi H.; Mizuta T.; Kanaseki T.; et al. Discovery of Atg5/Atg7-independent alternative macroautophagy. *Nature* 2009;

461: 654–658.

[157] Zimmermann AC.; Zarei M.; Eiselein S.; Dengjel J. Quantitative proteomics for the analysis of spatio-temporal protein dynamics during autophagy. *Autophagy* 2010; 6: 1009–1016.

[158] Behrends C.; Sowa ME.; Gygi SP.; Harper JW. Network organization of the human autophagy system. *Nature* 2010; 466: 68–76.

[159] Dengjel J.; Høyer-Hansen M.; Nielsen MMO.; Eisenberg T.; Harder LM.; Schandorff S.; et al. Identification of autophagosome-associated proteins and regulators by quantitative proteomic analysis and genetic screens. *Mol Cell Proteomics* 2012; 11: 1–17.

[160] Zhao W.; Liu Z.; Yu X.; Lai L.; Li H.; Liu Z.; et al. iTRAQ proteomics analysis reveals that PI3K is highly associated with bupivacaine-induced neurotoxicity pathways. *Proteomics* 2016; 16: 564–575.

[161] Kraya AA.; Piao S.; Xu X.; Zhang G.; Herlyn M.; Gimotty P.; et al. Identification of secreted proteins that reflect autophagy dynamics within tumor cells. *Autophagy* 2015; 11: 60–74.

[162] Mathew R.; Khor S.; Hackett SR.; Rabinowitz JD.; Perlman DH.; White E.; et al. Functional Role of Autophagy-Mediated Proteome Remodeling in Cell Survival Signaling and Innate Immunity. *Mol Cell* 2014; 55: 916–930.

[163] Antonioli M.; Ciccosanti F.; Dengjel J.; Fimia GM. Methods to Study the BECN1 Interactome in the Course of Autophagic Responses. In: Galluzzi L, Pedro JMB-S,

Kroemer G (eds) *Methods in Enzymology*. Elsevier Inc., pp. 429–445.

[164] Zhong L.; Zhou J.; Chen X.; Liu J.; Liu Z.; Chen Y.; et al. Quantitative proteomics reveals EVA1A-related proteins involved in neuronal differentiation. *Proteomics* 2016; 17: 1600294.

[165] Yu Y.; Li T.; Wu N.; Jiang L.; Ji X.; Huang H. The Role of Lipid Droplets in Mortierella alpina Aging Revealed by Integrative Subcellular and Whole-Cell Proteome Analysis. *Sci Rep* 2017; 7: 43896.

[166] Wang J.; Zhang J.; Lee YM.; Koh PL.; Ng S.; Bao F.; et al. Quantitative chemical proteomics profiling of de novo protein synthesis during starvation-mediated autophagy. *Autophagy* 2016; 12: 1931–1944.

[167] Zhang J.; Wang J.; Lee Y-M.; Lim T-K.; Lin Q.; Shen H-M. Proteomic Profiling of De Novo Protein Synthesis in Starvation- Induced Autophagy Using Bioorthogonal Noncanonical Amino Acid Tagging. In: *Methods in Enzymology*. Elsevier Inc., pp. 41–59.

[168] Zhang X.; Belkina N.; Jacob HKC.; Maity T.; Biswas R.; Venugopalan A.; et al. Identifying novel targets of oncogenic EGF receptor signaling in lung cancer through global phosphoproteomics. *Proteomics* 2015; 15: 340–355.

[169] Rodolfo C.; Rocco M.; Cattaneo L.; Tartaglia M.; Sassi M.; Aducci P.; et al. Ophiobolin A Induces Autophagy and Activates the Mitochondrial Pathway of Apoptosis in Human Melanoma Cells. *PLoS One* 2016; 11: e0167672.

[170] Wang J.; Zhang J.; Zhang C-J.; Wong YK.; Lim TK.; Hua Z-C.; et al. In situ

Proteomic Profiling of Curcumin Targets in HCT116 Colon Cancer Cell Line. *Sci Rep* 2016; 6: 22146.

[171] Kang J-H.; Li M.; Chen X.; Yin X-M. Proteomics Analysis of Starved Cells Revealed Annexin A1 as an Important Regulator of Autophagic Degradation. *Biochem Biophys Res Commun* 2011; 407: 581–586.

[172] Zhuo C.; Ji Y.; Chen Z.; Kitazato K.; Xiang Y.; Zhong M.; et al. Proteomics analysis of autophagy-deficient Atg7-/- MEFs reveals a close relationship between F-actin and autophagy. *Biochem Biophys Res Commun* 2013; 437: 482–488.

[173] Sanjuan MA.; Dillon CP.; Tait SWG.; Moshiach S.; Dorsey F.; Connell S.; et al. Toll-like receptor signalling in macrophages links the autophagy pathway to phagocytosis. *Nature* 2007; 450: 1253–7.

[174] Delgado MA.; Elmaoued RA.; Davis AS.; Kyei G.; Deretic V. Toll-like receptors control autophagy. *The EMBO Journal* 2008; 27: 1110–1121.

[175] Bertin S.; Samson M.; Pons C.; Guigonis J.; Gavelli A.; Baqué P.; et al. Comparative Proteomics Study Reveals That Bacterial CpG Motifs Induce Tumor Cell Autophagy in Vitro and in Vivo. *Mol Cell Proteomics* 2008; 7: 2311–2322.

[176] Delgado MA.; Deretic V. Toll-like receptors in control of immunological autophagy. *Cell Death Differ* 2009; 16: 976–983.

[177] Li1 L.; Jin H.; Wang H.; Cao Z.; Feng N.; Wang J.; et al. Autophagy is highly targeted among host comparative proteomes during infection with different virulent RABV strains. *Oncotarget* 2017; 8: 21336–21350.

[178] Patella F.; Neilson LJ.; Athineos D.; Erami Z.; Anderson KI.; Blyth K.; et al. In-Depth Proteomics Identifies a Role for Autophagy in Controlling Reactive Oxygen Species Mediated Endothelial Permeability. *J Proteome Res* 2016; 15: 2187–2197.

[179] Lampugnani MG.; Corada M.; Andriopoulou P.; Esser S.; Risau W.; Dejana E. Cell confluence regulates tyrosine phosphorylation of adherens junction components in endothelial cells. *J Cell Sci* 1997; 110: 2065–2077.

[180] Hu J.; Li G.; Qu L.; Li N.; Liu W.; xia D.; et al. TMEM166/EVA1A interacts with ATG16L1 and induces autophagosome formation and cell death. *Cell Death Dis* 2016; 7: e2323.

[181] Klionsky D. Guidelines for the use and interpretation of assays for monitoring autophagy. *Autophagy* 2012; 8: 445–544.

[182] Nilsson P.; Saido TC. Dual roles for autophagy: Degradation and secretion of Alzheimer's disease Aβ peptide. *BioEssays* 2014; 36: 570–578.

[183] Ubersax JA.; Ferrell JE.; Jr JEF. Mechanisms of specificity in protein phosphorylation. *Nat Rev Mol Cell Biol* 2007; 8: 530–41.

[184] Kim J-Y.; Stewart P.; Borne A.; Fang B.; Welsh E.; Chen Y.; et al. Activity-Based Proteomics Reveals Heterogeneous Kinome and ATP-Binding Proteome Responses to MEK Inhibition in KRAS Mutant Lung Cancer. *Proteomes* 2016; 4: 16.

[185] Yao W.; Yue P.; Zhang G.; Owonikoko TK.; Khuri FR.; Sun SY. Enhancing

therapeutic efficacy of the MEK inhibitor, MEK162, by blocking autophagy or inhibiting PI3K/Akt signaling in human lung cancer cells. *Cancer Lett* 2015; 364: 70–78.

[186] Eng C.; Abraham R. The autophagy conundrum in cancer: influence of tumorigenic metabolic reprogramming. *Oncogene* 2011; 30: 4687–4696.

[187] Watson AS.; Riffelmacher T.; Stranks A.; Williams O.; De Boer J.; Cain K.; et al. Autophagy limits proliferation and glycolytic metabolism in acute myeloid leukemia. *Cell Death Discov* 2015; 1: 15008.

[188] Beevers CS.; Chen L.; Liu L.; Luo Y.; Webster NJG.; Huang S. Curcumin disrupts the mammalian target of rapamycin-raptor complex. *Cancer Res* 2009; 69: 1000–1008.

[189] Høyer-Hansen M.; Bastholm L.; Szyniarowski P.; Campanella M.; Szabadkai G.; Farkas T.; et al. Control of Macroautophagy by Calcium, Calmodulin-Dependent Kinase Kinase-β, and Bcl-2. *Molecular Cell* 2007; 25: 193–205.

[190] Tavera-Mendoza LE.; Westerling T.; Libby E.; Marusyk A.; Cato L.; Cassani R.; et al. Vitamin D receptor regulates autophagy in the normal mammary gland and in luminal breast cancer cells. *Proc Natl Acad Sci* 2017; E2186–E2194.

[191] Mohammed H.; Santos CD.; Serandour AA.; Ali HR.; Brown GD.; Atkins A.; et al. Endogenous Purification Reveals GREB1 as a Key Estrogen Receptor Regulatory Factor. *Cell Rep* 2013; 3: 342–349.

[192] Anders Øverbye.; Monica Fengsrud.; Per O. Seglen. Proteomic Analysis of

Membrane-Associated Proteins from Rat Liver Autophagosomes. *Autophagy* 2007; 3: 300–322.

[193] Gao W.; Kang JH.; Liao Y.; Ding WX.; Gambotto AA.; Watkins SC.; et al. Biochemical isolation and characterization of the tubulovesicular LC3-positive autophagosomal compartment. *J Biol Chem* 2010; 285: 1371–1383.

[194] Strømhaug PE.; Berg TO.; Fengsrud M.; Seglen PO. Purification and characterization of autophagosomes from rat hepatocytes. *Biochem J* 1998; 335: 217–224.

[195] Suzuki K.; Nakamura S.; Morimoto M.; Fujii K.; Noda NN.; Inagaki F.; et al. Proteomic profiling of autophagosome cargo in Saccharomyces cerevisiae. *PLoS One* 2014; 9: 1–9.

[196] Bagshaw RD.; Mahuran DJ.; Callahan JW. A Proteomic Analysis of Lysosomal Integral Membrane Proteins Reveals the Diverse Composition of the Organelle. *Mol Cell Proteomics* 2004; 4: 133–143.

[197] Lübke T.; Lobel P.; Sleat DE. Proteomics of the lysosome. *Biochim Biophys Acta - Mol Cell Res* 2009; 1793: 625–635.

[198] Lieberman AP.; Puertollano R.; Raben N.; Slaugenhaupt S.; Walkley SU.; Ballabio A. Autophagy in lysosomal storage disorders. *Autophagy* 2012; 8: 719–730.

[199] Schröder BA.; Wrocklage C.; Hasilik A.; Saftig P. The proteome of lysosomes. *Proteomics* 2010; 10: 4053–4076.

[200] Jaquinod SK.; Chapel A.; Garin J.; Journet A. Affinity purification of soluble lysosomal proteins for mass spectrometric identification. *Methods Mol Biol* 2008; 432: 243–258.

[201] Naureckiene S.; Sleat DE.; Lackland H.; Fensom A.; Vanier MT.; Wattiaux R.; et al. Identification of HE1 as the Second Gene of Niemann-Pick C Disease. *Science (80-)* 2000; 290: 2298–2301.

[202] Sleat DE.; Zheng H.; Qian M.; Lobel P. Identification of Sites of Mannose 6-Phosphorylation on Lysosomal Proteins. *Mol Cell Proteomics* 2006; 5: 686–701.

[203] Journet A.; Chapel A.; Kieffer S.; Roux F.; Garin J. Proteomic analysis of human lysosomes: Application to monocytic and breast cancer cells. *Proteomics* 2002; 2: 1026–1040.

[204] Leighton F.; Poole B.; Beaufay H.; Baudhuin P.; Coffey JW.; Fowler S.; et al. The large-scale separation of peroxisomes, mitochondria, and lysosomes from the livers of rats injected with triton WR-1339. Improved isolation procedures, automated analysis, biochemical and morphological properties of fractions. *J Cell Biol* 1968; 37: 482–513.

[205] Chapel A.; Kieffer-Jaquinod S.; Sagne C.; Verdon Q.; Ivaldi C.; Mellal M.; et al. An Extended Proteome Map of the Lysosomal Membrane Reveals Novel Potential Transporters. *Mol Cell Proteomics* 2013; 12: 1572–1588.

[206] Della Valle MC.; Sleat DE.; Zheng H.; Moore DF.; Jadot M.; Lobel P. Classification of subcellular location by comparative proteomic analysis of native and density-shifted lysosomes. *Mol Cell Proteomics* 2011; 10: M110.006403.

[207] Vines DJ.; Warburton MJ. Classical late infantile neuronal ceroid lipofuscinosis fibroblasts are deficient in lysosomal tripeptidyl peptidase I. *FEBS Lett* 1999; 443: 131–135.

[208] Cotman SL.; Karaa A.; Staropoli JF.; Sims KB. Neuronal ceroid lipofuscinosis: Impact of recent genetic advances and expansion of the clinicopathologic spectrum topical collection on genetics. *Curr Neurol Neurosci Rep* 2013; 13: 1–17.

[209] Kollmann K.; Mutenda KE.; Balleininger M.; Eckermann E.; Von Figura K.; Schmidt B.; et al. Identification of novel lysosomal matrix proteins by proteome analysis. *Proteomics* 2005; 5: 3966–3978.

[210] Tharkeshwar AK.; Trekker J.; Vermeire W.; Pauwels J.; Sannerud R.; Priestman DA.; et al. A novel approach to analyze lysosomal dysfunctions through subcellular proteomics and lipidomics: the case of NPC1 deficiency. *Sci Rep* 2017; 7: 41408.

[211] Garver W.; Heidenreich R. The Niemann-Pick C Proteins and Trafficking of Cholesterol Through the Late Endosomal / Lysosomal System. *Curr Mol Med* 2002; 2: 485–505.

[212] Gao Y.; Chen Y.; Zhan S.; Zhang W.; Xiong F.; Ge W. Comprehensive proteome analysis of lysosomes reveals the diverse function of macrophages in immune responses. *Oncotarget* 2017; 8: 7420–7440.

[213] Wang CW. Lipid droplets, lipophagy, and beyond. *Biochim Biophys Acta - Mol Cell Biol Lipids* 2016; 1861: 793–805.

[214] Singh R.; Kaushik S.; Wang Y.; Xiang Y.; Novak I.; Komatsu M.; et al. Autophagy regulates lipid metabolism. *Nature* 2009; 458: 1131–5.

[215] Filomeni G.; Zio D De.; Cecconi F.; De Zio D.; Cecconi F. Oxidative stress and autophagy: the clash between damage and metabolic needs. *Cell Death Differ* 2015; 22: 377–388.

[216] Shui W.; Sheu L.; Liu J.; Smart B.; Petzold CJ.; Hsieh T-Y.; et al. Membrane proteomics of phagosomes suggests a connection to autophagy. *Proc Natl Acad Sci* 2008; 105: 16952–7.

[217] Davis S.; Wang J.; Zhu M.; Stahmer K.; Lakshminarayan R.; Ghassemian M.; et al. Sec24 phosphorylation regulates autophagosome abundance during nutrient deprivation. *Elife* 2016; 5: e21167.

[218] Ratajczak J.; Wysoczynski M.; Hayek F.; Janowska-Wieczorek A.; Ratajczak MZ. Membrane-derived microvesicles: important and underappreciated mediators of cell-to-cell communication. *Leukemia* 2006; 20: 1487–1495.

[219] Becker A.; Thakur BK.; Weiss JM.; Kim HS.; Peinado HH.; Lyden D. Extracellular Vesicles in Cancer: Cell-to-Cell Mediators of Metastasis. *Cancer Cell* 2016; 30: 836–848.

[220] Ko J.; Carpenter E.; Issadore D. Detection and isolation of circulating exosomes and microvesicles for cancer monitoring and diagnostics using micro-/nano-based devices. *Analyst* 2016; 141: 450–460.

[221] Deretic V.; Jiang S.; Dupont N. Autophagy intersections with conventional and

unconventional secretion in tissue development, remodeling and inflammation. *Trends Cell Biol* 2012; 22: 397–406.

[222] Karagiannis GS.; Pavlou MP.; Diamandis EP. Cancer secretomics reveal pathophysiological pathways in cancer molecular oncology. *Mol Oncol* 2010; 4: 496–510.

[223] Lodish H.; Berk A.; Zipursky SL.; Matsudaira P.; Baltimore D.; Darnell J. Translocation of Secretory Proteins across the ER Membrane. In: *Molecular Cell Biology*. New York: W. H. Freeman, 2000.

[224] Lodish H.; Berk A.; Zipursky SL.; Matsudaira P.; Baltimore D.; Darnell J. Overview of the Secretory Pathway. In: *Molecular Cell Biology*. New York: W. H. Freeman, 2000.

[225] Manjithaya R.; Subramani S. Autophagy: A broad role in unconventional protein secretion? *Trends Cell Biol* 2011; 21: 67–73.

[226] Zhang Q-Y.; Wu L-Q.; Zhang T.; Han Y-F.; Lin X. Autophagy-mediated HMGB1 release promotes gastric cancer cell survival via RAGE activation of extracellular signal-regulated kinases 1/2. *Oncol Rep* 2015; 1630–1638.

[227] Ohman T.; Teirila L.; Lahesmaa-Korpinen A-M.; Cypryk W.; Veckman V.; Saijo S.; et al. Dectin-1 Pathway Activates Robust Autophagy-Dependent Unconventional Protein Secretion in Human Macrophages. *J Immunol* 2014; 192: 5952–5962.

[228] Duran JM.; Anjard C.; Stefan C.; Loomis WF.; Malhotra V. Unconventional secretion of Acb1 is mediated by autophagosomes. *J Cell Biol* 2010; 188: 527–

536.

[229] Tooze SA.; Yoshimori T. The origin of the autophagosomal membrane. *Nat Cell Biol* 2010; 12: 831–835.

[230] Kang G-Y.; Bang JY.; Choi AJ.; Yoon J.; Lee W-C.; Choi S.; et al. Exosomal proteins in the aqueous humor as novel biomarkers in patients with neovascular age-related macular degeneration. *J Proteome Res* 2014; 13: 581–95.

[231] Brown GD.; Gordon S. Immune recognition. A new receptor for beta-glucans. *Nature* 2001; 413: 36–37.

[232] Apel A.; Herr I.; Schwarz H.; Rodemann HP.; Mayer A. Blocked autophagy sensitizes resistant carcinoma cells to radiation therapy. *Cancer Res* 2008; 68: 1485–1494.

[233] Chakradeo S.; Sharma K.; Alhaddad A.; Bakhshwin D.; Le N.; Harada H.; et al. Yet another function of p53 - The switch that determines whether radiation-induced autophagy will be cytoprotective or nonprotective: implications for autophagy inhibition as a therapeutic strategy. *Mol Pharmacol* 2015; 87: 803–814.

[234] Bristol ML.; Emery SM.; Maycotte P.; Thorburn A.; Chakradeo S.; Gewirtz DA. Autophagy inhibition for chemosensitization and radiosensitization in cancer: do the preclinical data support this therapeutic strategy? *J Pharmacol Exp Ther* 2013; 344: 544–52.

[235] Liu F.; Liu D.; Yang Y.; Zhao S. Effect of autophagy inhibition on chemotherapy-induced apoptosis in A549 lung cancer cells. *Oncol Lett* 2013; 5:

1261–1265.

[236] Thorburn A.; Thamm DH.; Gustafson DL. Autophagy and Cancer Therapy. *Mol Pharmacol* 2014; 85: 830–838.

[237] Sui X.; Chen R.; Wang Z.; Huang Z.; Kong N.; Zhang M.; et al. Autophagy and chemotherapy resistance: a promising therapeutic target for cancer treatment. *Cell Death Dis* 2013; 4: e838.

[238] White E. The role for autophagy in cancer. *J Clin Invest* 2015; 125: 42–46.

[239] Gewirtz DA. The Challenge of Developing Autophagy Inhibition as a Therapeutic Strategy. *Cancer Res* 2016; 76: 5610–5614.

[240] Lewis PD.; Parry JM. In silico p53 mutation hotspots in lung cancer. *Carcinogenesis* 2004; 25: 1099–1107.

[241] Ahrendt SA.; Hu Y.; Buta M.; Mcdermott MP.; Benoit N.; Yang SC.; et al. p53 Mutations and Survival in Stage I Non-Small-Cell Lung Cancer : Results of a Prospective Study. 2003; 95: 961–970.

[242] Mogi A.; Kuwano H. TP53 mutations in nonsmall cell lung cancer. *J Biomed Biotechnol* 2011; 2011: 1–9.

[243] World Health Organization. Cancerhttp://www.who.int/mediacentre/factsheets/fs297/en/ (2014, accessed 8 October 2014).

[244] American Cancer Society. *Cancer Facts & Figures 2014*. Atlanta, 2014.

[245] National Cancer Institute. Lung and Bronchus Cancer. *SEER Stat Fact*

*Sheets*http://seer.cancer.gov/statfacts/html/lungb.html (2014).

[246] Zappa C.; Mousa SA. Non-small cell lung cancer: current treatment and future advances. *Transl Lung Cancer Res* 2016; 5: 288–300.

[247] Alberg AJ.; Brock M V.; Ford JG.; Samet JM.; Spivack SD. Epidemiology of lung cancer: Diagnosis and management of lung cancer, 3rd ed: American college of chest physicians evidence-based clinical practice guidelines. *Chest* 2013; 143: e1S–e29S.

[248] Gumireddy K.; Li A.; Chang DH.; Liu Q.; Kossenkov A V.; Yan J.; et al. AKAP4 is a circulating biomarker for non-small cell lung cancer. *Oncotarget*; 6.

[249] Pérez-Soler R. Individualized therapy in non-small-cell lung cancer: future versus current clinical practice. *Oncogene* 2009; 28: S38–S45.

[250] Herbst RS.; Morgensztern D.; Boshoff C. Review The biology and management of non-small cell lung cancer. *Nat Publ Gr* 2018; 553: 446–454.

[251] Gazdar AF.; Bunn PA.; Minna JD. Small-cell lung cancer: What we know, what we need to know and the path forward. *Nat Rev Cancer* 2017; 17: 725–737.

[252] American Cancer Society. What Are Lung Carcinoid Tumors?https://www.cancer.org/cancer/lung-carcinoid-tumor/about/what-is-lung-carcinoid-tumor.html (2016, accessed 25 January 2018).

[253] Mary C Mancini. Carcinoid Lung Tumors: Background, Anatomy, Pathophysiology. *Medscape*https://emedicine.medscape.com/article/426400-overview (2017, accessed 25 January 2018).

[254] Zeman KG.; Brzezniak CE.; Carter CA. Recalcitrant small cell lung cancer: The argument for optimism. *J Thorac Dis* 2017; 9: E295–E296.

[255] Fasano M.; Maria C.; Corte D.; Papaccio F. Pulmonary Large-Cell Neuroendocrine Carcinoma From Epidemiology to Therapy. *J Thorac Oncol* 2015; 10: 1133–1141.

[256] Koshy M.; Goloubeva O.; Suntharalingam M. Impact of neoadjuvant radiation on survival in stage III non-small-cell lung cancer. *Int J Radiat Oncol Biol Phys* 2011; 79: 1388–1394.

[257] Non-small Cell Lung Cancer Collaborative Group. Chemotherapy in non-small celi lung cancer: a meta-analysis usmg updated data on individual patients from 52 randomised clinical trials. *BMJ* 1995; 311: 899–909.

[258] Masters GA.; Temin S.; Azzoli CG.; Giaccone G.; Baker S.; Brahmer JR.; et al. Systemic therapy for stage IV non-small-cell lung cancer: American society of clinical oncology clinical practice guideline update. *J Clin Oncol* 2015; 33: 3488–3515.

[259] American Cancer Society. Treatment Choices for Non-Small Cell Lung Cancer, by Stage. *cancer.org*https://www.cancer.org/cancer/non-small-cell-lung-cancer/treating/by-stage.html (2018, accessed 28 January 2018).

[260] Schreiber RD.; Old LJ.; Smyth MJ. Cancer Immunoediting : Integrating Suppression and Promotion. *Science (80-)* 2011; 331: 1565–1570.

[261] Brahmer J.; Reckamp KL.; Baas P.; Crinò L.; Eberhardt WEE.; Poddubskaya E.;

et al. Nivolumab versus Docetaxel in Advanced Squamous-Cell Non–Small-Cell Lung Cancer. *N Engl J Med* 2015; 373: 123–135.

[262] National Cancer Institute. Radiation Therapy for Cancerhttps://www.cancer.gov/about-cancer/treatment/types/radiation-therapy/radiation-fact-sheet (2010, accessed 28 January 2018).

[263] Borrego-Soto G.; Ortiz-López R.; Rojas-Martínez A. Ionizing radiation-induced DNA injury and damage detection in patients with breast cancer. *Genet Mol Biol* 2015; 38: 420–432.

[264] Lomax ME.; Folkes LK.; O'Neill P. Biological consequences of radiation-induced DNA damage: Relevance to radiotherapy. *Clin Oncol* 2013; 25: 578–585.

[265] Huisman C.; Giaccone G.; Smit EF.; Postmus PE. Second-line chemotherapy in relapsing or refractory non-small-cell lung cancer: a review. *J Clin Oncol* 2000; 18: 3722–3730.

[266] Hanahan D.; Weinberg RA. Hallmarks of cancer: The next generation. *Cell* 2011; 144: 646–674.

[267] Karagounis I V.; Kalamida D.; Mitrakas A.; Pouliliou S.; Liousia M V.; Giatromanolaki A.; et al. Repression of the autophagic response sensitises lung cancer cells to radiation and chemotherapy. *Br J Cancer* 2016; 115: 312–321.

[268] Lee JG.; Shin JH.; Shim HS.; Lee CY.; Kim DJ.; Kim YS. Autophagy contributes to the chemo- resistance of non-small cell lung cancer in hypoxic conditions. *Respir Res* 2015; 1–9.

[269] Dexter DL.; Leith JT. Tumor heterogeneity and drug resistance. *J Clin Oncol* 1986; 4: 244–57.

[270] Lee JG.; Shin JH.; Shim HS.; Lee CY.; Kim DJ.; Kim YS.; et al. Autophagy contributes to the chemo-resistance of non-small cell lung cancer in hypoxic conditions. *Respir Res* 2015; 16: 138.

[271] Fulda S.; Gorman AM.; Hori O.; Samali A. Cellular Stress Responses: Cell Survival and Cell Death. *Int J Cell Biol* 2010; 2010: 1–23.

[272] Kerr JF.; Wyllie AH.; Currie AR. Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. *Br J Cancer* 1972; 26: 239–57.

[273] Meyn RE. Apoptosis and response to radiation: implications for radiation therapy. *Oncology* 1997; 11: 349–356.

[274] Su Z.; Yang Z.; Xu Y.; Chen Y.; Yu Q. Apoptosis, autophagy, necroptosis, and cancer metastasis. *Mol Cancer* 2015; 14: 48.

[275] Chaabane W.; User SD.; El-Gazzah M.; Jaksik R.; Sajjadi E.; Rzeszowska-Wolny J.; et al. Autophagy, Apoptosis, Mitoptosis and Necrosis: Interdependence Between Those Pathways and Effects on Cancer. *Arch Immunol Ther Exp (Warsz)* 2013; 61: 43–58.

[276] Elmore S. Apoptosis: a review of programmed cell death. *Toxicol Pathol* 2007; 35: 495–516.

[277] Los M.; Wesselborg S.; Schulze-Osthoff K. The role of caspases in development, immunity, and apoptotic signal transduction: lessons from knockout mice.

Immunity 1999; 10: 629–39.

[278] Los M.; Mozoluk M.; Ferrari D.; Stepczynska A.; Stroh C.; Renz A.; et al. Activation and caspase-mediated inhibition of PARP: a molecular switch between fibroblast necrosis and apoptosis in death receptor signaling. *Mol Biol Cell* 2002; 13: 978–88.

[279] Lips J.; Kaina B. DNA double-strand breaks trigger apoptosis in p53-deficient fibroblasts. *Carcinogenesis* 2001; 22: 579–585.

[280] Bhosle S.; Huilgol N.; Mishra K. Programmed cell death as a prognostic indicator for radiation therapy in cervical carcinoma patients: A pilot study. *J Cancer Res Ther* 2005; 1: 41–45.

[281] Rödel C.; Grabenbauer GG.; Papadopoulos T.; Bigalke M.; Günther K.; Schick C.; et al. Apoptosis as a cellular predictor for histopathologic response to neoadjuvant radiochemotherapy in patients with rectal cancer. *Int J Radiat Oncol* 2002; 52: 294–303.

[282] Milisav I. Cellular Stress Responses. In: Wislet-Gendebien S (ed) *Advances in Regenerative Medicine*. InTech, pp. 215–232.

[283] Gewirtz DA. Autophagy and senescence: A partnership in search of definition. *Autophagy* 2013; 9: 808–812.

[284] Starokadomskyy P.; Dmytruk K V. A bird’s-eye view of autophagy. *Autophagy* 2013; 9: 1121–1126.

[285] Mathew R.; Karantza-Wadsworth V.; White E. Role of autophagy in cancer. *Nat*

Rev Cancer 2010; 7: 961–967.

[286] Young ARJ.; Narita M.; Ferreira M.; Kirschner K.; Sadaie M.; Darot JFJ.; et al. Autophagy mediates the mitotic senescence transition. *Genes Dev* 2009; 23: 798–803.

[287] Hayflick L.; Moorhead PS. The serial cultivation of human diploid cell strains. *Exp Cell Res* 1961; 25: 585–621.

[288] Chakradeo S.; Elmore LW.; Gewirtz DA. Is Senescence Reversible? *Curr Drug Targets*; 17.

[289] Kuilman T.; Michaloglou C.; Mooi WJ.; Peeper DS. The essence of senescence. *Genes Dev* 2010; 24: 2463–2479.

[290] Sikora E.; Mosieniak G.; Sliwinska MA. Morphological and functional characteristic of senescent cancer cells. *Curr Drug Targets*http://www.ncbi.nlm.nih.gov/pubmed/26477465 (2015, accessed 2 November 2015).

[291] Fagagna F d'Adda di.; Reaper PM.; Clay-Farrace L.; Fiegler H.; Carr P.; Von Zglinicki T.; et al. A DNA damage checkpoint response in telomere-initiated senescence. *Nature* 2003; 426: 194–8.

[292] Sharpless NE.; Ramsey MR.; Balasubramanian P.; Castrillon DH.; DePinho RA. The differential impact of p16(INK4a) or p19(ARF) deficiency on cell growth and tumorigenesis. *Oncogene* 2004; 23: 379–85.

[293] Goehe RW.; Di X.; Sharma K.; Bristol ML.; Henderson SC.; Valerie K.; et al. The

autophagy-senescence connection in chemotherapy: must tumor cells (self) eat before they sleep? *J Pharmacol Exp Ther* 2012; 343: 763–78.

[294] Knizhnik A V.; Roos WP.; Nikolova T.; Quiros S.; Tomaszowski KH.; Christmann M.; et al. Survival and Death Strategies in Glioma Cells: Autophagy, Senescence and Apoptosis Triggered by a Single Type of Temozolomide-Induced DNA Damage. *PLoS One* 2013; 8: 1–12.

[295] Singh K.; Matsuyama S.; Drazba JA.; Almasan A. Autophagy-dependent senescence in response to DNA damage and chronic apoptotic stress. *Autophagy* 2012; 8: 236–51.

[296] Kang HT.; Lee KB.; Kim SY.; Choi HR.; Park SC. Autophagy impairment induces premature senescence in primary human fibroblasts. *PLoS One* 2011; 6: 1–12.

[297] Yang D.; Li L.; Liu H.; Wu L.; Luo Z.; Li H.; et al. Induction of autophagy and senescence by knockdown of ROC1 E3 ubiquitin ligase to suppress the growth of liver cancer cells. *Cell Death Differ* 2013; 20: 235–47.

[298] White E. Autophagy and p53. *Cold Spring Harb Perspect Med* 2016; 6: 1–10.

[299] Sui X.; Han W.; Pan H. P53-Induced Autophagy and Senescence. *Oncotarget* 2015; 6: 11723–11724.

[300] Yu X.; Harris SL.; Levine AJ. The regulation of exosome secretion: A novel function of the p53 protein. *Cancer Res* 2006; 66: 4795–4801.

[301] Chenau J.; Michelland S.; de Fraipont F.; Josserand V.; Coll J.; Favrot M-C.; et al. The Cell Line Secretome, a Suitable Tool for Investigating Proteins Released in

Vivo by Tumors: Application to the Study of p53-Modulated Proteins Secreted in Lung Cancer Cells. *J Proteome Res* 2009; 8: 4579–4591.

[302] Hu R.; Huffman KE.; Chu M.; Zhang Y.; Minna JD.; Yu Y. Quantitative Secretomic Analysis Identifies Extracellular Protein Factors That Modulate the Metastatic Phenotype of Non-Small Cell Lung Cancer. *J Proteome Res* 2016; 15: 477–486.

[303] Habener JF.; Rich A.; Jr JTP.; Kemper B.; Habener JF.; Rich A.; et al. Parathyroid Secretion: Discovery of a Major Calcium-Dependent Protein. *Science (80-)* 1974; 184: 167–169.

[304] Dupont N.; Jiang S.; Pilli M.; Ornatowski W.; Bhattacharya D.; Deretic V. Autophagy-based unconventional secretory pathway for extracellular delivery of IL-1β. *EMBO J* 2011; 30: 4701–11.

[305] Ko A.; Kanehisa A.; Martins I.; Senovilla L.; Chargari C.; Dugue D.; et al. Autophagy inhibition radiosensitizes in vitro, yet reduces radioresponses in vivo due to deficient immunogenic signaling. Cell Death Differ 2014. (11): Comment. *Cell Death Differ* 2014; 21: 92–99.

[306] Maecker HL.; Koumenis C.; Giaccia AJ. p53 promotes selection for Fas-mediated apoptotic resistance. *Cancer Res* 2000; 60: 4638–4644.

[307] Wiśniewski JR.; Zougman A.; Nagaraj N.; Mann M. Universal sample preparation method for proteome analysis. *Nat Methods* 2009; 6: 359–62.

[308] Pathan M.; Keerthikumar S.; Ang CS.; Gangoda L.; Quek CYJ.; Williamson NA.; et al. FunRich: An open access standalone functional enrichment and interaction

network analysis tool. *Proteomics* 2015; 15: 2597–2601.

[309] Yang J.; Everett AD. Hepatoma-derived growth factor binds DNA through the N-terminal PWWP domain. *BMC Mol Biol* 2007; 8: 101.

[310] Nakamura H.; Izumoto Y.; Kambe H.; Kuroda T.; Mori T.; Kawamura K.; et al. Molecular cloning of complementary DNA for a novel human hepatoma-derived growth factor: Its homology with high mobility group-1 protein. *J Biol Chem* 1994; 269: 25143–25149.

[311] Vaca Jacome AS.; Rabilloud T.; Schaeffer-reiss C.; Rompais M.; Ayoub D.; Lane L.; et al. N-terminome analysis of the human mitochondrial proteome. *Proteomics* 2015; 2519–2524.

[312] Honorй B.; Rasmussen HH.; Celis A.; Leffers H.; Madsen P.; Celis JE. The molecular chaperones HSP28, GRP78, endoplasmin, and calnexin exhibit strikingly different levels in quiescent keratinocytes as compared to their proliferating normal and transformed counterparts: cDNA cloning and expression of calnexin. *Electrophoresis* 1994; 15: 482–490.

[313] Gyorffy B.; Surowiak P.; Budczies J.; Lanczky A. Online Survival Analysis Software to Assess the Prognostic Value of Biomarkers Using Transcriptomic Data in Non-Small-Cell Lung Cancer. *PLoS One* 2013; 8: 1–8.

[314] Nickel W. The mystery of nonclassical protein secretion: A current view on cargo proteins and potential export routes. *Eur J Biochem* 2003; 270: 2109–2119.

[315] Cordani M.; Pacchiana R.; Butera G.; D'Orazi G.; Scarpa A.; Donadelli M. Mutant

p53 proteins alter cancer cell secretome and tumour microenvironment: Involvement in cancer invasion and metastasis. *Cancer Lett* 2016; 376: 303–309.

[316] Eichelbaum K.; Winter M.; Diaz MB.; Herzig S.; Krijgsveld J. Selective enrichment of newly synthesized proteins for quantitative secretome analysis. *Nat Biotechnol* 2012; 30: 984–990.

[317] Surinova S.; Schiess R.; Hüttenhain R.; Cerciello F.; Wollscheid B.; Aebersold RR.; et al. On the Development of Plasma Protein Biomarkers. *J Proteome Res* 2011; 10: 5–16.

[318] Chapman JD.; Goodlett DR.; Masselon CD. Multiplexed and Data-Independent Tandem Mass Spectrometry for Global Proteome Profiling. *Mass Spectrom Rev* 2013; 33: 452–470.

[319] Bhutia SK.; Kegelman TP.; Das SK.; Azab B.; Su Z-Z.; Lee S-G.; et al. Astrocyte elevated gene-1 induces protective autophagy. *Proc Natl Acad Sci* 2010; 107: 22243–22248.

[320] Gong F.; Peng X.; Sang Y.; Qiu M.; Luo C.; He Z.; et al. Dichloroacetate induces protective autophagy in LoVo cells: involvement of cathepsin D/thioredoxin-like protein 1 and Akt-mTOR-mediated signaling. *Cell Death Dis* 2013; 4: e913.

[321] Giatromanolaki A.; Kalamida D.; Sivridis E.; Karagounis I V.; Gatter KC.; Harris AL.; et al. Increased expression of transcription factor EB (TFEB) is associated with autophagy, migratory phenotype and poor prognosis in non-small cell lung cancer. *Lung Cancer* 2015; 90: 98–105.

[322] Jiang Y.; Woosley AN.; Sivalingam N.; Natarajan S.; Howe PH. Cathepsin-B-mediated cleavage of Disabled-2 regulates TGF- β -induced autophagy. *Nat Cell Biol*; 2. Epub ahead of print 2016. DOI: 10.1038/ncb3388.

[323] Taupenot L.; Harper KL.; O'Connor DT. The chromogranin–secretogranin family. *N Engl J Med* 2003; 348: 1134–1149.

[324] Kirchmair R.; Gander R.; Egger M.; Hanley A.; Silver M.; Ritsch A.; et al. The Neuropeptide Secretoneurin Acts as a Direct Angiogenic Cytokine In Vitro and In Vivo. *Circulation* 2004; 109: 777–783.

[325] Shyu WC.; Lin SZ.; Chiang MF.; Chen DC.; Su CY.; Wang HJ.; et al. Secretoneurin promotes neuroprotection and neuronal plasticity via the Jak2/Stat3 pathway in murine models of stroke. *J Clin Invest* 2008; 118: 133–148.

[326] Ischia R.; Gasser RW.; Fischer-Colbrie R.; Eder U.; Pagani A.; Cubeddu LX.; et al. Levels and molecular properties of secretoneurin-immunoreactivity in the serum and urine of control and neuroendocrine tumor patients. *J Clin Endocrinol Metab* 2000; 85: 355–360.

[327] Yasuda D.; Iguchi H.; Funakoshi A.; Wakasugi H.; Sekiya K.; Misawa T.; et al. Comparison of Plasma Pancreastatin and GAWK Concentrations, Presumed Processing Products of Chromogranin A and B, in Plasma of Patients with Pancreatic Islet Cell Tumors. *Horm Metab Res* 1993; 25: 593–595.

[328] Schleusener JT.; Tazelaar HD.; Jung SH.; Cha SS.; Cera PJ.; Myers JL.; et al. Neuroendocrine differentiation is an independent prognostic factor in chemotherapy-treated nonsmall cell lung carcinoma. *Cancer* 1996; 77: 1284–

1291.

[329] Carles J.; Rosell R.; Ariza A.; Pellicer I.; Sanchez JJ.; Fernandez-Vasalo G.; et al. Neuroendocrine differentiation as a prognostic factor in non-small cell lung cancer. *Lung Cancer* 1993; 10: 209–19.

[330] Renal C.; Carcinoma C.; Lucarelli G.; Rutigliano M.; Sanguedolce F.; Galleggiante V.; et al. Increased expression of the autocrine motility factor is associated with poor prognosis in patients with clear cell-renal cell carcinoma. *Med (United States)* 2015; 94: e2117.

[331] Li Y.; Li Y.; Che Q.; Che Q.; Bian Y.; Bian Y.; et al. Autocrine motility factor promotes epithelial-mesenchymal transition in endometrial cancer via MAPK signaling pathway. *Int J Oncol* 2015; 47: 1017–1024.

[332] Hofman ER.; Boyanapalli M.; Lindner DJ.; Weihua X.; Hassel BA.; Jagus R.; et al. Thioredoxin Reductase Mediates Cell Death Effects of the Combination of Beta Interferon and Retinoic Acid. *Mol Cell Biol* 1998; 18: 6493–6504.

[333] Damdimopoulos AE.; Miranda-Vizuete A.; Treuter E.; Gustafsson J-A.; Spyrou G. An Alternative Splicing Variant of the Selenoprotein Thioredoxin Reductase Is a Modulator of Estrogen Signaling. *J Biol Chem* 2004; 279: 38721–38729.

[334] Cadenas C.; Franckenstein D.; Schmidt M.; Gehrmann M.; Hermes M.; Geppert B.; et al. Role of thioredoxin reductase 1 and thioredoxin interacting protein in prognosis of breast cancer. *Breast Cancer Res* 2010; 12: R44.

[335] Zhu Y-H.; Zhang B.; Li M.; Huang P.; Sun J.; Fu J.; et al. Prognostic significance

of FAM3C in esophageal squamous cell carcinoma. *Diagn Pathol* 2015; 10: 192.

[336] Gao Z.; Lu C.; Wang Z.; Song Y.; Zhu J.; Gao P.; et al. ILEI: A novel marker for Epithelial-Mesenchymal Transition and poor prognosis in colorectal cancer. *Histopathology* 2014; 65: 527–38.

[337] Ling-Zhi Wang, Ross A. Soo, Win Lwin Thuya, Ting Ting Wang, Tiannan Guo, Jieying Amelia Lau, Fang Cheng Wong, Andrea Li Ann Wong, Soo Chin Lee, Siu Kwan Sze BCG. Exosomal protein FAM3C as a potential novel biomarker for non-small cell lung cancer. In: *American Society of Clinical Oncology Annual Meeting*. American Society of Clinical Oncologyhttp://meetinglibrary.asco.org/content/132756-144 (2014).

[338] Song Q.; Sheng W.; Zhang X.; Jiao S.; Li F. ILEI drives epithelial to mesenchymal transition and metastatic progression in the lung cancer cell line A549. *Tumor Biol* 2014; 35: 1377–1382.

[339] Tjoelker LW.; Seyfried CE.; Eddy RL.; Byers MG.; Shows TB.; Calderon J.; et al. Human, mouse, and rat calnexin cDNA cloning: identification of potential calcium binding motifs and gene localization to human chromosome 5. *Biochemistry* 1994; 33: 3229–36.

[340] Núñez A.; Dulude D.; Jbel M.; Rokeach LA. Calnexin is essential for survival under nitrogen starvation and stationary phase in Schizosaccharomyces pombe. *PLoS One* 2015; 10: 1–27.

[341] Kobayashi M.; Nagashio R.; Jiang SX.; Saito K.; Tsuchiya B.; Ryuge S.; et al. Calnexin is a novel sero-diagnostic marker for lung cancer. *Lung Cancer* 2015;

90: 342–345.

[342] Quidville V.; Alsafadi S.; Goubar A.; Commo F.; Scott V.; Pioche-Durieu C.; et al. Targeting the deregulated spliceosome core machinery in cancer cells triggers mTOR blockade and autophagy. *Cancer Res* 2013; 73: 2247–2258.

[343] Novotny I.; Podolska K.; Blazikova M.; Valasek LS.; Svoboda P.; Stanek D. Nuclear LSm8 affects number of cytoplasmic processing bodies via controlling cellular distribution of Like-Sm proteins. *Mol Biol Cell* 2012; 23: 3776–3785.

[344] Jensen LJ.; Kuhn M.; Stark M.; Chaffron S.; Creevey C.; Muller J.; et al. STRING 8 - A global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res* 2009; 37: 412–416.

[345] STRING CONSORTIUM. LSM8 STRING Analysis.

[346] Clement PMJ.; Johansson HE.; Wolff EC.; Park MH. Differential expression of eIF5A-1 and eIF5A-2 in human cancer cells. *FEBS J* 2006; 273: 1102–1114.

[347] Klier H.; Csonga R.; Joao HC.; Eckerskorn C.; Auer M.; Lottspeich F.; et al. Isolation and structural characterization of different isoforms of the hypusine-containing protein eIF-5A from HeLa cells. *Biochemistry* 1995; 34: 14693–14702.

[348] Li AL.; Li HY.; Jin BF.; Ye QN.; Zhou T.; Yu XD.; et al. A novel eIF5A complex functions as a regulator of p53 and p53-dependent apoptosis. *J Biol Chem* 2004; 279: 49251–49258.

[349] Patel PH.; Costa-Mattioli M.; Schulze KL.; Bellen HJ. The Drosophila deoxyhypusine hydroxylase homologue nero and its target eIF5A are required for

cell growth and the regulation of autophagy. *J Cell Biol* 2009; 185: 1181–1194.

[350] Xu G.; Shi X.; Sun L.; Zhou Q.; Zheng D.; Shi H.; et al. Down-regulation of eIF5A-2 prevents epithelial-mesenchymal transition in non-small-cell lung cancer cells. *J Zhejiang Univ Sci B* 2013; 14: 460–467.

[351] He LR.; Zhao HY.; Li BK.; Liu YH.; Liu MZ.; Guan XY.; et al. Overexpression of eIF5A-2 is an adverse prognostic marker of survival in stage I non-small cell lung cancer patients. *Int J Cancer* 2011; 129: 143–150.

[352] Fujimura K.; Choi S.; Wyse M.; Strnadel J.; Wright T.; Klemke R. Eukaryotic translation initiation factor 5A (EIF5A) regulates pancreatic cancer metastasis by modulating RhoA and rho-associated kinase (ROCK) protein expression levels. *J Biol Chem* 2015; 290: 29907–29919.

[353] Grundner-culemann K.; Dybowski JN.; Klammer M.; Tebbe A.; Schaab C.; Daub H. Comparative proteome analysis across non-small cell lung cancer cell lines. *J Proteomics* 2015; 130: 1–10.

[354] Polanski M.; Anderson NL. A List of Candidate Cancer Biomarkers for Targeted Proteomics. *Biomark Insights* 2006; 2: 1–48.

[355] Knowles B.; Howe C.; Aden D. Human hepatocellular carcinoma cell lines secrete the major plasma proteins and hepatitis B surface antigen. *Science (80-)*; 209http://science.sciencemag.org/content/209/4455/497 (1980, accessed 4 August 2017).

[356] Kohler JJ.; Patrie SM. *Mass Spectrometry of Glycoproteins: Methods and*

Protocols. Humana Press, 2013.

[357] Krasnova L.; Wong C-H. Understanding the Chemistry and Biology of Glycosylation with Glycan Synthesis. *Annu Rev Biochem* 2016; 85: 599–630.

[358] Stowell SR.; Ju T.; Cummings RD. Protein glycosylation in cancer. *Annu Rev Pathol* 2015; 10: 473–510.

[359] Moremen KW.; Tiemeyer M.; Nairn A V. Vertebrate protein glycosylation: diversity, synthesis and function. *Nat Rev Mol Cell Biol* 2012; 13: 448–462.

[360] Defaus S.; Gupta P.; Andreu D.; Gutiérrez-Gallego R. Mammalian protein glycosylation - structure versus function. *Analyst* 2014; 139: 2944–67.

[361] Stastna M.; Eyk JE Van. Investigating the Secretome: Lessons About the Cells that Comprise the Heart. *Circ Cardiovasc Genet* 2012; 5: 1–19.

[362] Huang DW.; Sherman BT.; Lempicki R a. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009; 4: 44–57.

[363] Huang DW.; Sherman BT.; Lempicki RA. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 2009; 37: 1–13.

[364] Yamashita R.; Fujiwara Y.; Ikari K.; Hamada K.; Otomo A.; Yasuda K.; et al. Extracellular proteome of human hepatoma cell, HepG2 analyzed using two-dimensional liquid chromatography coupled with tandem mass spectrometry. *Mol Cell Biochem* 2007; 298: 83–92.

[365] Nanjappa V.; Thomas JK.; Marimuthu A.; Muthusamy B.; Radhakrishnan A.; Sharma R.; et al. Plasma Proteome Database as a resource for proteomics research: 2014 update. *Nucleic Acids Res* 2014; 42: D959–D965.

[366] Kaji H.; Saito H.; Yamauchi Y.; Shinkawa T.; Taoka M.; Hirabayashi J.; et al. Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nat Biotechnol* 2003; 21: 667–672.

[367] Yoshida M.; Stadler J.; Bertholdt G.; Gerisch G. Wheat germ agglutinin binds to the contact site A glycoprotein of Dictyostelium discoideum and inhibits EDTA-stable cell adhesion. *EMBO J* 1984; 3: 2663–70.

[368] Li Y.; Shah P.; Marzo AM De.; Eyk JE Van.; Li Q.; Daniel W. Chan.; et al. Identification of Glycoproteins Containing Specific Glycans Using a Lectin-Chemical Method. *Anal Chem* 2015; 87: 4683–4687.

[369] Johnson PJ.; Poon TC.; Hjelm NM.; Ho CS.; Ho SK.; Welby C.; et al. Glycan composition of serum alpha-fetoprotein in patients with hepatocellular carcinoma and non-seminomatous germ cell tumour. *Br J Cancer* 1999; 81: 1188–95.

[370] Kim H.; Kim K.; Jin J.; Park J.; Yu SJ.; Yoon JH.; et al. Measurement of glycosylated alpha-fetoprotein improves diagnostic power over the native form in hepatocellular carcinoma. *PLoS One*; 9. Epub ahead of print 2014. DOI: 10.1371/journal.pone.0110366.

[371] Ajdukiewicz AB.; Kelleher PC.; Krawitt EL.; Walters CJ.; Mason PB.; Koff RS.; et al. Alpha-fetoprotein glycosylation is abnormal in some hepatocellular carcinoma, including white patients with a normal alpha-fetoprotein concentration. *Cancer*

Lett 1993; 74: 43–50.

[372] Peterson CB.; Blackburn MN. Isolation and characterization of an antithrombin III variant with reduced carbohydrate content and enhanced heparin binding. *J Biol Chem* 1985; 260: 610–615.

[373] Fu D.; van Halbeek H. N-glycosylation site mapping of human serotransferrin by serial lectin affinity chromatography, fast atom bombardment-mass spectrometry, and 1H nuclear magnetic resonance spectroscopy. *Anal Biochem* 1992; 206: 53–63.

[374] Hülsmeier AJ.; Paesold-Burda P.; Hennet T. N-Glycosylation Site Occupancy in Serum Glycoproteins Using Multiple Reaction Monitoring Liquid Chromatography-Mass Spectrometry. *Mol Cell Proteomics* 2007; 6: 2132–2138.

[375] Büll C.; Stoel MA.; Den Brok MH.; Adema GJ. Sialic acids sweeten a tumor's life. *Cancer Res* 2014; 74: 3199–3204.

[376] Pearce OMT.; Läubli H. Sialic acids in cancer biology and immunity. *Glycobiology* 2015; 26: 111–128.

[377] Hawkridge AM.; Muddiman DC. Mass spectrometry-based biomarker discovery: toward a global proteome index of individuality. *Annu Rev Anal Chem* 2009; 2: 265–77.

[378] Sung NS.; Crowley WF.; Genel M.; Salber P.; Sandy L.; Sherwood LM.; et al. Central Challenges Facing the National Clinical Research Enterprise. *J Am Med Assoc* 2003; 289: 1278–1287.

[379] Butler D. Translational research: crossing the valley of death. *Nature* 2008; 453: 840–842.

[380] Anderson NL.; Anderson NG. The Human Plasma Proteome. *Mol Cell Proteomics* 2002; 1: 845–867.

[381] Ortea I.; Rodríguez-Ariza A.; Chicano-Gálvez E.; Arenas Vacas MS.; Jurado Gámez B. Discovery of potential protein biomarkers of lung adenocarcinoma in bronchoalveolar lavage fluid by SWATH MS data-independent acquisition and targeted data extraction. *J Proteomics* 2016; 138: 106–114.

[382] Planque C.; Kulasingam V.; Smith CR.; Reckamp K.; Goodglick L.; Diamandis EP. Identification of five candidate lung cancer biomarkers by proteomics analysis of conditioned media of four lung cancer cell lines. *Mol Cell Proteomics* 2009; 8: 2746–2758.

[383] Lewis JA.; Dennis WE.; Hadix J.; Jackson DA. Analysis of secreted proteins as an in vitro model for discovery of liver toxicity markers. *J Proteome Res* 2010; 9: 5794–5802.

Vita

Emmanuel Kenneth Cudjoe Jr is a Ghanaian citizen born on September 4, 1986 in Shama, Ghana. He attended and graduated from Mfantsipim School, Ghana in 2004 where he received his high school diploma in General Science. Emmanuel had his undergraduate education at the Kwame Nkrumah University of Science and Technology, Kumasi, Ghana where he received his Bachelor of Pharmacy degree. He then pursued a Master's degree in Pharmaceutical Chemistry in the same university from 2010 and 2013 during which time his interest in bioanalysis was kindled leading him to apply and pursue a PhD. Following his graduation from the master's program, he joined the PhD program in the School of Pharmacy in the department of Pharmacotherapy & Outcomes Science in August 2013 under the advisorship of Dr. Adam M. Hawkridge.