Mass spectrometry analysis of the secretome in malignant lung adenocarcinoma transformation

by

Jennifer K. Luu

B.Sc., The University of British Columbia, 2017

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Pathology and Laboratory Medicine)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August 2022

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, a thesis entitled:

Mass spectrometry analysis of the secretome in malignant lung adenocarcinoma transformation

submitted by Jennifer K. Luu in partial fulfillment of the requirements for

the degree of Master of Science

in Pathology and Laboratory Medicine

Examining Committee:

William Lockwood, Associate Professor, Pathology and Laboratory Medicine, UBC
Supervisor

Kevin Bennewith, Associate Professor, Pathology and Laboratory Medicine, UBC
Supervisory Committee Member

Cathie Garnis, Associate Professor, Surgery, UBC
Additional Examiner

Wan Lam, Professor, Pathology and Laboratory Medicine, UBC
Additional Examiner

Additional Supervisory Committee Members:

Ninan Abraham, Professor, Microbiology and Immunology, UBC
Supervisory Committee Member

Abstract

Lung cancer and its most common subtype, lung adenocarcinoma (LUAD), is the leading cause of global cancer mortality. Mortality is partially attributed to late stage diagnosis, where curative options are limited and less effective for patient survival. Current screening guidelines are prone to false positive results, and there are no lung cancer-specific biomarkers to aid detection. This is especially relevant for never smokers, a lung cancer patient demographic often characterized by the *EGFR* oncogenic mutation. Never smokers lack concrete screening guidelines; this makes biomarker discovery and validation crucial for earlier detection.

The collective set of secreted proteins derived from a cell, known as the secretome, has been explored as a potential source of cancer biomarkers. However, LUAD secretome studies have been limited to late stage or metastatic tumors; investigation of secretome changes during malignant transformation may thus be more suitable for biomarker discovery.

This thesis investigated changes in the secretome in an *EGFR*-driven model of LUAD malignant transformation. In this study, I generated a stepwise model of *EGFR*-driven malignant transformation by generating stable lung epithelial cell lines and selecting *EGFR* mutant NSCLC cell lines. I also developed and optimized a mass spectrometry protocol to profile the secretome of my model. With this pipeline, I identified differentially expressed proteins in advanced LUAD. I then validated these findings by performing differential gene expression (DGE) analysis on an *EGFR* mutant LUAD patient cohort. Finally, validated secreted proteins were assessed for effects on overall patient survival resulting in 4 *EGFR*-specific and 1 non-specific biomarker candidates. This work provides insight into potential secretome changes during

LUAD malignant transformation, biomarker candidates for further validation, and illustrates the potential of the secretome in *EGFR* mutant LUAD detection.

Lay Summary

Lung adenocarcinoma (LUAD) is a subtype of lung cancer that is often detected during late stages of progression, where patient survival is more limited. By studying changes during LUAD progression, especially when a cell transforms into a cancer cell, this may identify molecular markers to screen for early LUAD. Cells release proteins into their environment as part of routine biological processes. Analyzing changes in proteins outside of a cell as it becomes a cancer cell may then provide candidates for LUAD screening. In this thesis, I investigated the changes in released proteins between normal lung cell lines and LUAD cell lines using a technique called “mass spectrometry”. My findings identified 4 proteins that are differentially expressed in cell lines and in clinical patient data, specific to LUAD with an *EGFR* gene mutation, and 1non-specific protein. These findings may help complement existing or future screening methods for LUAD.

Preface

The work in this thesis was done under the guidance and mentorship of Dr. William W. Lockwood. Conditioned media collection for the optimized secretome experiment in Chapter 3 was done with the assistance of Tianna Sihota and Rocky Shi. Mass spectrometry sample processing and labeling were performed by Dr. Fraser Johnson. Mass spectrometry sample collection optimization and data-dependent acquisition experiments were performed with assistance from Dr. Sandra Spencer-Miko. Mass spectrometry data processing, normalization, and principal component analysis (including visualization) were performed by Dr. Gian Negri. All other experiments and data analysis were performed by me.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AAH	Atypical adenomatous hyperplasia
AIS	Adenocarcinoma *in situ*
BPE	Bovine pituitary extract
CCLE	Cancer Cell Line Encyclopedia
CM	Conditioned media
COPD	Chronic obstructive pulmonary disease
CT	C-terminal
DAPI	4′,6-diamidino-2-phenylindole
DGE	Differential gene expression
DN	Dominant negative
DPE	Differential protein expression
DTT	Dithiothreitol
DUSP6	Dual specificity phosphatase 6
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
ER	Endoplasmic reticulum
ERK	Extracellular signal-regulated kinase
EV	Extracellular vesicle
FBS	Fetal bovine serum
FDR	False discovery rate
FGF	Fibroblast growth factor 2

GEMM	Genetically engineered mouse model
GEO	Gene Expression Omnibus
GFP	Green fluorescent protein
GO	Gene Ontology
HBEC	Human bronchial epithelial cell
IAA	Iodoacetamide
IL-1β	Interleukin 1 beta
KRAS	Kirsten rat sarcoma virus
KSFM	Keratinocyte serum-free media
LDCT	Low-dose computed tomography
LFC	Log_2 fold change
LT	Large T antigen
LUAD	Lung adenocarcinoma
MAPK	Mitogen-activated protein kinase
MEK	Mitogen-activated protein kinase kinase
MIA	Minimally invasive adenocarcinoma
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
NLST	National Lung Screening Trial
NSCLC	Non-small cell lung cancer
P-	Phospho-
PCA	Principal component analysis

PI	Propidium iodide
PSM	Peptide-spectrum match
RAS	Rat sarcoma virus
RFP	Red fluorescent protein
RTK	Receptor tyrosine kinase
SCC	Squamous cell carcinoma
SCLC	Small cell lung cancer
SP	Signal peptide
STAGE	STop And Go Extraction
SV40	Simian virus 40
TFA	trifluoroacetic acid
TKD	Tyrosine kinase domain
TKI	Tyrosine kinase inhibitor
TMT	Tandem mass tag
TP53	Tumor protein p53
WT	Wild type

Acknowledgements

This thesis was made possible by the mentorship and support of my supervisor, Dr. William Lockwood. Dr. Lockwood was a guiding force in navigating the complexity of my work by providing sound scientific advice and encouragement after dead-end experiments (of which there were many). I owe my growth as a researcher to his mentorship. I am also incredibly grateful to my lab members, both past and present, including Dr. Fraser Johnson, Daniel Lu, Tianna Sihota, Rocky Shi, and Dylan Farnsworth for their willingness to be a captive audience for experimental troubleshooting, assistance, and camaraderie during the long days, nights, and weekends spent at the lab. Research is not easy, but having others to celebrate successes and commiserate failure with makes the process more manageable.

I am also grateful to my committee members, Drs. Marcel Bally, Kevin Bennewith, and Ninan Abraham for their expertise and constructive criticism during this work.

I would also like to extend my gratitude to the Canadian Institutes of Health Research for making this research possible.

Finally, I am eternally grateful to my family and friends for their support during my studies. There are too many names to write here, but their patience, kindness, and endless love made the last 4 years less of a panicked fever dream and more of an enriching journey. I hope to pay it forward in the future!

This thesis is dedicated to my parents, an immigrant and a refugee.

Chapter 1: Introduction

The purpose of this thesis is to develop an experimental pipeline to identify biomarker candidates for the detection of lung adenocarcinoma, a subtype of lung cancer. This pipeline was developed using an *in vitro* cell line model of lung adenocarcinoma transformation and a mass spectrometry protocol optimized to detect secreted proteins. To provide context to this thesis, I will provide an overview of the following topics: lung cancer; detection methods of lung cancer; lung cancer classification; malignant transformation, specifically in lung adenocarcinoma; and the secretome.

1.1 Overview of lung cancer

Lung cancer is the leading cause of global cancer mortality in men and women, resulting in an estimated 1.8 million deaths in 2020 (Sung *et al.*, 2021). In Canada, it is the most commonly diagnosed cancer and cause of cancer mortality; an estimated 25% of cancer deaths are attributed to this disease (Canadian Cancer Statistics Advisory Committee *et al.*, 2021). Lung cancer is often detected in the advanced or metastatic stages when potentially curative surgery is no longer an option, leading to a reliance on common therapeutic options such as chemotherapy and radiotherapy which are less effective for long-term patient survival (Herbst *et al.*, 2018; Lemjabbar-Alaoui *et al.*, 2015). Due to the direct relationship between clinical staging and prognosis, there is a strong incentive for earlier lung cancer detection to improve long-term patient survival (Lemjabbar-Alaoui *et al.*, 2015). Tobacco smoking accounts for the majority of lung cancer cases, however there are other known risk factors. These include biomass combustion, air pollution, occupational exposure to carcinogens such as arsenic, and environmental sources of radon (Bade & Dela Cruz, 2020). Genetic predisposition via one's family history of cancer and chronic pulmonary diseases such as chronic obstructive pulmonary

disease (COPD) are also risk factors in an individual's susceptibility to lung cancer (Bade & Dela Cruz, 2020).

1.2 Lung cancer detection methods for lung cancer

Detection of lung cancer can be divided into conventional and emerging techniques. Conventional techniques consist of radiographic imaging, sputum cytology, bronchoscopy, and tissue biopsy; emerging techniques include biosensors, biomarkers, radiogenomics, artificial intelligence-based methods, and liquid biopsies (Nooreldeen & Bach, 2021; Prabhakar *et al.*, 2018). As lung cancer normally presents in advanced stages, detection is typically based upon cytology and biopsy samples (Nicholson *et al.*, 2022). Earlier detection of lung cancer has been explored by screening high risk individuals prior to symptom development; techniques that have been applied include sputum cytology, chest radiography, and computed tomography (Manser *et al.*, 2003).

Of the conventional techniques described, the standard screening method for lung cancer is low-dose computed tomography (LDCT). LDCT is a radiographic technique used to image abnormal growths in the lung, characterized by its reduced scanning time, decreased radiation exposure, and increased diagnostic accuracy relative to conventional radiography (Nooreldeen & Bach, 2021). The effectiveness of LDCT screening was highlighted in the National Lung Screening Trial (NLST) conducted in 2002-2009. In this study, annual LDCT screening was found to decrease lung cancer mortality by 20% compared to conventional chest radiography in high risk individuals, defined as 55-74 year old individuals who had either smoked a minimum of 30 pack years or previous smokers who had ceased smoking within 15 years (Gierada *et al.*, 2020; Tammemagi & Lam, 2014). However, LDCT demonstrates a lack of specificity, with 96% of positive NSLT LDCT scans ultimately being false positives (The National Lung Screening

Trial Research Team, 2011). Additionally, this study did not include never smokers, defined as individuals who have been exposed to less than 100 cigarettes in their lifetime; currently, the effectiveness of LDCT screening for never smokers is unclear (Lam, 2019; Sun *et al.*, 2007). Over diagnosis of benign pulmonary nodules and the potential harm of repeated radiation exposure over time are additional limitations (Tammemagi & Lam, 2014). Currently, LDCT is paired with other diagnostic techniques such as sputum cytology or needle biopsy to detect and diagnose lung cancer (Prabhakar *et al.*, 2018). Biomarker inclusion within LDCT scanning framework has been proposed as a way to address these issues. In particular, biomarkers present in blood are particularly desirable, as blood collection is inexpensive and less invasive than assessing lung tissue of bronchoalveolar lavage (BAL) (Sozzi & Boeri, 2014). Current blood serum biomarker candidates are neuron-specific enolase (NSE), progastrin-releasing peptide (proGRP), carcinoembryonic antigen (CEA), squamous cell carcinoma antigen (SCCA), and CYRFA 21-1. However, none are specific to lung cancer and do not distinguish between lung cancer histologies or molecular subtypes, which are important considerations for disease management strategies (Duffy & O'Byrne, 2018).

1.3 Lung cancer classification

Lung cancer is divided into two major subtypes, small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). These subtypes comprise 15% and 85% of lung cancer cases, respectively (Herbst *et al.*, 2008). Each subtype can be further classified into pathologic subtypes based on tumor histological and molecular characteristics in order to tailor therapeutic strategies (Travis *et al.*, 2015; Zheng, 2016).

1.3.1 Small cell lung cancer

SCLC is a high-grade neuroendocrine tumor that occurs almost exclusively among current or previous smokers (Rudin *et al.*, 2021). SCLC presents as a tumor consisting of small cells that have a round to spindle shape, limited cytoplasm and nucleoli, and granulated nuclear chromatin (Travis, 2012). To identify SCLC, hematoxylin and eosin staining is commonly performed, with supplemental histological markers CD56, chromogranin, and synaptophysin for unclear cases (Rudin *et al.*, 2021). The most common genetic alterations observed in SCLC are the frequent loss of function of tumor suppressor genes *TP53* and *RB1* (Semenova *et al.*, 2015). Therapeutic options vary by tumor stage, yet as patients often present with advanced disease, the combination of chemotherapies cisplatin and etoposide is often the treatment standard (Rudin *et al.*, 2021; Semenova *et al.*, 2015).

1.3.2 Non-small cell lung cancer

NSCLC is a highly diverse subtype of lung cancer, reflected in the developing molecular classifications as research on genetic drivers and targeted therapies increases (Nicholson *et al.*, 2022). The most common histological subtypes of NSCLC are lung adenocarcinoma (LUAD) and squamous cell carcinoma (SSC), comprising 60% and 15% of NSCLC cases. Other histological subtypes, such as large cell carcinoma and mixed histology tumors, are less common (Arbour & Riely, 2019). NSCLC can arise in either smokers or never smokers; however, there is a stronger association between smoking and SCC (Dela Cruz *et al.*, 2011). In comparison, LUAD is more frequently observed in never smokers; this may be attributed to factors such as second-hand smoke, pollution, and genetic predisposition (Dela Cruz *et al.*, 2011; Herbst *et al.*, 2018).

The diversity of NSCLC is most reflected in LUAD, which comprises nearly half of all lung cancer cases (Travis *et al.*, 2011). Key features of LUAD tumors are the presence of glandular differentiation or mucin production (Zheng, 2016). LUAD histology varies and commonly presents as five histological subtypes: lepidic, acinar, papillary, micropapillary, and solid (Travis *et al.*, 2011, 2015). LUAD can also be identified via histological staining for thyroid transcription factor-1(TTF-1) and Napsin A (Zheng, 2016). *KRAS* and *EGFR* oncogenic mutations are frequently observed in LUAD as are mutations in tumor suppressor genes *TP53*, *KEAP1*, *STK11*, and *NF1* (Herbst *et al.*, 2018). Treatment options for NSCLC, and by extension LUAD, often depend on variables such as tumor stage, histology, patient clinical status, and tumor molecular profile (Alexander *et al.*, 2020). Surgical resection is often recommended for early stage tumors (stages I-II) and select advanced stage (stage IIIA) tumors. However, approximately 30% of patients present with locally advanced disease (stage IIIA-C) where most tumors are unresectable. In these cases, concurrent chemoradiotherapy and subsequent immunotherapy are recommended (Alexander *et al.*, 2020; Herbst *et al.*, 2018). Patients with advanced stage disease are also suitable candidates for molecular targeted therapy, often the first-line treatment when specific tumor genetic alterations have been identified (Gridelli *et al.*, 2015).

1.3.2.1 Oncogenic mutations in lung adenocarcinoma

Sequencing studies of LUAD patient subsets, specifically those where there are tumor dependencies on distinct genes, have identified key mutations in LUAD that form the basis of targeted therapies (Hirsch *et al.*, 2016; Lamberti *et al.*, 2020). The most common oncogenic driver mutations are *KRAS* and *EGFR* in early and metastatic stage LUAD (Skoulidis & Heymach, 2019)(Figure 1.1). These genes are involved in driving mitogen-activated protein kinase (MAPK) and other signaling pathways that promote cancer development and progression

(Roberts & Der, 2007). Up to 30% of NSCLC cases, and by extension LUAD, have a *KRAS* mutation. *EGFR* mutation incidence varies by patient demographic, with 50% frequency in Asian patients and 15% in Western patients (Chevallier *et al.*, 2021). Other oncogenic drivers in LUAD include mutations in *BRAF, MET,* and *ALK* (Chevallier *et al.*, 2021; Lamberti *et al.*, 2020). There is additional diversity within defined oncogene subgroups, seen in co-occurring genetic alterations. An example of this includes recurrent mutations in genes including *TP53, RB1,* and *PIK3CA* in tandem with *EGFR* oncogenic mutations in advanced LUAD (Skoulidis & Heymach, 2019).

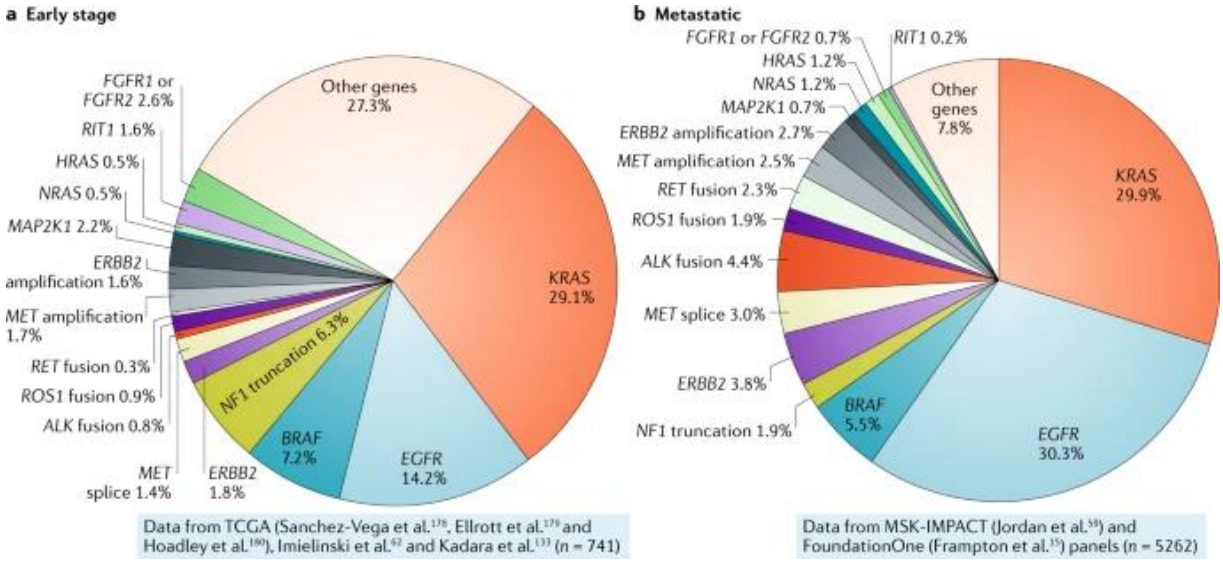


Figure 1.1 Distribution of driver oncogenes observed in early stage and metastatic lung adenocarcinoma. Reprinted by permission from Springer Nature: Springer Nature, Nature Reviews Cancer, Co-occurring genomic alterations in non-small-cell lung cancer biology and therapy (Skoulidis & Heymach, 2019). Data used to generate this reprinted figure are acknowledged here. Reprinted from Cell, Volume 150, Issue 6, Imielinski *et al.*, Mapping the Hallmarks of Lung Adenocarcinoma with Massively Parallel Sequencing, 1107-1220, 2012, with permission from Elsevier (Imielinski *et al.*, 2012). Reprinted from Annals of Oncology, Volume 28, Issue 8, Kadara *et al.*, Whole-exome sequencing and immune profiling of early-stage lung adenocarcinoma with fully annotated clinical follow-up, 75-82.,2017, with permission from Elsevier (Kadara *et al.*, 2017). Reprinted from Cancer Discovery, 2015, Volume 5, Issue 8, 850-859, Frampton *et al.*, Activation of MET via Diverse Exon 14 Splicing Alterations Occurs in Multiple Tumor Types and Confers Clinical Sensitivity to MET Inhibitors, with permission from AACR(Frampton *et al.*, 2015). Reprinted from Cancer Discovery, 2017, Volume 7, Issue 6, 596-609, Jordan *et al.*, Prospective Comprehensive Molecular Characterization of Lung Adenocarcinomas for Efficient Patient Matching to Approved and Emerging Therapies, with permission from AACR (Jordan *et al.*, 2017).Work not directly cited in this caption are acknowledged through the Creative Commons license.

1.4 Malignant transformation in lung adenocarcinoma

The process under which cells progress from histologically non-malignant into a malignant tumor is known as malignant transformation. LUAD is understood to be a process of multiple genetic and epigenetic changes that promote the progression of lung epithelial cells into tumor cells (Larsen & Minna, 2011). This is observed in a clinicopathological fashion, where cells progress through a series of stages from atypical adenomatous hyperplasia (AAH), to adenocarcinoma *in situ* (AIS), minimally invasive adenocarcinoma (MIA), and finally invasive adenocarcinoma (Inamura, 2018; Travis *et al.*, 2015)(Figure 1.2). During AAH, the precursor to LUAD, multiple molecular changes are observed; mutations in genes such as *EGFR* and *KRAS* and epigenetic repression and low expression of *CDKN2A* are detected. As the histology progresses towards invasive adenocarcinoma, alterations are associated with tumor invasion such as increased *TP53* mutation frequency (Inamura, 2018). Transforming cells will acquire traits that promote tumor formation, known as the hallmarks of cancer (Hanahan & Weinberg, 2011). Of the 10 commonly understood traits, these include the ability to sustain proliferative signaling and evade growth suppression (Hanahan, 2022).

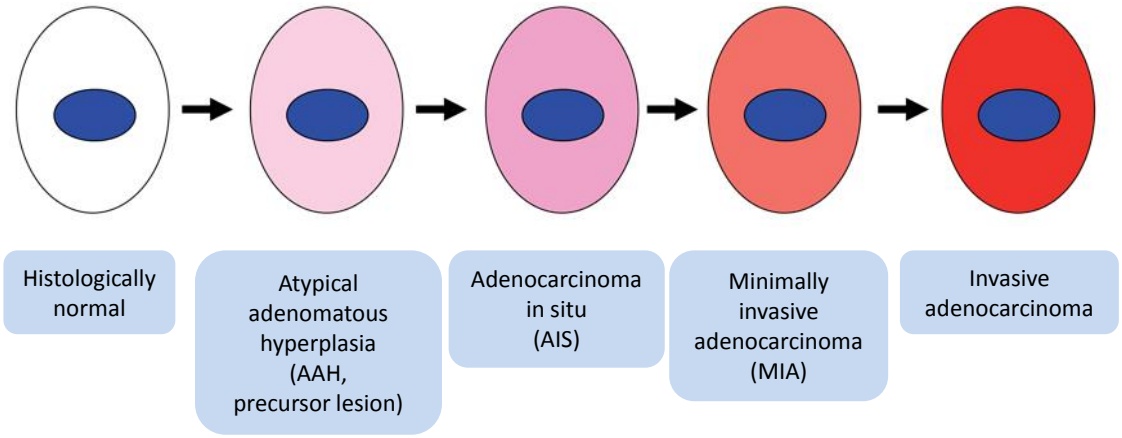


Figure 1.2 Stepwise model of lung adenocarcinoma transformation.
Adapted from Cancer Prevention Research, 2008, Volume 1, Issue 3, 156-160, Gazdar and Minna, Deregulated EGFR signaling during lung cancer progression: mutations, amplicons, and autocrine loops, with permission from AACR (Gazdar & Minna, 2008). Stages are derived from 2015 WHO guidelines (Travis *et al.*, 2015).

1.4.1 EGFR pathway signaling and malignant transformation

MAPK signaling cascades are key regulators of cellular processes such as cell proliferation, survival, and differentiation (Roberts & Der, 2007). The RAS-RAF-MEK-ERK pathway is one MAPK cascade, mainly activated by growth factor-stimulating receptor tyrosine kinase (RTK) proteins that include EGFR. Through RTK activation, proteins are sequentially phosphorylated until the terminal MAPK ERK1/2 is phosphorylated, which then leads to transcription factor activation and downstream cellular processes (Katz *et al.*, 2007).

EGFR is an initiator of the RAS-RAF-MEK-ERK signaling cascade and an early oncogenic mutation of lung adenocarcinoma (Skoulidis & Heymach, 2019). General EGFR activation occurs via the binding of specific ligands to its extracellular region, normally EGF (Lemmon *et al.*, 2014). The ligands induce receptor dimerization, and conformational changes that enable phosphorylation of dimerized tyrosine kinase domains (TKD). The phosphorylated TKD results in an active EGFR, which triggers the subsequent phosphorylation of the RAS-RAF-MEK-ERK pathway (Du & Lovly, 2018).

EGFR mutations have been frequently observed in LUAD. Mutations often occur in exons 18-21, where the TKD is located (exons 18-24) (Du & Lovly, 2018; Siegelin & Borczuk, 2014)(Figure 1.3). The most common mutations are an in-frame deletion in exon 19 or a point mutation in exon 21, L858R, that result in constitutive EGFR activation regardless of ligand binding (Siegelin & Borczuk, 2014). This can promote excessive cell growth, proliferation, and dependence on dysregulated signalling for cell survival, known as "oncogene addiction" (Larsen & Minna, 2011). As excessive proliferative signalling is a known cancer hallmark, deregulated EGFR signalling would be expected to promote malignant transformation (Hanahan & Weinberg, 2011). In fact, EGFR mutations were observed in non-invasive and invasive tumor lesions that included AAH, highlighting its role in tumor initiation (Soh *et al.*, 2008).

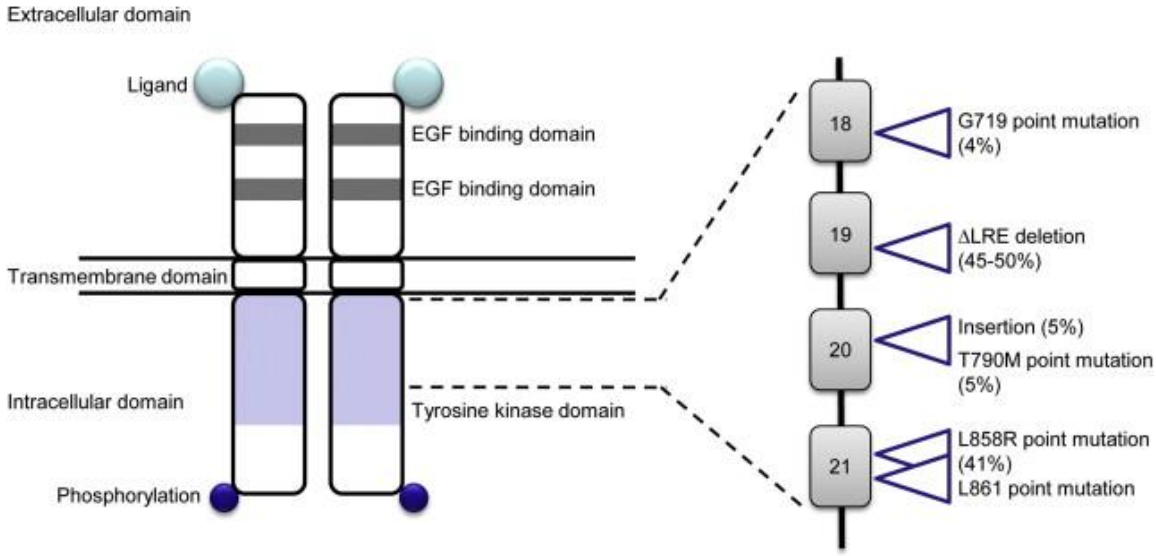


Figure 1.3 Structure of EGFR and mutations observed in lung cancer.
Reprinted from Clinics in Chest Medicine, Volume 32, Issue 4, Larsen and Minna, Molecular biology of lung cancer: clinical implications, 703-740, 2011, with permission from Elsevier (Larsen & Minna, 2011).

1.4.2 Models of malignant transformation

Multiple studies have attempted to model stepwise changes that are thought to occur during cellular transformation to lung adenocarcinoma. While there are no primary or immortalized cultures derived from lesions such as AAH, well-characterized methods to study

cellular transformation include genetically engineered mouse models (GEMMs) and immortalized cell lines *in vitro* (Hynds *et al.*, 2021). The widespread use of cell lines has led to the generation of immortalized "normal" lung epithelial cell lines initially with either simian virus 40 (SV40) or large T antigen (LT), then non-viral oncoprotein methods (Sato *et al.*, 2020). These cell lines have been used as a background to introduce mutations in oncogenes or tumor suppressor genes to identify the number and types of genetic alterations necessary for cellular transformation (Hynds *et al.*, 2021; Sato *et al.*, 2020).

The most commonly used cell lines to study cellular transformation are BEAS-2B, immortalized with SV40, and human bronchial epithelial cells (HBEC), immortalized with non-viral proteins hTERT and CDK4; both cell lines are derived from bronchial epithelial cells (Ramirez *et al.*, 2004; Reddel *et al.*, 1988). Due to SV40 immortalization, BEAS-2B has been observed to undergo spontaneous transformation *in vivo* after approximately 30 passages (Reddel *et al.*, 1993). This is attributed to LT in SV40 targeting proteins which include tumor suppressors p53 and Rb family members, but makes BEAS-2B less suitable for experimentally controlled models of transformation (Ahuja *et al.*, 2005). Recently, a study analyzing BEAS-2B epithelial origin identified mesenchymal stem cell marker expression of CD73, CD90, CD105, and CD44, further suggesting that BEAS-2B may not be a suitable LUAD transformation model (Han *et al.*, 2020). In contrast, HBEC cells do not show a transformation phenotype either *in vitro* with anchorage-independent growth or *in vivo* with mouse tumor formation. HBEC cells also retain wild-type p53 and epithelial properties, such as expression of epithelial markers cytokeratins 7, 14, 17, and 19 (Ramirez *et al.*, 2004).

1.4.3 Current understanding of genetic alterations in malignant transformation *in vitro*

Immortalized lung epithelial cell lines have been important in understanding the genetic alterations necessary to promote transformation. Sato *et al.* demonstrated that the combination of *TP53* knockdown, *KRAS G12V* overexpression, and *c-MYC* overexpression were sufficient to fully transform HBECs and form tumors *in vivo*. They showed that a total of five genetic alterations in *hTERT*, *CDK4*, *TP53*, *KRAS*, and *MYC* were adequate for stepwise bronchial epithelial transformation (Sato *et al.*, 2013). Sasai *et al.* also demonstrated a similar combination of genetic alterations in a human small airway epithelial cell (HSAEC) line model; they noted that *hTERT* overexpression, *CDK4* overexpression, dominant negative expression of *TP53*, *KRAS G12V* expression, and expression of either an active form of *PIK3CA*, active *CCND1*, or dominant negative *LKB1* were sufficient to form tumors in nude mice (Sasai *et al.*, 2011). Collectively, these studies illustrate that five genetic alterations are needed to transform bronchial epithelial cells: *hTERT* overexpression, *TP53* and *p16/CDK4/Rb* pathway inactivation, and two further alterations (Sato *et al.*, 2020)(Figure 1.4).

Studies analyzing stepwise alterations in LUAD transformation have mainly focused on *KRAS*-driven transformation, with less attention on *EGFR*. Studies on *EGFR*-driven LUAD have modeled clinical observations that *TP53* mutations frequently occur in *EGFR*-driven LUAD (Skoulidis & Heymach, 2019). In an initial study of HBEC transformation, HBEC cells expressing either *EGFR L858R* or *EGFR E746_A750* (exon 19 deletion) formed anchorage-independent colonies, and colonies increased in combination with *TP53* knockdown for cells expressing *EGFR L858R*. However, the combination of *EGFR L858R* expression and *TP53* knockdown was insufficient to form tumors *in vivo*; this suggests additional alterations are needed for transformation (Sato *et al.*, 2006). More recently, a study on a *EGFR L858R*, *TP53*-

mutated LUAD GEMM identified that *APC*, *RB1*, and *RBM10* loss promoted tumor growth (Foggetti *et al.*, 2021). These results suggest further genetic alterations to be studied in generating an EGFR-driven model of transformation.

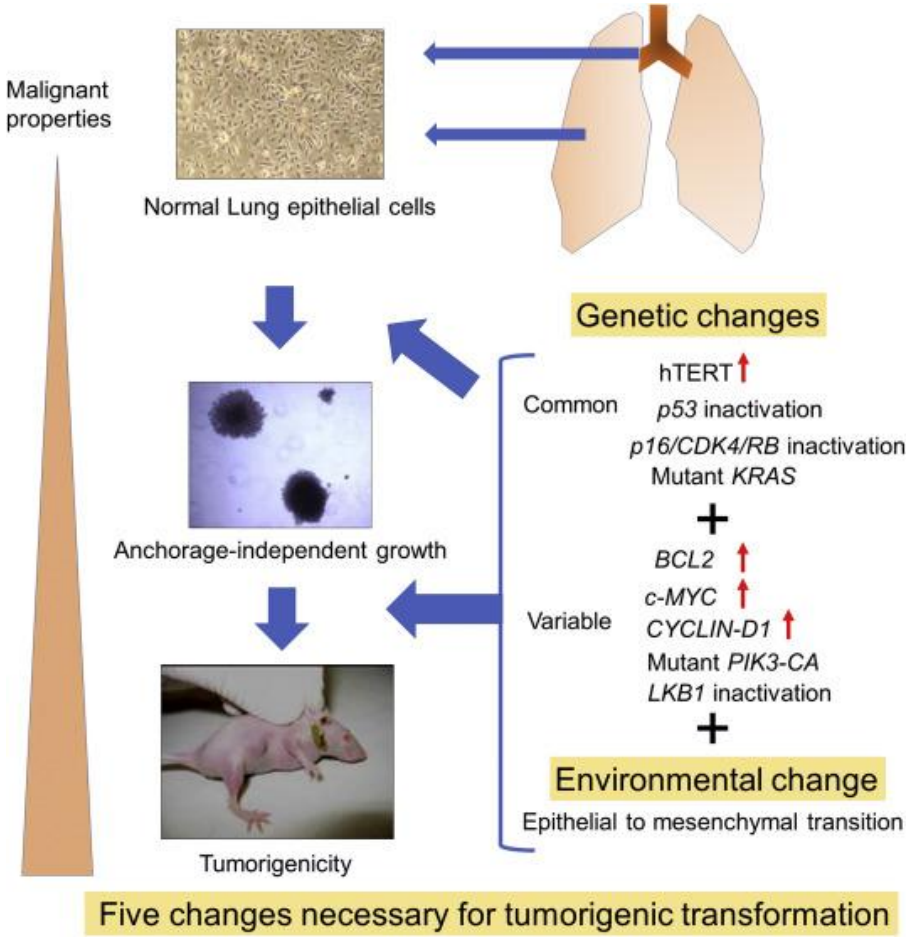


Figure 1.4 Current understanding of genetic alterations necessary for tumorigenic transformation in lung epithelial cells.
Reprinted from Respiratory Investigation, Volume 58, Issue 5, Sato *et al.*, Immortalized normal human lung epithelial cell models for studying lung cancer biology, 344-354, 2020, with permission from Elsevier (Sato *et al.*, 2020).

1.5 Overview of the secretome

The collective set of proteins that are transported from within a cell to the extracellular space is known as the "secretome", estimated to comprise 15% of all human protein-encoding genes (Uhlén *et al.*, 2019). The secretome plays an important role in local cell and systemic signalling; these proteins include growth factors, hormones, extracellular matrix-degrading

proteinases, immunomodulatory cytokines, and cell motility factors (Uhlén *et al.*, 2019; Xue *et al.*, 2008). Given the role of the secretome in many biological processes, dysregulated protein secretion can affect disease progression. This is observed in cancer, where tumor cells can secrete growth factors, angiogenic factors, and proteases during tumor progression (Brown *et al.*, 2013).

1.5.1 Secreted proteins and secretory pathways

Secreted proteins are defined proteins that are actively transported through secretory pathways beyond the cell cytoplasm (Uhlén *et al.*, 2019). There are currently two categories of protein secretion understood in the literature. The first is known as classical secretion, where secretion occurs via the cleavage of a signal peptide (SP) domain during translation and through the Golgi apparatus to the extracellular space (Cohen *et al.*, 2020). The SP domain is located at the protein N-terminus and generally has approximately 25-30 amino acid residues, although longer SPs of up to 140 residues have been observed (Owji *et al.*, 2018). The SP typically consists of three components: a positively charged amino terminal (n-region); a hydrophobic core (h-region); and a carboxyl cleavage site (c-region) (Owji *et al.*, 2018; von Heijne, 1990). During protein translation, the SP domain is recognized by the signal recognition particle and the ribosome-mRNA complex is guided to the endoplasmic reticulum; the SP domain is removed during this coupled process of translation and translocation (Cohen *et al.*, 2020; Steringer *et al.*, 2015). Protein folding then occurs in the ER lumen, and then the mature protein moves through the Golgi apparatus, and ultimately enters the extracellular space via vesicular transport (Cohen *et al.*, 2020). However, not all secreted proteins will possess a signal peptide; examples of this are interleukin 1beta (IL-1β) and fibroblast growth factor 2 (FGF2) (Steringer *et al.*, 2015).

The second category of protein secretion is known as unconventional secretion and consists of four types of secretion. These consist of Type I, Type III, and Type IV secretion; Type II occurs when there is acylated peptide or yeast mating peptide secretion and is relatively under characterized (Rabouille, 2017). These pathways can be further categorized into two subcategories, where proteins either lack an SP or transmembrane domain or contain an SP and/or transmembrane domain. The first subcategory, proteins lacking an SP or transmembrane domain, consists of direct protein translocation across the plasma membrane and is considered "leaderless". Type I, II, and III secretion constitute "leaderless" secretion. The second subcategory, proteins that contain an SP and/or a transmembrane domain, consists of proteins entering the ER without proceeding through the Golgi apparatus as in classical secretion; Type IV pathway secretion comprises this subcategory (Dimou & Nickel, 2018; Rabouille, 2017)(Figure 1.5).

The more well-characterized pathways are Types I, III, and IV secretion. While each pathway has distinct characteristics, they are often induced upon cellular stress. This is thought to be a response to impaired classical secretion under cellular stress (Rabouille, 2017). Type I secretion occurs through the formation of plasma membrane pores, experimentally observed in FGF2 secretion. During FGF2 secretion, FGF2 in its folded conformation is recruited to the plasma membrane, interacts with phosphoinositides at the plasma membrane inner leaflet, triggering FGF2 oligomerization, and ultimately the formation of a lipidic pore for translocation to the extracellular space (Dimou & Nickel, 2018; Steringer *et al.*, 2015).

Type III secretion occurs via membrane-bound organelles that are thought to include autophagosomes, lysosomes, endosomes, and exosomes. Of these organelles, autophagy-related vesicular structures are thought be important in Type III secretion; this has been shown with IL-

1β (J. Kim *et al.*, 2018). Recent studies have found that IL-1β secretion appears to occur through a vesicle intermediate, mediated by autophagy structures and the inflammosome (Dupont *et al.*, 2011; Zhang *et al.*, 2015). However, this may be context-dependent as IL-1β has been observed to be secreted with the Type I pathway when inflammation is the stimulus (Rabouille, 2017).

Type IV secretion occurs as a modified version of classical secretion, where the Golgi apparatus is avoided. While secreted proteins in Type IV secretion will have a signal peptide, translocation routes will vary (J. Kim *et al.*, 2018). A form of Type IV secretion occurs when a transmembrane protein, typically directed for the plasma membrane, is mutated. The presence of a point mutation can hinder full protein folding, resulting in excess unfolded protein and ER stress activation. ER stress activation then promotes exit of proteins from the ER to the plasma membrane (Rabouille, 2017). This is observed in a mutated form of pendrin, H723R-pendrin, a monovalent anion transporter (Jung *et al.*, 2016).

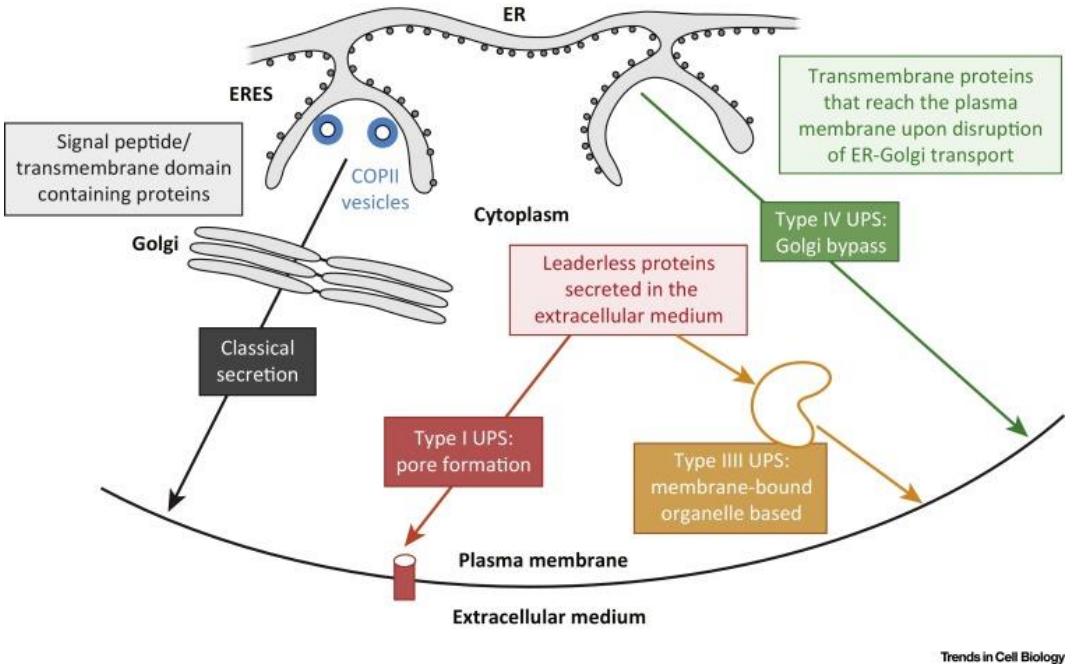


Figure 1.5 Schematic image of conventional and unconventional secretion pathways (excluding Type II). Reprinted from Trends in Cell Biology, Volume 27, Issue 3, Rabouille, Pathways of unconventional protein secretion, 230-240, 2017, with permission from Elsevier (Rabouille, 2017).

Another facet of unconventional protein secretion occurs through the production and release of extracellular vesicles (EVs). EVs are membrane bound vesicles generated via differing cell trafficking processes, where the most commonly known types are exosomes and ectosomes (Meldolesi, 2022; van Niel *et al.*, 2018). EV characteristics vary per type, but all consist of a lipid bilayer exterior and range between 50 nm to 2000 nm in diameter (Cheng & Hill, 2022). Exosomes are produced through the endocytic pathway; an intraluminal vesicle (ILV) is generated and released into the extracellular space via fusion of a larger multivesicular body containing multiple ILVs with the plasma membrane (Meldolesi, 2022; van Niel *et al.*, 2018). In contrast, ectosomes are produced from the outward budding of the plasma membrane into the extracellular space (van Niel *et al.*, 2018). While EV cargo varies per type, cargo can include unconventionally secreted proteins, cytosolic proteins, various RNA types including microRNAs and long non-coding RNAs, DNA sequences, lipids, and metabolic molecules (Meldolesi, 2022).

1.5.2 Applications of the secretome in cancer

The secretome is thought to be a valuable supply of biomarkers for cancer, as tumor secreted proteins can be detected in bodily fluids such as blood or urine (Xue *et al.*, 2008). Various studies have aimed to characterize secreted proteins to identify potential candidates. Such methods include using serial analysis of gene expression, DNA microarray, RNA sequencing, antibody and bead arrays, mass spectrometry, secretion traps, and bioinformatic prediction (Mukherjee & Mani, 2013). In practice, mass spectrometry is the predominant technique as it is unbiased, direct, and sensitive enough to detect dilute protein concentrations in the nanomolar range (Basisty *et al.*, 2020; Brandi *et al.*, 2018). A recent study by Robinson *et al.* used RNA-sequencing of 32 cancer types and 30 healthy tissues from biological databases, repositories for biological data generated from high throughput technologies (Zou *et al.*, 2015).

This was done to profile changes in the transcriptional level of genes encoding proteins that are predicted to be secreted. In 20 cancer types, they identified that matrix metalloproteinase (MMP) and collagen (COL) genes had elevated expression in tumor compared to non-tumor samples (Robinson *et al.*, 2019). These observations reflect the increased levels of MMP family members in various cancers, including lung cancer, and previous biomarker validation of certain collagens in cancer such as type IV collagen in pancreatic cancer (Izbicka *et al.*, 2012; Öhlund *et al.*, 2009). Robinson *et al.* also found that the vast changes in common tumor type secretome expression changes were in proteins involved in functions such as adhesion and immune response, as well as tumor suppressors involved in negative regulation of proliferation and migration (Robinson *et al.*, 2019). Identification of such proteins can not only generate diagnostic biomarkers but also prognostic biomarkers, as suggested with angiopoietin-like protein 1 (ANGPTL1) in hepatocellular carcinoma (Yan *et al.*, 2017).

1.5.3 Approaches to study the secretome in non-small cell lung cancer

Studies performed to characterize the cancer secretome often apply mass spectrometry (MS), where the model system may vary (Xue *et al.*, 2008). Often the cancer model consists of immortalized cell lines, due to their wide availability and ease of growth and manipulation; NSCLC studies often consist of cell lines derived from patient tumors (Brandi *et al.*, 2018; Hynds *et al.*, 2021). Cancer secretome studies, and by extension NSCLC secretome studies, then largely consist of studying the conditioned media (CM) of cells (Mukherjee & Mani, 2013). While cell lines cannot fully model the tumor microenvironment, such as stroma and paracrine signaling, they are easy to genetically alter and their CM can provide insight into processes *in vivo* (Brandi *et al.*, 2018; Méndez & Villanueva, 2015). Although blood appears to be a suitable alternative to cell line secretome research, it is a complex bodily fluid; it consists of various

proteins with a vast dynamic range of at minimum, 9-10 orders of magnitude (Xue *et al.*, 2008). Further, the blood proteome may be overrepresented by abundant proteins such as albumin, masking less abundant proteins that may be informative as potential biomarkers (Omenn *et al.*, 2005; Xue *et al.*, 2008). In contrast, cell line CM is relatively less complex; this may enable identification of less abundant secreted proteins (Brandi *et al.*, 2018). Therefore, cell lines serve as a suitable model to study the secretome for discovery of potential biomarkers. In NSCLC secretome studies, cell lines established from primary and metastatic tumors such as H2073, H1993, PC-9, and H1975 are often used. These studies have identified secreted proteins involved in metastasis, resistance to the EGFR tyrosine kinase inhibitor erlotinib, and cisplatin response prediction (Bosse *et al.*, 2016; Böttger *et al.*, 2019; Hu *et al.*, 2016).

A key challenge in cancer secretome research is the use of serum (Brandi *et al.*, 2018). Many cell lines require serum-containing media for regular culture conditions, usually fetal bovine serum (FBS), which interferes with secretome assessment. FBS proteins can mask low abundance proteins (in the range of ng/mL) such that thorough proteomic investigation of the secretome is challenging (Shin *et al.*, 2019). To avoid this issue, cell line culture media is often changed to serum-free media to decrease secretome complexity, improve overall analysis, and minimize non-human contamination (Hu *et al.*, 2016; Shin *et al.*, 2019). However, serum starvation can increase cell cytolysis and subsequent contamination of CM by intracellular proteins (Brandi *et al.*, 2018). Protocols that collect serum-free CM for secretome analysis attempt to minimize cytolysis by optimizing serum-free incubation at 60-70% cell confluency for 24 hours, generally (Mbeunkui *et al.*, 2006; Severino *et al.*, 2013). While this strategy can minimize cell cytolysis, cytolysis could be further minimized by using cell lines that do not require the use of serum. This is observed in the application of HBEC3KT, a cell line routinely

cultured under serum-free conditions, as a control to study secretome changes during NSCLC metastatic progression (Hu *et al.*, 2016).

1.6 Thesis rationale

NSCLC secretome research has been limited to the study of established cancer cell lines, limiting further understanding of the changes in secreted proteins that may occur during cellular transformation. Studying the changes in secreted proteins that occur during stepwise *EGFR*-driven transformation of lung epithelial cells may yield potential biomarker candidates for diagnostic or prognostic use. This is important in the larger context of LUAD detection, as *EGFR*-driven transformation is under characterized and screening guidelines have not been established for never smokers where *EGFR*-driven LUAD is frequently observed (Lam, 2019; Rudin *et al.*, 2009).

1.7 Thesis hypothesis and objective

I **hypothesize** that secreted proteins will differ between wild-type, non-transformed lung epithelial cells, compared to those expressing clinically relevant mutations observed in *EGFR*-driven models of transformation or *EGFR* mutant cancer cell lines.

The **objective** of this thesis is to establish an experimental pipeline to quantitatively assess the secretome of *EGFR* mutant "malignant" (defined as *EGFR* mutant NSCLC cell lines) and "pre-malignant" (defined as non-transformed lung epithelial cell lines expressing clinically relevant mutations in *EGFR*-driven LUAD) stages of transformation. This consists of generating a model and a mass spectrometry method to analyze and quantify secreted proteins. To achieve this objective, there are three aims:

Aim 1: Develop an *in vitro* oncogenic model of *EGFR* cellular transformation, consisting of normal, pre-malignant, and malignant stages.

Aim 2: Develop a mass spectrometry-based protocol to enrich for and quantify secreted proteins from conditioned media of the model developed in Aim 1.

Aim 3: Apply the mass spectrometry-based protocol developed in Aim 2 to identify and quantify changes in secreted proteins for potential biomarker candidate validation.

After this introductory chapter, Chapter 2 will describe the development of the *in vitro* model and mass spectrometry protocol. Chapter 3 will detail the application of the protocol and identified biomarker candidates. Chapter 4 will discuss the research performed and suggest future avenues for further investigation.

Chapter 2: Establishment of cell line experimental model system and optimization of secretome collection conditions

2.1 Introduction

Current cancer secretome protocols enable the identification and quantification of secreted proteins, often incubating cancer cell lines under serum-free conditions (Brandi *et al.*, 2018; Makridakis & Vlahou, 2010). The limitation of this approach is that this may alter the cellular secretome due to cellular stress. Under serum starvation conditions, cells may undergo decreased proliferation, increased apoptosis, or activate signaling pathways not representative of normal culture conditions to maintain survival (Makridakis & Vlahou, 2010). Current NSCLC secretome studies have focused on advanced NSCLC, due to the nature of cell lines being derived from primary or metastatic tumors, investigating the role of secreted proteins in areas such as metastasis and TKI resistance (Bosse *et al.*, 2016; Hu *et al.*, 2016). However, solely profiling the secretome of cancer cell lines may not show gradual changes in secreted proteins occurring during tumor development, which could serve as potential biomarker candidates (J. E. Kim *et al.*, 2008). Yet changes in the secretome during tumor progression may vary, due to the dynamic nature of the secretome in local and systemic signalling, and thus affect biomarker development and implementation (Uhlén *et al.*, 2019). A study on untransformed mammary epithelial, non-metastatic and metastatic breast cancer cell lines implied secretome changes may indeed vary during tumor progression (Ziegler *et al.*, 2016). Malignant cell lines secreted more glycolytic enzymes compared to untransformed control, yet a non-metastatic cell line secreted more proteases relative to a metastatic cell line; these observations indicate stage-dependent

secretome profiles (Ziegler *et al.*, 2016). These findings suggest a temporal approach to biomarker development, where biomarker candidates may be most effective for a specific tumor stage or early or late stage tumor progression.

I sought to complement existing studies by analyzing the secretome of an *in vitro* oncogenic model of *EGFR* LUAD malignant transformation, consisting of normal, pre-malignant, and malignant stages. To achieve this, I have generated a model of LUAD transformation to study the secretome and have developed a serum-free mass spectrometry protocol to profile the secretome. The model consists of HBEC stable cell lines and selected *EGFR* mutant NSCLC cell lines, while the protocol can identify and enrich for secreted proteins with minimal effect on cell viability under nutrient deprivation. The novelty of this approach is the modeling of stepwise changes with HBEC stable cell lines, as noted clinically, and the application of serum-free HBEC culture conditions to profile the secretome (Inamura, 2018).

2.2 Materials and Methods

2.2.1 Cell lines and standard culture conditions

All cell lines used in this work were either obtained from the American Type Culture Collection (ATCC), or were gifted from Dr. Adi Gazdar. PC-9, H1975 (NCI-H1975), HCC4006, HCC4011, H3255 (NCI-H3255) were cultured in RPMI-1640 (Thermo Fisher Scientific), supplemented with 10% fetal bovine serum (FBS; Thermo Fisher Scientific) and 1% Penicillin-Streptomycin (Thermo Fisher Scientific). HBEC (HBEC3-KT) cells were cultured in Keratinocyte serum-free medium (KSFM; Thermo Fisher Scientific), supplemented with accompanying bovine pituitary extract (BPE; Thermo Fisher Scientific), human recombinant epidermal growth factor (EGF; Thermo Fisher Scientific), and 1% Pencillin-Streptomycin. All cell lines were grown in a humidified incubator at 37°C, 95% air and 5% CO_2.

2.2.2 Plasmids and generation of stable cell lines

HBEC stable cell lines expressing *EGFR L858R, GFP* vector control, in combination with *TP53* c-terminal fragment (*p53* CT) or *TP53* CT vector control were generated using the pLenti CMV (*EGFR L858R*, *GFP*) and pCX4 vectors (*TP53* CT, vector control). Vectors used to generate the EGFR L858R (Plasmids #82906, #17451) and GFP (Plasmid #17445) constructs were sourced from Addgene (Addgene, Cambridge, MA). The *TP53* vector and vector control construct (pCX4 hisD empty) were kindly provided by R. Somwar, Memorial Sloan Kettering Cancer Centre. Lentivirus used to generate cell lines were produced using HEK 239TD cells (ATCC), psPAX2 (Addgene Plasmid #12260) and pMD2.G (Plasmid #12259) or Phoenix-AMPHO cells (ATCC). Cell lines were selected with 5µg/mL blasticidin (Thermo Fisher Scientific), and 2mg/mL L-histidinol (Thermo Fisher Scientific). Unless specified otherwise, only the cell line expressing *EGFR L858R* and *TP53* CT (HBEC EGFR $^{L858R;}$ p53 ^{CT}) was also selected in 10µM Nutlin-3a for 6 days (SelleckChem).

2.2.3 Western blot analysis

Protein lysates were collected by washing cells with cold Dulbecco's phosphate-buffered saline (DPBS) (Thermo Fisher Scientific), then lysed with RIPA buffer (VWR) containing Halt protease and phosphatase inhibitor cocktail (Thermo Fisher Scientific) on ice and vortexed briefly. Samples were frozen at -80°C, sonicated, and supernatants separated via centrifugation at 15 000xg, at 4°C for 10 minutes. Protein concentrations were determined with a Pierce BCA protein assay kit (Thermo Fisher Scientific). Samples were processed by boiling in 1x diluted NuPAGE LDS sample buffer (Thermo Fisher Scientific) containing 1:10 diluted 2-Mercaptoethanol (MilliporeSigma) at 75°C, for 10 minutes. 20-25µg of samples were loaded onto NuPAGE 4-12% Bis-Tris protein gels (Thermo Fisher Scientific) and run in NuPAGE

MOPS SDS running buffer (Thermo Fisher Scientific) at 200V, for 50 minutes. Samples were transferred to Immobilon-P PVDF (MilliporeSigma) either at 70V, 4°C, for 2 hours or 30V, 4°C overnight, then blocked in TBS-T (0.1% Tween-20) (TBS, Bio-Rad; Tween-20, Thermo Fisher Scientific) containing 5% BSA (MilliporeSigma). Primary antibodies were prepared according to manufacturer's recommendations in TBS-T containing 5% BSA or 5% milk (MKP3). Membranes were incubated in primary antibodies at 4°C, overnight then incubated with HRP-conjugated secondary antibodies as per manufacturer's recommendations (Cell Signalling Technology). Membranes were then incubated with ECL, SuperSignal West Pico Plus Chemiluminescent Substrate (Thermo Fisher Scientific) or SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Fisher Scientific) and proteins detected with the ChemiDoc MP imager (Bio-Rad). The following antibodies were prepared to manufacturer's recommendations unless otherwise specified: p-ERK1/2(Thr202/Tyr204) (p-p44/42 (Thr202/Tyr204); Cell Signaling Technology, 9101); ERK1/2(p44/p42; Cell Signaling Technology, 4695); p-MEK1/2(Ser217/221) (Cell Signaling Technology, 9121); MEK1/2 (Cell Signaling Technology, 9122); p-EGFR(Tyr1068) (Cell Signaling Technology, 2234); EGFR L858R (Cell Signaling Technology, 3197); EGFR (Cell Signaling Technology, 2232); MKP3 (DUSP6) (Santa Cruz Biotechnology, sc-377070, 1:200); p53(Cell Signaling Technology, 2527); p53 (used to probe for p53 CT) (MilliporeSigma, SAB4503011); GFP (Cell Signaling Technology, 2956); β-Actin (Cell Signaling Technology, 12620, 1:2000).

2.2.4 Short term Nutlin-3a treatment

Prior to this experiment, HBEC cell lines expressing *p53* CT (GFP; $p53^{CT}$, $EGFR^{L858R}$;$p53^{CT}$) had been selected in 10μM Nutlin-3a for 6 days. 300 000 HBEC and 200 000 H1975 cells were seeded in standard culture media conditions in a 6 well plate and left to adhere overnight. Cells

were washed twice with PBS and media was changed. For HBECs, media was changed to KSFM, supplemented with BPE, EGF, 1% Penicillin-Streptomycin, and 10μM Nutlin-3a or 0.1% DMSO (Thermo Fisher Scientific) for 24 hours. H1975 cells were treated under the same conditions, using a base culture media of RPMI-1640 supplemented with 10% FBS, 1% Penicillin-Streptomycin.

2.2.5 Short term EGF treatment

250 000 HBEC cells were seeded in standard culture media in a 6 well plate and left to adhere overnight (with the exception of HBEC $EGFR^{L858R}$;$p53^{CT}$, cell lines were not previously selected with 10μM Nutlin-3a). Cells were then washed twice with PBS to remove supplements and media was replaced with supplement-free KSFM, 1% Penicillin-Streptomycin for 24 hours. After the elapsed time, cells were washed twice with PBS and media was changed to KSFM containing EGF (5ng/mL), 1% Pencillin-Streptomycin or supplement-free KSFM, 1% Penicillin-Streptomycin for 30 minutes.

2.2.6 Mutation analysis of *EGFR* mutant NSCLC cell lines

EGFR mutant NSCLC cell line mutation data (PC-9, H1975, HCC4006, H3255) was retrieved from The Cancer Cell Line Encyclopedia (CCLE) via The Cancer Dependency Map project (DepMap, release: DepMap Public 21Q4). Where mutation information was unavailable, appropriate literature references were used.

2.2.7 Trypan Blue viability assay

Cell lines were seeded at the following densities and incubated overnight in standard media conditions in a 6-well plate: 350 000 (HBEC); 340 000 (PC-9); 340 000 (H1975); 1 200 000 (HCC4006); 2 400 000 (HCC4011); 850 000 (H3255). Cells were washed twice with PBS to remove supplements, then switched to supplement-free or serum-free conditions for 24 hours.

HBECs were incubated with both supplement-free KSFM and 1% Penicillin-Streptomycin, or KSFM, containing BPE (50μg/mL), EGF (5ng/mL), and 1% Penicillin-Streptomycin. *EGFR* mutant NSCLC cell lines were incubated with serum-free RPMI-1640, and 1% Penicillin-Streptomycin or RPMI-1640 containing 10% FBS and 1% Penicillin-Streptomycin. After the elapsed time, cells were trypsinized with either 0.05% Trypsin-EDTA (Thermo Fisher Scientific; HBEC) or 0.25% Trypsin-EDTA (Thermo Fisher Scientific; *EGFR* mutant NSCLC), neutralized with trypsin neutralizer (Thermo Fisher Scientific; HBEC) or RPMI-1640 containing 10% FBS (*EGFR* mutant NSCLC), suspended in a 1:1 ratio with 0.4% Trypan Blue solution (Thermo Fisher Scientific) and live cell population was counted on a TC20 automated cell counter (Bio-Rad). Average percent live cell population was calculated from the average of 3 wells (2 counts per well).

2.2.8 Propidium iodide staining viability assay

Cell lines were seeded and treated under the same conditions as in Section 2.2.7. Post-incubation, cells were stained with 1μg/mL Hoescht 33342 (Thermo Fisher Scientific) and 1μg/mL propidium iodide (Thermo Fisher Scientific). Cells were incubated for each dye for 30 minutes and 10 minutes, respectively. Stained cells were visualized with an EVOS FL fluorescence microscope (Thermo Fisher Scientific). Average staining quantification of images was calculated from the average of 2 images with ImageJ software. The formula $100\% - \left(\frac{\text{number of PI stained cells}}{\text{number of Hoescht 33342 stained cells}} \times 100\right)$ was used to calculate live cell population per sample and condition.

2.2.9 Secretome optimization experiment collection

Secretome collection was performed under differing conditions, described in the following sub-sections.

2.2.9.1 Initial secretome optimization experiment

For the first secretome optimization experiment, 900 000 HBEC *GFP p53* c-terminal(GFP;$p53^{CT}$) cells were seeded in a 6 cm plate under standard culture conditions overnight with another 6cm plate containing only culture media; this was done in singlet. Cells were incubated for 24 hours, then conditioned media was collected by centrifuging media at 1000 RPM for 5 minutes, at 20°C, and filtered with a 0.2µm syringe filter (Sarstedt) to remove cell debris. A 1mL fraction (of 6mL collected) of the filtered media was concentrated using an Amicon Ultra-0.5 3kDa centrifugal filter unit (MilliporeSigma), to approximately $1/5^{th}$-$2/5^{th}$ of the original volume by centrifuging media at 3220xg, 4°C. Concentrated media was buffer exchanged to 50mM HEPES buffer, pH 7.0, by centrifuging twice with 1mL HEPES in the Amicon Ultra-0.5 3kDa centrifugal filter unit under the same centrifuge conditions to a final volume of approximately the same volume, 200µL. Samples were stored at -80°C until processing for mass spectrometry.

2.2.9.2 Follow-up secretome optimization experiment

For the second secretome optimization experiment, 900 000 HBEC GFP cells were seeded in 6cm plates, under standard culture conditions and incubated overnight along with 6cm plates containing only culture media; this was done in triplicate. Media was changed at 48 hours in a 96 hour period, then plates were washed thrice with 1mL PBS, then switched to KSFM containing EGF (5ng/mL), and 1% Penicillin-Streptomycin and incubated for 72 hours. Conditioned media was centrifuged at 1000 RPM for 5 minutes, at 20°C, and filtered with a 0.2µm syringe filter (Sarstedt) to remove cell debris. The complete volume of filtered media, 5mL, was concentrated with a Vivaspin Turbo 4 3kDa ultrafiltration unit (Sartorius) to approximately $1/25^{th}$ of the original volume by centrifuging media at 3220xg, at 4°C. Concentrated media was buffer

exchanged to 50mM HEPES buffer, pH 7.0, by centrifuging twice with 1mL HEPES in the Vivaspin Turbo 4 3kDa ultrafiltration unit under the same centrifuge conditions to a final volume of approximately 200-400µL. Samples were stored at -80°C until processing for mass spectrometry.

2.2.10 Secretome MS/MS analysis

Secretome protein samples were processed for mass spectrometry using a standard protocol consisting of reduction, alkylation, and digestion. Briefly, samples were reduced by adding 16µL of 200 mM dithiothreitol (DTT, Bio-Rad) and incubated for 30 minutes at 55 °C, then alkylated by adding 32 µL of 400 mM iodoacetamide (IAA, Bio-Rad) and incubating for 30 minutes at room temperature. The reaction was quenched by adding another 16 µL of 200 mM DTT. Proteins were digested in Typsin/Lys-C mix, which was prepared by adding 200 µL of 200 mM HEPES pH 8.0 into 20 µg Trypsin/Lys-C (Promega). Into each sample, 16 µL Typsin/Lys-C mix was added and samples were incubated on a ThermoMixer overnight at 1000 rpm, 37°C.

One tenth of each sample was pooled and prepared for quality check. The peptides were acidified by adding 10% (v/v) trifluoroacetic acid (TFA, Thermo Fisher Scientific) and diluting to a concentration of 1% TFA. Peptides were desalted using Stop And Go Extraction (STAGE) tip protocol (Rappsilber *et al.*, 2003). After elution, solvent was evaporated by centrifuging samples in a SpeedVac Vaccum Concentrator, until samples were dried. Peptides were then reconstituted in an aqueous solution containing 0.1% formic acid, 1% DMSO. Reconstituted peptides were run on a LTQ Orbitrap Velos™ and assessed for quality.

Following a successful quality check, the remaining digested peptides were labeled with a tandem mass tag (TMT) 6-plex isobaric mass tagging kit (Thermo Fisher Scientific) according to manufacturer instructions. After TMT-labeling, samples were pooled, speed vacuumed to

evaporate excess solvent, and acidified by adding TFA as described above. Peptides were then desalted by STAGE tip protocol, and speed vacuumed to evaporate remaining solvent. Peptides were reconstituted in an aqueous solution containing 0.1% formic acid, 1% DMSO and run on an Orbitrap Fusion^{TM} mass spectrometer set to MS2 mode. In the first secretome experiment, singlet samples were run in 2 6-plex TMT channels, while in the second experiment technical triplicate samples were run in 6 6-plex TMT channels (1 sample per channel). MS spectra were searched in the Proteome Discoverer suite (v.2.4.0.305, Thermo Fisher Scientific). Peptide-spectrum matches (PSMs) were validated with Percolator, where only PSMs with false discovery rate (FDR) < 0.05 were retained in the analysis.

2.2.11 Secretome enrichment and bioinformatic analysis

Proteins were assessed for enrichment relative to media control by calculating the average fold change by dividing the raw, untransformed signal intensity of cell line conditioned media by media control (cell line/media). Enriched proteins were defined as those with a relative fold change of 2 or greater. Enriched proteins were analyzed for functional and cellular component annotation with Gene Ontology (GO) databases (GO release: 2021-11-16, DOI: 10.5281/zenodo.5725227; 2022-01-13, DOI: 10.5281/zenodo.5874355) and UniProt (UniProt release: 2021_3; 2021_4). GO functional analysis was performed with the Fisher's exact test, and false discovery rate (FDR) calculated.

2.2.12 Statistical analysis

Statistical analysis, unpaired two-tailed *t*-test, was performed with Prism6 (GraphPad) software for Figure 2.4 and Figure 2.5. Unless indicated, data was calculated as the average of three biological replicate experiments and presented as mean +- SEM. *P < 0.05, ** P < 0.01, n.s. not significant.

2.3 Results

2.3.1 Validation of gene expression in HBEC stable cell lines

In developing a model of mutant EGFR LUAD transformation, I generated stable HBEC cell lines expressing $EGFR^{L858R}$ or GFP vector control, with or without p53 CT (GFP; $p53^{wt}$, GFP; $p53^{CT}$, $EGFR^{L858R}$;$p53^{wt}$, $EGFR^{L858R}$;$p53^{CT}$) as to represent the commonly mutated genes in *EGFR*-driven LUAD (Inamura, 2018; Skoulidis & Heymach, 2019). Expression of these genes was confirmed by treating cells for 24 hours with 10μM Nutlin-3a, an inhibitor for the upstream p53 inhibitor MDM2 and comparing relative protein expression via Western blot as seen in Figure 2.1 (Kojima *et al.*, 2006; Zhao *et al.*, 2015). Figure 2.1 shows that most cell lines expressing p53 CT have little relative change in p53 protein expression when treated with Nutlin-3a, as expected of a mutant p53 phenotype when comparing to a mutant p53 cell line, H1975 (Inoue *et al.*, 2020). The exception was HBEC $EGFR^{L858R}$; $p53^{CT}$, where p53 expression increased. This may be due to basal wild-type p53 expression, despite the introduction of p53 CT (Ramirez *et al.*, 2004). In contrast, cell lines expressing the vector control for p53 CT show an increase in relative p53 protein levels; this is expected of wild-type p53 cell lines. $EGFR^{L858R}$ and total EGFR expression are also confirmed in Figure 2.1. Interestingly, total EGFR expression was lower in HBEC cell lines expressing EGFR L858R relative to GFP control cell lines. This could be due to a combination of insufficient EGFR L858R antibody stripping and cross-reactivity with wild-type EGFR from epitope similarity (Cell Signaling Technology, 2022).

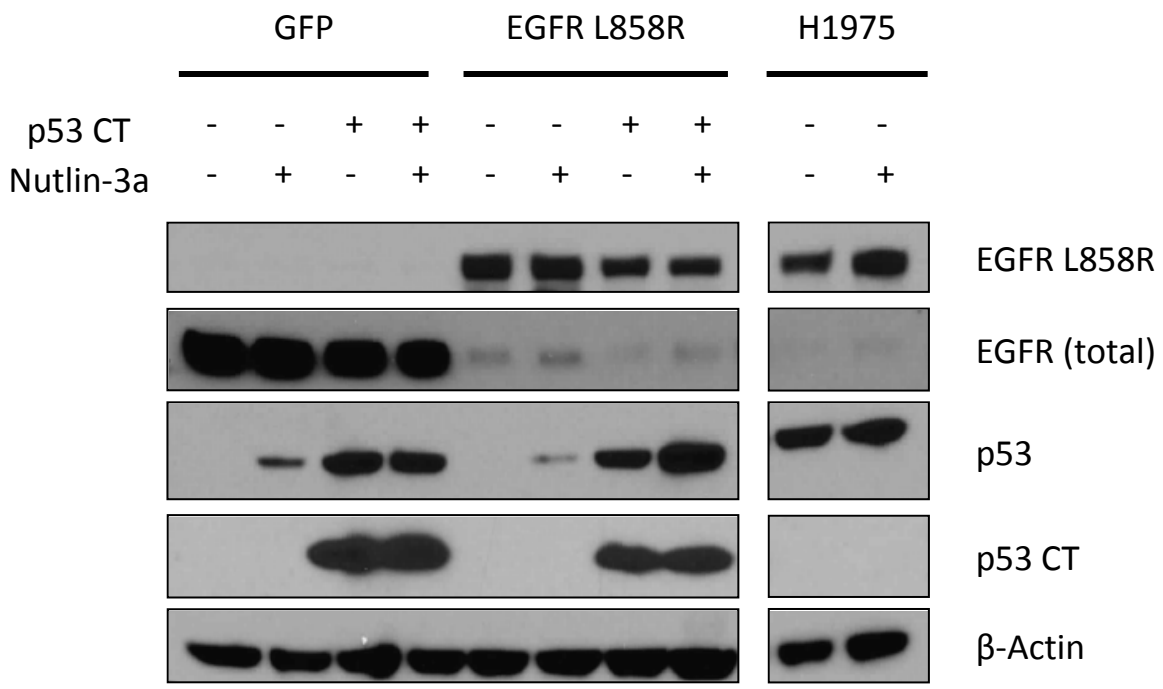


Figure 2.1 Western blot of HBEC stable cell lines following 24 hour incubation with 10μM Nutlin-3a. HBEC stable cell lines expressing GFP or EGFR L858R, in combination with p53 CT were treated with 10μM Nutlin-3a for 24 hours to confirm expression of mutant p53. Basal expression of EGFR L858R was also detected. H1975 was used as a positive control for mutant p53 expression. All samples were probed on the same membrane (extraneous samples were removed). Blot is representative of one biological replicate; HBEC cell lines expressing p53 CT were previously selected with 10μM Nutlin for 6 days.

2.3.2 Validation of ERK signaling pathway expression in HBEC stable cell lines

To further validate that my HBEC cell lines had signaling pathway expression consistent with constitutive EGFR activation, I analyzed relative phosphorylation status of downstream ERK signaling pathway proteins after treating cells with EGF for 30 minutes. As seen in Figure 2.2, P-EGFR (Y1068) and P-ERK (T202/Y204) were increased in cell lines treated with EGF; this was expected of EGF stimulation. However, in HBEC $EGFR^{L858R}$; $p53^{wt}$, P-ERK was not noticeably different between untreated and treated cells. As it was unclear whether ERK signaling via P-ERK was greater in HBEC $EGFR^{L858R}$ cell lines relative to GFP vector control, I also assessed relative protein levels of P-MEK (S217/221), upstream of ERK. Figure 2.2 shows increased levels of P-MEK in HBEC $EGFR^{L858R}$ cell lines relative to GFP vector control,

suggesting that ERK pathway signalling is increased in mutant EGFR-expressing cell lines, despite an unclear P-ERK output.

While relative protein expression of P-ERK was greater in EGF-treated cell lines, it was unclear if there were differences between HBECs expressing GFP or EGFR ^{L858R}; constitutive activation of EGFR would likely have higher expression of P-ERK. To investigate whether there was inhibition of ERK phosphorylation, I assessed DUSP6 (MKP3) expression between EGF-treated and untreated cell lines. In Figure 2.2, there is no appreciable difference between treated and untreated conditions, but interestingly, HBEC EGFR^{L858R};p53^{wt} showed greater levels of DUSP6 relative to other cell lines regardless of treatment. Another interesting observation was that total EGFR protein expression in EGF-treated cell lines was decreased relative to untreated cell lines. This may be due to decreased availability of the overall EGFR intracellular domain upon phosphorylation of the intracellular tyrosine kinase domain, due to receptor endocytosis (Sorkin & Goh, 2009).

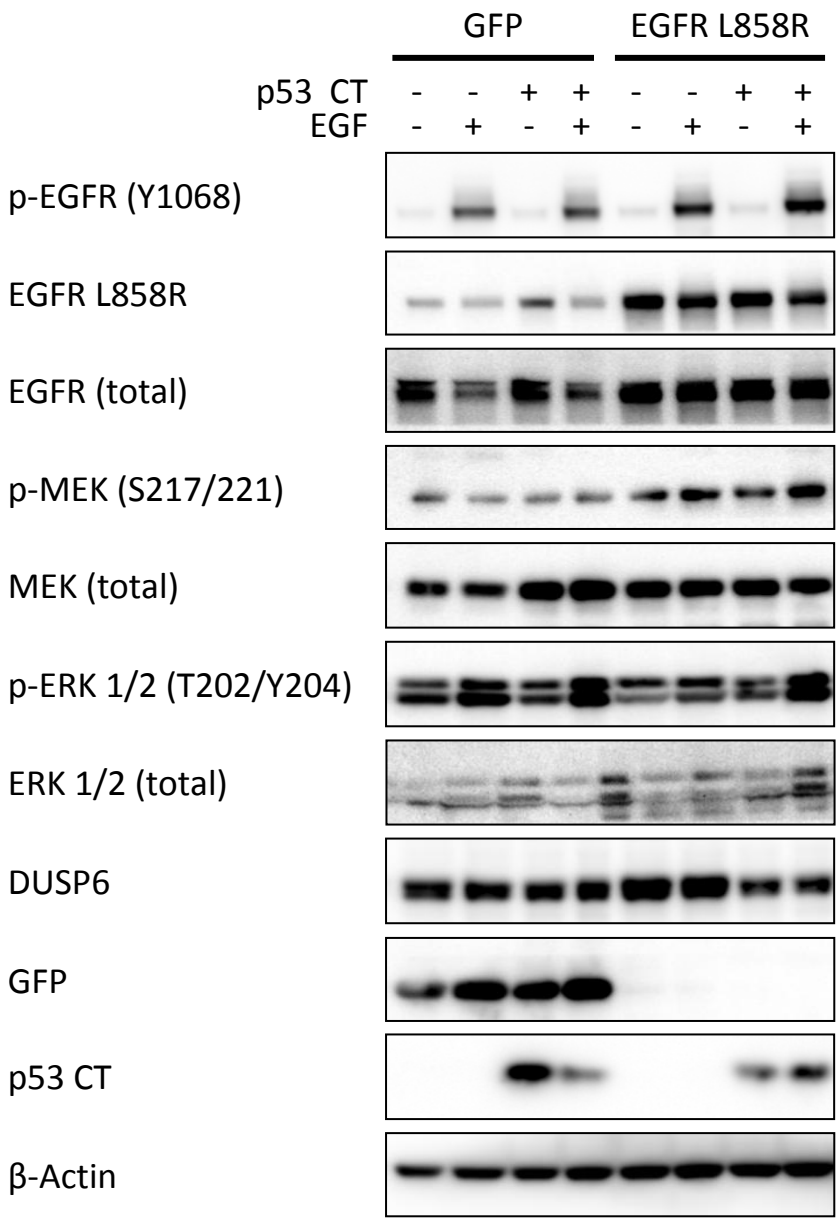


Figure 2.2 Western blot of HBEC stable cell lines after 30 min incubation with EGF.
HBEC stable cell lines expressing GFP or EGFR L858R, in combination with p53 CT were starved in supplement-free KSFM for 24 hours then treated with EGF for 30 minutes to assess ERK signaling pathway activity between EGFR L858R-expressing HBEC cell lines and GFP control. Blot is representative of two biological replicates.

2.3.3 Confirmation of EGFR NSCLC cell line *EGFR* and *TP53* mutations

In addition to generating HBEC stable cell lines, I also selected NSCLC cell lines expressing mutant EGFR and of varying *TP53* status to act as the transformed, malignant state of the experimental model. I utilized The Cancer Cell Line Encyclopedia (CCLE) to confirm genetic alterations in these genes, with the exception of HCC4011. Mutation information for HCC4011 was confirmed with previous publications (Kubo *et al.*, 2009; Pikor *et al.*, 2013). Table 2.1 summarizes mutant *EGFR* and *p53* status for the selected cell lines.

Table 2.1 *EGFR* mutant NSCLC cell lines and their *EGFR* and *TP53* status.

Cell line	Primary EGFR mutation	Secondary EGFR mutation	TP53 status	Cancer Cell Line Encyclopedia DepMap ID
PC-9	Ex19Del	NA	WT	ACH-000779
HCC4006	Ex19Del	NA	MUT (Y205H)	ACH-000066
H1975	L858R	T970M	MUT (R273H)	ACH-000587
HCC4011	L858R (Kubo *et al.*, 2009)	NA	MUT (Deletion; (Pikor *et al.*, 2013))	NA
H3255	L858R	NA	WT	ACH-000109

2.3.4 Initial secretome collection optimization experiment

To determine whether I could detect secreted proteins, I performed a secretome collection optimization experiment under standard media conditions to reflect ideal viability conditions. Recent studies have been performed under serum-free or limited supplement conditions, which

may interfere with viability and contaminate results with intracellular proteins (Shin *et al.*, 2019). Therefore, I applied the serum-free culture conditions of the HBEC cell line to perform my secretome collection experiment.

In my initial secretome collection optimization experiment, I collected the conditioned media of the HBEC GFP; $p53^{CT}$ cell line. Figure 2.3 illustrates the methodology and results of this experiment. 591 proteins were detected from a 1mL fraction collected after a 24 hour incubation period. To assess if proteins were enriched in HBEC GFP; $p53^{CT}$ relative to media control, I calculated relative fold enrichment from the raw signal intensity values; 45 proteins were greater than or equal to an enrichment of 2-fold. As shown in Figure 2.3B, the maximum fold enrichment observed was 8-fold. This suggests that while proteins can be detected from a partial sample volume and short-term incubation period, the dynamic range was limited and less abundant proteins may be difficult to detect.

To determine if detected proteins were secreted, I performed a cellular component analysis of the 45 proteins enriched at 2-fold or greater using the UniProt database to identify protein subcellular localization. I analyzed keywords associated with protein subcellular location of the 45 proteins, and 19 were annotated as "secreted" (UniProt KW-0964). Figure 2.3C illustrates all cellular component annotations associated with the 45 protein input. Interestingly, an equal amount of proteins were annotated as "cytoplasmic". This may be due to a basal level of intracellular protein contamination during sample preparation, or basal membrane protein ectodomain shedding due to growth factors in the incubation media (Dowling & Clynes, 2011; Hayashida *et al.*, 2010). Alternatively, the UniProt database may not have the full scope of secreted proteins to adequately annotate them, as subcellular annotations can be inferred from multiple evidence sources. Sources include experimental evidence - which may be derived from

different cellular models - curator inference, and sequence similarity; accordingly, there may be insufficient information to annotate a protein as “secreted” (UniProt Consortium, 2022). Other cellular components included “nucleus” and “membrane”, further suggesting intracellular protein contamination and membrane shedding.

While 19 of the 45 identified enriched proteins were annotated as “secreted”, 6 proteins, Annexin A2 (ANXA2), 14-3-3- protein sigma (SFN), Annexin A1 (ANXA1), Macrophage migration inhibitory factor (MIF), Galectin-1 (LGALS1), and Peptidyl-prolyl cis-trans isomerise A (PPIA) do not possess the signal peptide sequence commonly associated with conventional protein secretion (Owji *et al.*, 2018). This indicates the presence of alternative secretion pathways, yet these proteins were observed without culture conditions that would promote stress (Rabouille, 2017). I also observed that 9 of the 19 proteins identified had multiple subcellular annotations beyond the scope of secretion, which were defined as the annotations “secreted”, “extracellular space”, “extracellular matrix”, and “basement membrane”. These included annotations such as “cytoplasm” and “nucleus”. This reflects the broad scope of the UniProt annotation system, where protein subcellular location can be inferred from multiple sources; this includes different cell lines that have varying secretion pathways based on experimental conditions.

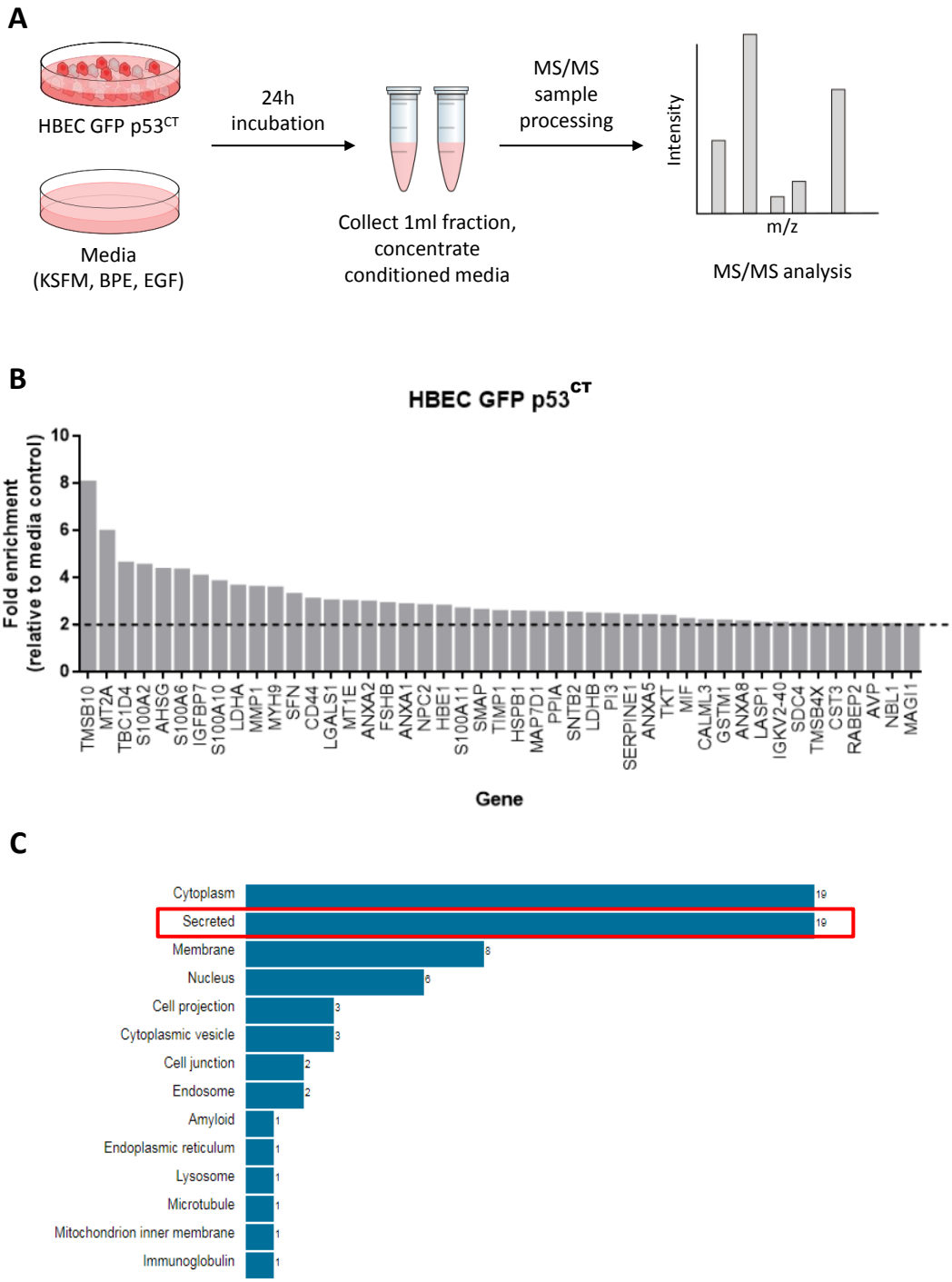


Figure 2.3 Summary of first secretome collection optimization experiment.
(A) Schematic overview of secretome collection conditions. HBEC GFP;p53^{CT} cells were incubated under standard media conditions, in parallel with a media control, for 24 hours. A 1mL fraction was collected and subjected to sample processing and MS/MS analysis. (B) Bar graph indicating proteins (referred to by gene symbol) that were a minimum of 2-fold enriched in HBEC GFP;p53^{CT} sample (relative to media control). (C) Complete UniProt database cellular component identification of enriched proteins.

2.3.5 Functional annotation analysis of enriched proteins from initial secretome collection optimization experiment with GO

Although there were no additional cell lines to compare the secretome composition of HBEC GFP; $p53^{CT}$, I performed a GO Slim Biological Process analysis of the secreted proteins identified to gain a broad overview of their biological functions. Table 2.2 displays the GO Slim term output sorted by FDR (FDR < 0.05). Interestingly, proteolysis terms consist of half of the overrepresented terms. This is reflected in the terms that include either general or negative regulation of peptidase and endopeptidase activities. These annotations are consistent with the broad scope of protease functions, which include involvement in biological processes such as cell proliferation, autophagy, and the immune response (López-Otín & Bond, 2008).

Table 2.2 PANTHER GO-Slim Biological Process analysis of secreted proteins identified in first secretome optimization experiment (FDR < 0.05).

GO ID	GO term	Fold enrichment	FDR
GO:0051346	negative regulation of hydrolase activity	48.01	3.51E-03
GO:0043086	negative regulation of catalytic activity	26.19	1.79E-02
GO:0045861	negative regulation of proteolysis	41.3	2.09E-02
GO:0010466	negative regulation of peptidase activity	43.88	2.11E-02
GO:0044092	negative regulation of molecular function	21.9	2.37E-02
GO:0010951	negative regulation of endopeptidase activity	45.3	2.41E-02
GO:0052547	regulation of peptidase activity	32.66	3.08E-02
GO:0052548	regulation of endopeptidase activity	33.43	3.29E-02

2.3.6 Cell viability assessment under secretome conditions

I next sought to assess whether secretome collection conditions affected cell viability for HBEC and *EGFR* mutant NSCLC cell lines. To minimize HBEC EGF ligand-mediated signaling from EGF-containing media, the HBEC secretome incubation media was changed to supplement-free KSFM. This was done to create an equivalent to serum-free media, which is

used in secretome experiments (Méndez & Villanueva, 2015). To determine if the selected collection conditions affected cell viability, I performed Trypan Blue staining, and propidium iodide (PI) staining as a corollary method to assess changes in cell viability. HBEC and *EGFR* mutant NSCLC cell lines were incubated under supplement-free or serum-free conditions for 24 hours prior to staining cells with the above reagents. The incubation conditions were consistent with the initial secretome collection optimization experiment and previous secretome experiment literature, where 24 hours was chosen to minimize cell death from serum starvation (Mbeunkui *et al.*, 2006; Shin *et al.*, 2019). Figure 2.4 and Figure 2.5 show the changes in cell viability in both assays, where live cell population was calculated with the formula $100\% - \left(\frac{number\ of\ PI\ stained\ cells}{number\ of\ Hoescht\ 33342\ stained\ cells} \times 100\right)$. In either assay, the majority of cell lines show no significant changes in cell viability, observed as percentage of live cell population. However, HCC4011 was difficult to assess for viability, as cells were prone to detachment during PI staining (Figure 2.5C). PI staining of H3255 yielded a significant ($P < 0.01$) change in the proportion of live cells stained; however, the change in proportion was minimal and did not signify a biologically significant change in cell viability (Figure 2.5B).

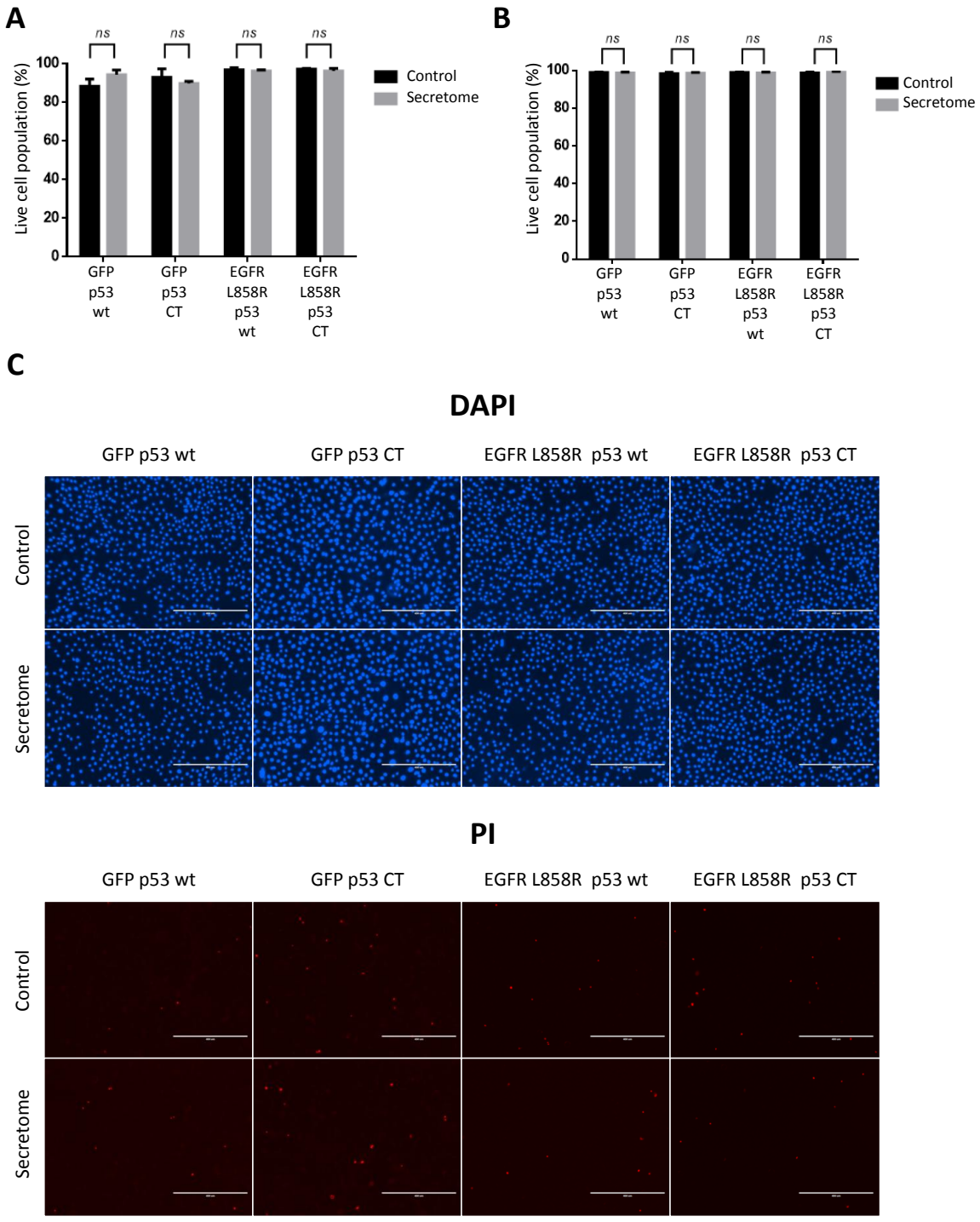


Figure 2.4 HBEC cell viability analysis under secretome collection conditions.
(A) Live cell population of HBEC cell lines incubated in supplement-free KSFM (secretome) or standard media conditions for 24 hours, as stained with Trypan Blue. (B) Live cell population of HBEC cell lines treated under the same conditions as in (A), as stained by PI. (C) Representative images used to quantify PI staining (DAPI and PI channels). Graphs are representative of three biological replicates.

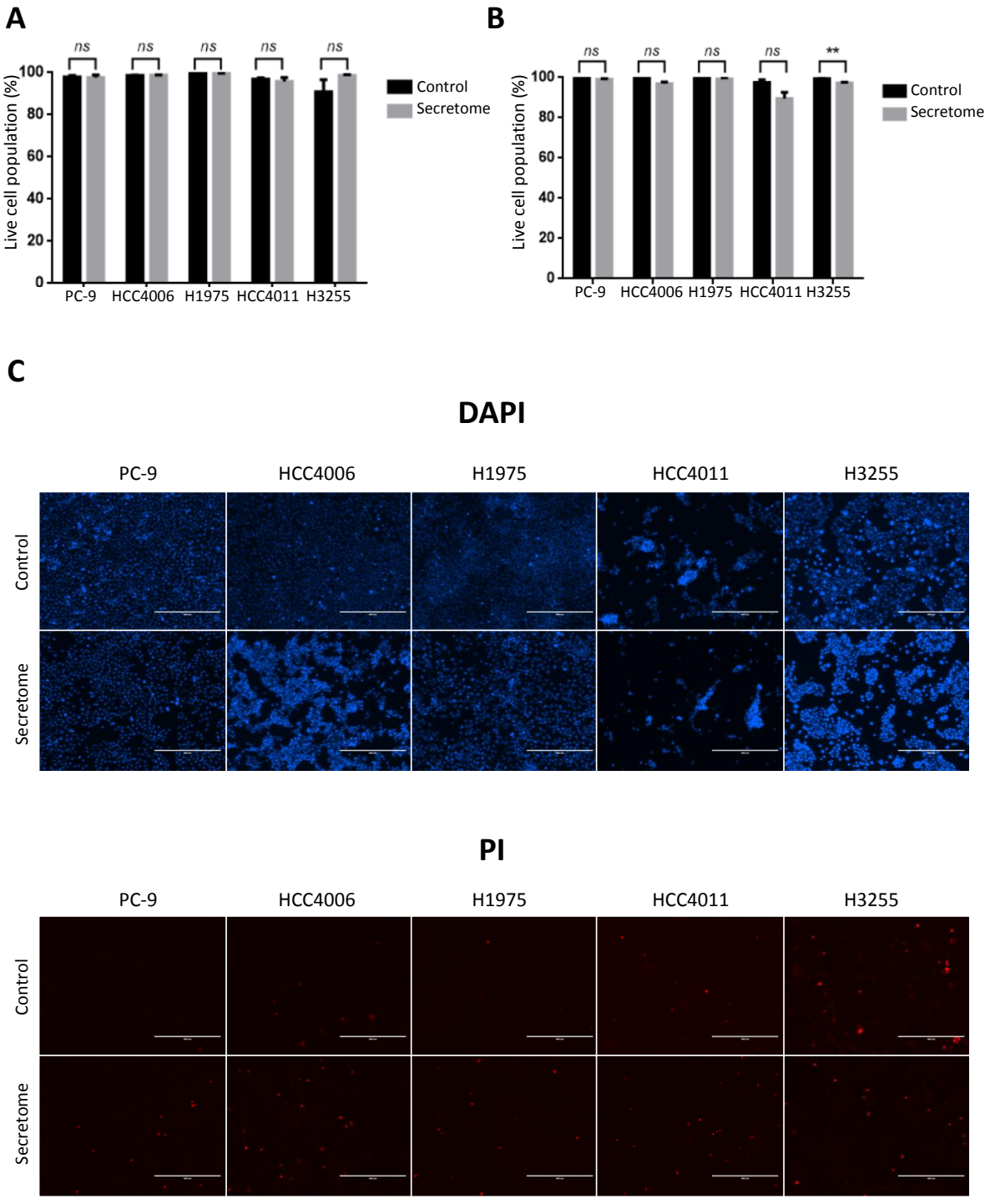


Figure 2.5 *EGFR* mutant NSCLC cell viability analysis under secretome conditions.
(A) Live cell population of *EGFR* mutant NSCLC cell lines incubated in serum-free RPMI-1640 (secretome) or standard media conditions for 24 hours, as stained with Trypan Blue. (B) Live cell population of *EGFR* mutant NSCLC cell lines treated under the same conditions as in (A), as stained by PI. (C) Representative images used to quantify PI staining (DAPI and PI channels). Graphs are representative of three biological replicates.

2.3.7 Follow-up secretome collection optimization experiment

I next sought to enhance the fold enrichment range of identified secreted proteins by performing a second secretome collection optimization experiment. To increase the fold enrichment range, the following variables were altered: incubation media, incubation time, cell confluency, and sample volume collected. The previous experiment had standard culture media containing BPE; the exact composition of BPE is unknown but includes growth factors such as EGF-like proteins, fibroblast growth factor (FGF), and platelet-derived growth factor (PDGF) that may mask less abundant proteins similar to FBS (Dowling & Clynes, 2011; Kent & Bomser, 2003). To minimize BPE protein masking, I changed the incubation media for collection to culture media without BPE. In addition, I changed the incubation time from 24 to 72 hours and pre-incubated the selected cell line HBEC GFP for 96 hours. These changes were made to increase the detection of low abundance proteins upon collection by increasing their concentration over a longer period. To further enable detection, I also collected the complete 5mL volume of conditioned media rather than 1mL fraction, for MS/MS analysis. Figure 2.6 summarizes the methodology and results of this experiment.

In the follow-up secretome collection optimization experiment, 438 proteins were identified. As the cell line used in this experiment differs from the previous HBEC GFP;$p53^{CT}$, direct comparisons cannot be made. Further, peptide signal intensities will vary between MS experiments due to the stochastic sampling during MS analysis. Variable sampling can be accounted for with a pooled standard or reference channel per experiment, yet neither were performed due to limited sample channels (Brenes *et al.*, 2019; Plubell *et al.*, 2017). Despite these constraints, 393 proteins were observed at a minimum 2-fold enrichment relative to the

media control. Figure 2.6B shows the top 50 enriched proteins from the second experiment, with a max 74-fold enrichment. The majority of proteins observed in this experiment were enriched at minimum 2-fold. This observation suggests that altering the incubation media, duration of incubation, and pre-incubation of cell line used may be conducive to detecting proteins commonly secreted in low abundance. I then determined the proportion of secreted proteins in the enriched protein subset, performing a cellular component analysis as done in the initial secretome optimization experiment. Selected cellular component annotations are shown in Figure 2.6C. Interestingly, the "membrane" annotation comprised the top cellular component annotation, consisting of 169 proteins, while "secreted" proteins comprised the second most abundant annotation, consisting of 146 proteins. Of the secreted protein subset, 16 proteins did not possess a signal peptide, and 80 proteins had multiple subcellular annotations beyond secretion-associated terms "extracellular space", "extracellular matrix", and "basement membrane"; "membrane" was the most common non-secretion subcellular annotation. The presence of both annotations suggests that membrane protein ectodomain shedding may be occurring (Hayashida *et al.*, 2010). Alternatively, multiple annotations of proteins further suggest that the UniProt annotation system is broad, due to multiple cellular models, multiple protein isoforms listed under a single entry, and other factors used to annotate protein subcellular location. The 72 hour incubation under supplement-limited culture conditions may have also led to cell death, and membrane and cytoplasmic protein release due to loss of plasma membrane integrity (Figure 2.6C) (Fulda *et al.*, 2010).

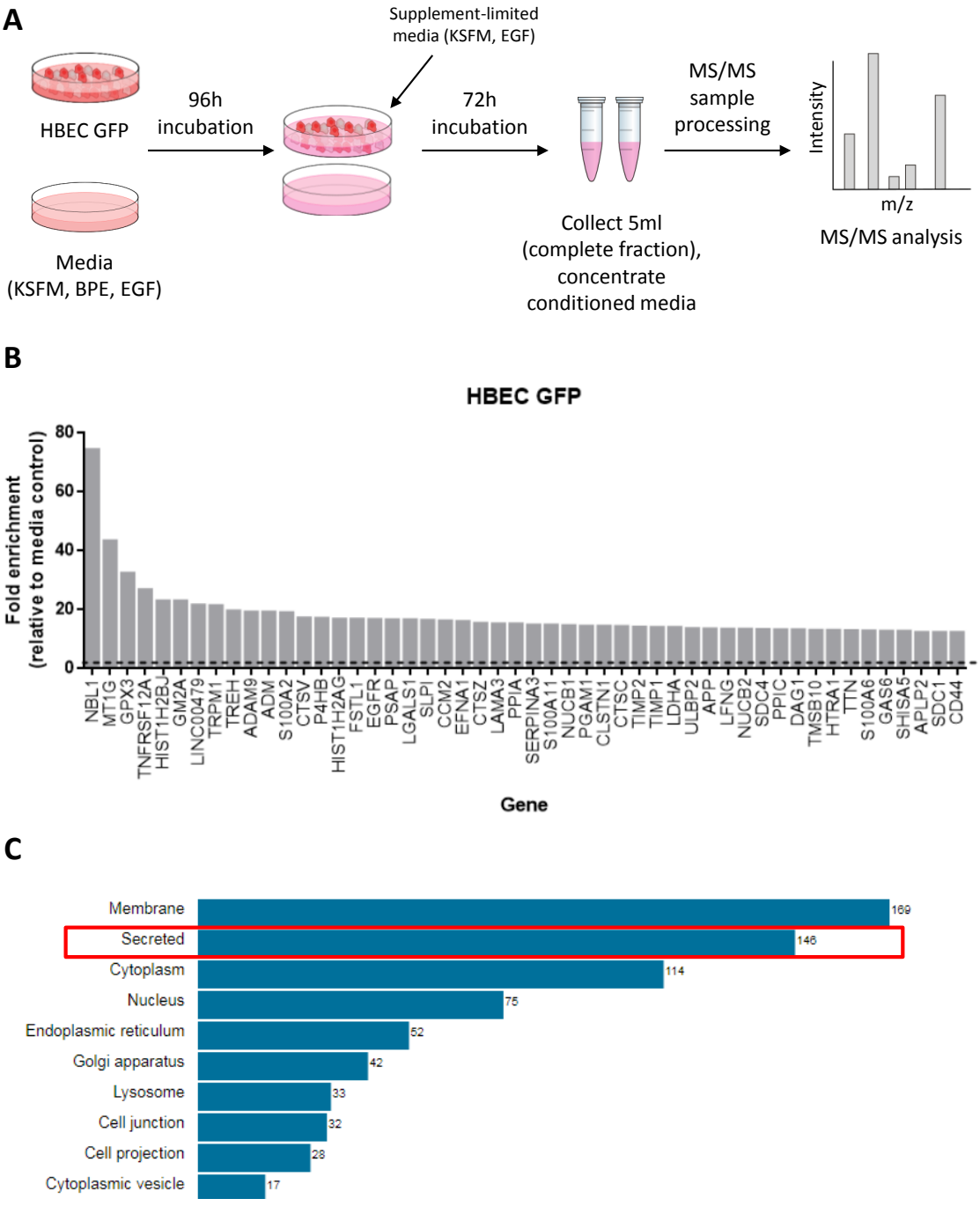


Figure 2.6 Summary of second secretome collection optimization experiment.
(A) Schematic overview of secretome collection conditions. HBEC GFP cells were incubated under standard media conditions, in parallel with a media control, for 96 hours then incubated under supplement-limited conditions for 72 hours. The complete 5mL conditioned media volume was collected and subjected to sample processing and MS/MS analysis. (B) Bar graph illustrating top 50 enriched proteins (referred to by gene symbol) by fold enrichment (relative to media control). Dotted line denotes 2-fold enrichment cut-off. (C) Selected UniProt database cellular component identification of enriched proteins.

2.3.8 Functional annotation analysis of enriched proteins from follow-up secretome collection optimization experiment with Gene Ontology

After identifying the secreted protein subset, I performed GO analysis as previously done to examine protein biological function. As there was no cell line processed in parallel, I performed a GO Slim Biological Process analysis to obtain a broad overview of biological functions represented. A selected subset of top GO terms is represented in Table 2.3. There was no distinct biological process overrepresented; however, terms associated with proteolysis and general biological development were commonly observed. These terms may reflect common cell signaling pathways in cell growth, division, and development, and the broad range of protease biological functions such as involvement in processes including the cell cycle and protein and organelle cycling, as noted from the terms in Table 2.2 (Turk *et al.*, 2012). The overrepresentation of these terms may be due to incubation under supplement-limited conditions, promoting signaling pathways involved in general biological processes that may be classified as developmental processes.

Table 2.3 Selected PANTHER GO Slim Biological Process analysis of secreted proteins identified in follow-up secretome optimization experiment (FDR < 0.05).

GO ID	GO term	Fold enrichment	FDR
GO:0009888	tissue development	12.61	7.72E-09
GO:0007167	enzyme linked receptor protein signaling pathway	7.08	6.86E-07
GO:0051346	negative regulation of hydrolase activity	16.93	8.57E-07
GO:0010951	negative regulation of endopeptidase activity	19.16	1.66E-06
GO:0010466	negative regulation of peptidase activity	18.57	1.71E-06
GO:0009887	animal organ morphogenesis	12	1.75E-06
GO:0007275	multicellular organism development	3.79	1.92E-06
GO:0045861	negative regulation of proteolysis	17.47	1.98E-06
GO:0048856	anatomical structure development	3.4	2.17E-06
GO:0048513	animal organ development	5.83	2.22E-06
GO:0052548	regulation of endopeptidase activity	14.14	6.77E-06
GO:0052547	regulation of peptidase activity	13.82	7.49E-06
GO:0040011	locomotion	5.5	9.54E-06
GO:0009653	anatomical structure morphogenesis	4.83	9.83E-06
GO:0032502	developmental process	3.01	1.70E-05
GO:0048583	regulation of response to stimulus	3.29	2.79E-05
GO:0048731	system development	3.61	3.11E-05
GO:0043086	negative regulation of catalytic activity	9.23	3.17E-05
GO:0030198	extracellular matrix organization	10.24	5.08E-05
GO:0043062	extracellular structure organization	10.16	5.17E-05

2.4 Discussion

Established secretome studies have utilized serum-free media methods to identify and quantify secreted proteins in cancer cell lines (Brandi *et al.*, 2018). The benefit of a serum-free approach is to minimize masking of low abundance proteins by highly abundant serum proteins. However, cell lines that are grown under serum-free conditions risk cellular stress and cell death due to nutrient deprivation (Shin *et al.*, 2019). Additionally, the use of cancer cell lines may not provide insight into changes occurring during malignant transformation, which may be a reservoir for biomarker candidates. Currently the HBEC cell line, routinely cultured under

serum-free conditions, is used as a non-tumorigenic control for secretome studies with established cancer cell lines (Hu *et al.*, 2016; X.-D. Wang *et al.*, 2019). Applying the non-transformed and serum-free culture conditions of HBECs is a suitable way to study secretome changes occurring during stepwise malignant transformation. To complement existing literature on secretome changes during NSCLC, which have been limited to advanced and metastatic stages of NSCLC, I developed a model of stepwise LUAD malignant transformation and a mass spectrometry protocol to profile the secretome (Böttger *et al.*, 2019; Hu *et al.*, 2016).

Firstly, I generated a set of stable HBEC cell lines expressing genetic alterations characteristic in NSCLC patients, *EGFR* and *TP53*. These cell lines express *EGFR L858R*, a common oncogenic mutation leading to constitutive *EGFR* signaling activation, and *p53* CT; *p53* CT was selected to emulate the frequent co-mutation of *TP53* in *EGFR*-driven LUAD and an early genetic alteration during tumorigenesis (Skoulidis & Heymach, 2019). The expression of *EGFR L858R* and *p53* CT and *EGFR*-mediated *ERK* signaling were confirmed by Western blot (Figure 2.1, Figure 2.2).

Secondly, I developed a method to optimally collect for secreted proteins in a serum-free context with HBEC and *EGFR* mutant NSCLC cell lines with two optimization experiments. In the first experiment, I confirmed the presence of secreted proteins with the HBEC GFP;$p53^{CT}$ cell line (Figure 2.3). Functional annotation of identified secreted proteins, which were the second most abundant subcellular location, revealed an overrepresentation of proteolysis terms, likely representative of the broad functions of proteases that include DNA replication and transcription, cell proliferation, and stress responses such as autophagy, senescence, and apoptosis (López-Otín & Bond, 2008) (Table 2.2). However, the max protein fold-enrichment was 8-fold. This suggests that while secreted proteins can be detected, abundant proteins in BPE

such as growth hormone, basic fibroblast growth factor, and epidermal growth factor-like proteins may mask less abundant proteins similar to albumin in serum, leading to the limited fold enrichment observed (Dowling & Clynes, 2011; Kent & Bomser, 2003). As I was interested in profiling the secretome in HBEC and *EGFR* mutant NSCLC cell lines, I sought to emulate serum-free conditions as done in previous secretome studies and assess cell viability under limited nutrient conditions (Brandi *et al.*, 2018). I chose to incubate HBEC cell lines in supplement-free media as an equivalent to serum-free conditions, and minimize signalling from BPE and EGF that may affect secretome composition (Kent & Bomser, 2003; Kerpedjieva *et al.*, 2012). As a 24 hour serum-free incubation period was observed to be optimal for secretome collection, I maintained a 24 hour incubation period as done with my initial experiment (Mbeunkui *et al.*, 2006). HBEC and *EGFR* mutant NSCLC cell lines were incubated under supplement-free or serum-free conditions for 24 hours, and I assessed cell viability by determining the live cell population with Trypan Blue and PI staining. There were no biologically significant changes to cell viability; this confirmed that the initial secretome collection conditions were suitable for secretome analysis with minimal risk of intracellular protein contamination due to cell death (Figure 2.4, Figure 2.5).

Because of the limited dynamic range of the first secretome experiment, I performed a follow-up experiment to optimize the depth of secretome profiled. I altered the incubation media, incubation period, and volume of conditioned media used for analysis. This consisted of pre-incubating the selected HBEC GFP cell line prior to collection and increasing incubation time, removing BPE from the incubation media to minimize signaling from growth factors and protein masking, and concentrating the complete volume of conditioned media to maximize detection of less abundant proteins (Dowling & Clynes, 2011; Kent & Bomser, 2003). The majority of

proteins identified were enriched at 2-fold or greater, suggesting that a low background incubation media, incubation period, and collection of the complete conditioned media may influence the breadth of the secretome that can be profiled (Figure 2.6). Secreted proteins also comprised the second most abundant subcellular location and functional annotation consisted of general cellular processes associated with development and proteolysis, which may be representative of general cellular pathways involved in proliferation and survival under supplement-limited incubation conditions (Figure 2.6, Table 2.3) (Ahmadiankia, 2020). However, cell viability was not assessed under supplement-limited or supplement-free conditions for a 72 hour incubation period for either HBEC or *EGFR* mutant NSCLC cell lines. Further analysis of cell viability under this incubation period will be useful in determining if prolonged nutrient deprivation will decrease cell viability and result in intracellular protein contamination (Brandi *et al.*, 2018). A study noted an increase in cytosolic protein contamination after incubating cells under serum-free conditions for 30 hours, indicating that cell viability would not be sufficiently optimal to profile the secretome after 24 hours (Mbeunkui *et al.*, 2006). Therefore, I finalized my incubation media and time to serum-free or supplement-free media for 24 hours, in addition to concentrating the complete volume of collected media.

2.5 Conclusion

In summary, I developed a model of *EGFR*-driven LUAD transformation and a serum-free mass spectrometry protocol to identify and quantify secreted proteins *in vitro*. The combination of stable HBEC cell lines generated and selected *EGFR* mutant NSCLC cell lines depict a model of *EGFR*-driven LUAD transformation, with clinically relevant genetic alterations. The serum-free protocol that was developed is optimized to profile cell lines normally cultured in serum-free and serum-containing conditions, enhance detection of low abundance secreted proteins, and minimize cell death under nutrient deprivation. With protein enrichment observed at a max fold-enrichment of 74-fold during optimization, and frequent subcellular annotation of secreted proteins, this protocol is a robust method to profile the secretome. Together, the model and protocol are a viable approach to study changes in the secretome during LUAD transformation.

Chapter 3: Application and analysis of secretome pipeline

3.1 Introduction

The cancer secretome is a valuable resource to identify diagnostic and prognostic biomarkers (Robinson *et al.*, 2019; Xue *et al.*, 2008). This is especially valuable in NSCLC, where there are no biomarkers available and proposed candidates are non-specific for NSCLC (Duffy & O'Byrne, 2018). Recent studies that have investigated the NSCLC secretome have primarily focused on advanced or metastatic LUAD, such as identifying erlotinib resistance, response prediction to cisplatin, or factors involved in metastatic self-sufficiency (Bosse *et al.*, 2016; Böttger *et al.*, 2019; Brady *et al.*, 2016). There is limited research on changes in the secretome that occur during LUAD malignant transformation, which may yield valuable biomarker candidates for early detection. The HBEC cell line has been used to model stepwise malignant transformation, employing key genetic alterations such as in *EGFR* and *TP53* (Sato *et al.*, 2006). In the previous Chapter, I generated a model of LUAD malignant transformation *in vitro*, consisting of HBEC stable cell lines and selected *EGFR* mutant NSCLC cell lines, and a serum-free mass spectrometry protocol optimized to identify secreted proteins. This Chapter applies the model and method developed in Chapter 2, collectively a pipeline, to identify and validate differentially expressed proteins during malignant transformation for further investigation as biomarkers.

3.2 Materials and Methods

3.2.1 Cell lines and standard culture conditions

Cell lines and culture conditions used in this Chapter are described in Section 2.2.1.

3.2.2 Secretome sample collection

Cell lines were seeded under standard culture conditions in a 6cm plate in triplicate and left to adhere overnight: 900 000 (HBEC); 1 000 000 (PC-9); 1 000 000 (H1975); 3 500 000 (HCC4006); 7 000 000 (HCC4011); 2 500 000 (H3255). Cells were then washed twice with DPBS and media was replaced with either supplement-free KSFM, containing 1% Penicillin-Streptomycin (HBEC) or serum-free RPMI-1640, containing 1% Pencillin-Streptomycin (*EGFR* mutant NSCLC). Cells were incubated under these conditions for 24 hours, in parallel with media-only plates that were prepared at this time. Conditioned media samples were centrifuged 1000 RPM for 5 minutes, at 4°C then filtered with a 0.45μM filter to remove cell debris. The complete volume of filtered conditioned media, 4mL, was concentrated to approximately 1/20th-1/25th of the original volume (150-200μL) with a Vivaspin Turbo 4 3kDa ultrafiltration unit by centrifuging at 3220xg, at 8°C. Concentrated media was buffer exchanged to 50mM HEPES buffer, pH 7.0, by centrifuging twice with 4mL HEPES, then once with 1mL HEPES to approximately 150-300uL under the same centrifuge conditions. Samples were stored at -80°C until processing for mass spectrometry.

3.2.3 Secretome MS/MS analysis

MS/MS sample preparation and analysis are outlined in Section 2.2.10 with the exceptions of an TMT 11-plex kit (Thermo Fisher Scientific) used for labeling, samples were run on an Orbitrap Eclipse^{TM} mass spectrometer, and only PSMs with FDR < 0.01 were retained in the analysis.

Technical replicates were pooled and run in the same TMT channel (1 cell line per channel). Further search settings can be found in Section A.2 (Figure S.5).

3.2.4 Data processing and statistical analysis

Raw signal intensity data was filtered, summarized to the peptide level, and aggregated to the protein level. Protein level data was then $\log_2$-transformed and median normalized by taking the median total signal and calculating a normalization factor for samples and media respectively, and imputed for missing values. Samples were contrasted against the appropriate media control on the $\log_2$ scale, and the intersection was analyzed to identify for enriched proteins. Proteins that were not observed in 2 or more technical replicates were removed from analysis. Prior to statistical testing, average fold changes were generated by subtracting the average $\log_2$ HBEC GFP;$p53^{wt}$ signal intensity from all samples. The moderated *t*-test statistical analysis was performed and adjusted *P*-values were calculated and adjusted with the Benjamini-Hochberg procedure, with the R package limma (Ritchie *et al.*, 2015) (version 3.50.0) , and results visualized with ggplot2 (Wickham, 2016) (version 3.3.5) in the R programming environment (version 4.0.5).

3.2.5 Bioinformatic analysis

Proteins of interest were queried in UniProt and GO databases as described in Section 2.2.11.

3.2.6 Gene expression data analysis

Z-score normalized Affymetrix gene expression data was sourced from Memorial Sloan Kettering Cancer Center (Chitale *et al.*, 2009) and analyzed with the Wilcoxon rank-sum test (Mann-Whitney *U* test) and adjusted *P*-values were calculated with the Benjamini-Hochberg procedure with base R functions in the R programming environment (version 4.0.5). Probes were annotated with the corresponding gene symbol with the R package hgu133a.db (Carlson, 2017a)

(version 3.2.3) and significant genes were filtered for optimal 1:1 probe:gene mapping with the R package jetset (Q. Li *et al.*, 2011) (version 3.4.0).

3.2.7 Survival analysis

Survival analysis of significant genes identified during Section 3.2.6 was performed on the NCBI GEO gene expression dataset GSE 31210, where probes were annotated with the R package hgu133plus2.db (Carlson, 2017b) (version 3.2.3). The Logrank test was performed, computing for median overall survival after applying a median split for gene expression.

3.3 Results

3.3.1 Confirmation of secreted proteins

After optimizing secretome collection conditions, I performed the secretome experiment consisting of HBEC and selected *EGFR* mutant NSCLC cell lines to investigate changes in secreted proteins between the normal, pre-malignant, and malignant states of the LUAD transformation model previously described. An overview of the secretome experiment is provided in Figure 3.1. Prior to performing differential protein expression analysis, I sought to verify the quality of my data and reduce any sources of variation during sample processing. This was performed with principal component analysis (PCA), a technique used to reduce dimensions of large datasets, but preserves the overall structure and variability inherent to a dataset (Jolliffe & Cadima, 2016) (Figure 3.2). Figure 3.2 shows that as expected of differing tumor or tissue origins, the cell lines and their technical replicates generally cluster together; cell lines also cluster away from media controls (Figure S.1). The exception to this was one PC-9 technical replicate, but the replicate was not sufficiently adjacent to the neighbouring HCC4006 and HBEC clusters to be considered an outlier. Interestingly, there were no distinct clusters within the HBEC group reflective of the *EGFR* or *TP53* alterations introduced in the previous Chapter.

This is likely due to PCA reflecting the entire scope of identified secreted proteins, which may not differ significantly in HBEC cell lines where one or two genetic changes have been introduced.

I then sought to confirm that secreted proteins were enriched in conditioned media collected. 1020 proteins were initially identified from the experiment output. Proteins enriched in the conditioned cell line media were confirmed by extracting the intersection of proteins between cell lines and media control. Of the 852 enriched proteins identified, previously identified proteins HSPG2, LAMA5, and AGRN were present (Hu *et al.*, 2016) (Figure 3.3). This validated the presence of secreted proteins in the conditioned media samples. EGFR was also enriched in conditioned media samples, reflective of EGF ligand secretion and associated signalling observed in NSCLC cells (Gazdar & Minna, 2008). I also validated the presence of secreted proteins within the enriched protein subset by performing cellular component analysis as done previously. As shown in Figure 3.3, secreted proteins comprise the second most abundant subcellular localization category. Cytoplasmic and membrane annotations comprised the most abundant and third most abundant subcellular localizations respectively. As mentioned previously, this could be attributed to reasons such as basal cell lysis and intracellular protein contamination, alternative pathways of secretion without a signal peptide, membrane protein ectodomain shedding, and the broad nature of UniProt annotations (Dowling & Clynes, 2011; Hayashida *et al.*, 2010; Rabouille, 2017; UniProt Consortium, 2022). An alternative reason for cytoplasmic proteins comprising the most abundant subcellular localization category may be due to the presence of EVs. While EV cargo can vary per type, cargo proteins include cytosolic and cytoskeleton proteins such as actin and annexin, and heat shock proteins (Meldolesi, 2018). EV cargo proteins may have been present in the conditioned media collected, contributing to the

abundance of cytoplasmic proteins. I then investigated whether EVs were potentially present and observed that common EV marker proteins CD9, CD81, but not CD63 were enriched (Bost *et al.*, 2022). This further suggested that EVs may contribute to the enrichment of cytoplasmic proteins found in the cellular component analysis.

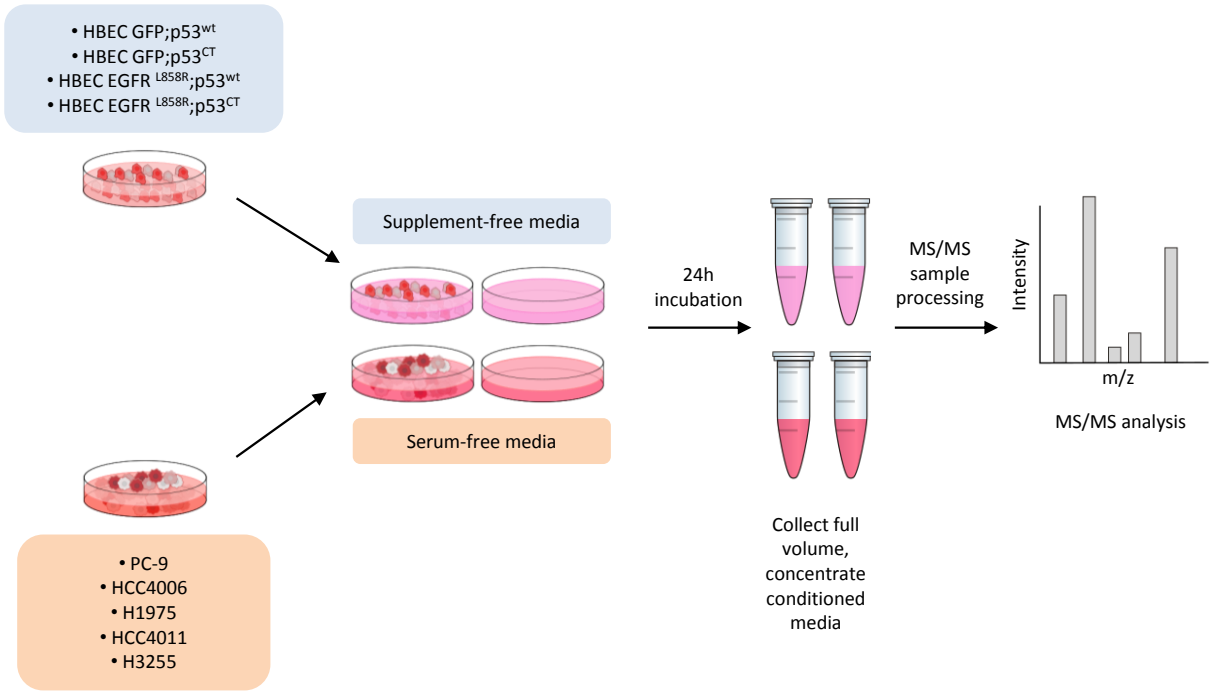


Figure 3.1 Schematic overview of secretome experiment.

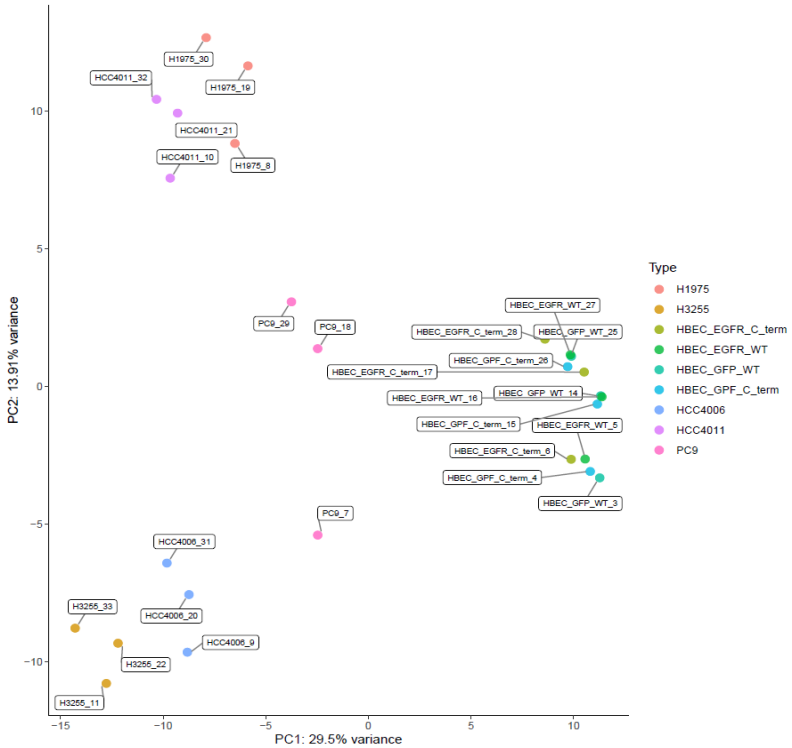
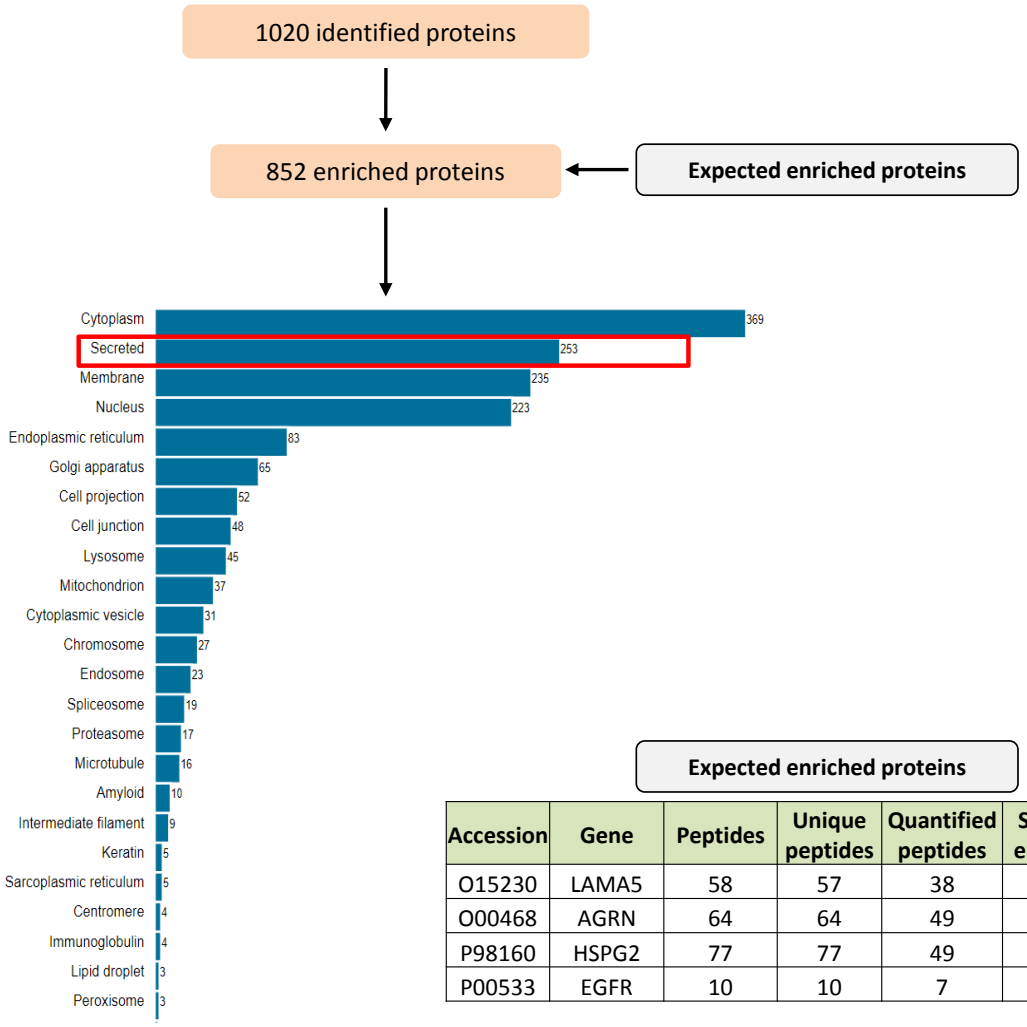


Figure 3.2 PCA of identified proteins (excluding media background controls).



Expected enriched proteins

Accession	Gene	Peptides	Unique peptides	Quantified peptides	Secretome enrichment
O15230	LAMA5	58	57	38	TRUE
O00468	AGRN	64	64	49	TRUE
P98160	HSPG2	77	77	49	TRUE
P00533	EGFR	10	10	7	TRUE

Figure 3.3 Cellular component analysis of identified proteins.
Overview of identified proteins, expected enriched proteins as seen previously, and cellular component analysis of enriched proteins identified (Hu *et al.*, 2016).

3.3.2 Differential protein expression analysis of pre-malignant and malignant states

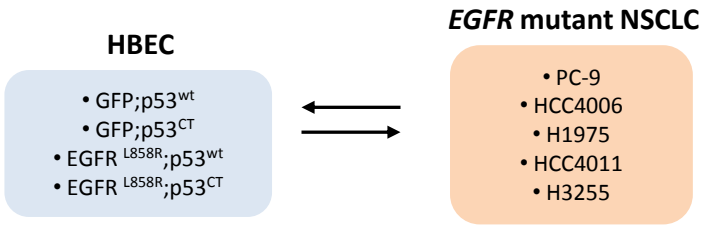
To identify differences in secreted proteins between pre-malignant and malignant states of my LUAD transformation model for further analysis and validation, I performed differential protein expression (DPE) analysis of *EGFR* mutant NSCLC cell lines relative to HBEC stable cell lines, assessing group-wise differences in $\log_2$ fold change. This analysis was done to compare the normal and pre-malignant stages of LUAD transformation to the malignant stage.

Figure 3.4 shows the comparison performed (Figure 3.4A) and the results of the DPE analysis (Figure 3.4B). 91 proteins were identified as differentially expressed, with more proteins underexpressed than overexpressed. As I was interested in the biological functions of these proteins, I performed GO functional analysis (Figure 3.5). The majority of biological processes identified consisted of immune functions and coagulation; of the immune functions represented, they consisted of MHC class I protein complex assembly, T cell mediated immunity, and the acute phase response. This suggests that there may be immunological changes during malignant transformation, as observed in the development of an immunosuppressive tumor microenvironment in *EGFR* mutant LUAD *in vivo* (Akbay *et al.*, 2013). To confirm that these findings were not solely statistically relevant, I reviewed the literature of the top 5 overexpressed and underexpressed proteins for potential biological relevance (Figure 3.4C). Two of the top overexpressed proteins, midkine (MDK) and NPC intracellular cholesterol transporter 2 (NPC2) were overexpressed in NSCLC patient serum and mouse LUAD plasma, respectively (Stern *et al.*, 2021; Taguchi *et al.*, 2011). This supports the idea that there may be biological relevance to the proteins identified from the DPE analysis.

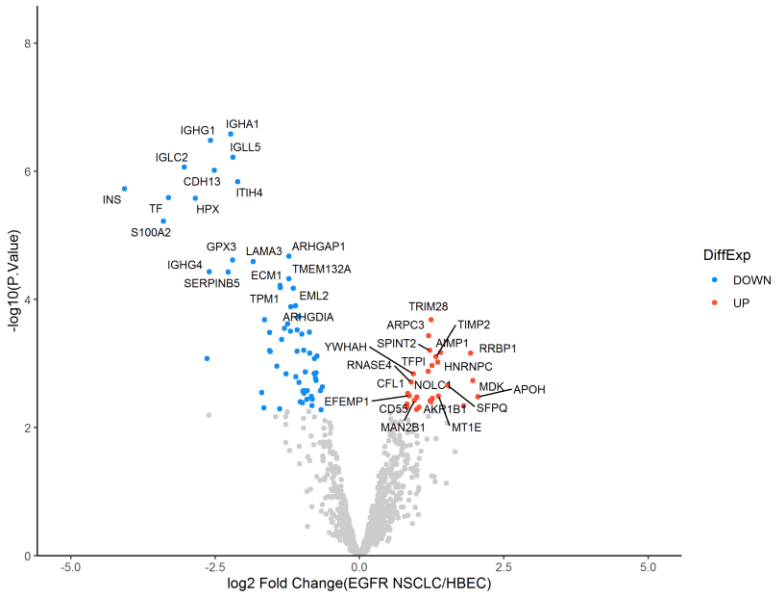
I next sought to confirm whether proteins identified in the DPE analysis were also differentially expressed in the stable HBEC cell lines, defined as the normal and pre-malignant states. I performed DPE analysis, where HBEC GFP; $p53^{CT}$, HBEC EGFR ^{L858R}; $p53^{wt}$, and HBEC EGFR ^{L858R}; $p53^{CT}$ were individually compared to HBEC GFP; $p53^{wt}$. A total of 4 differentially expressed proteins were identified: ATP-dependent 6-phosphofructokinase (platelet type) (PFKP), fibronectin (FN1), alpha-1-antichymotrypsin (SERPINA3), and serpin B7 (SERPINB7) (Figure S.2). SERPINA3 and SERPINB7 were observed in the initial DPE group-wise comparison, suggesting that these proteins may be differentially expressed during the stages

of malignant transformation. Interestingly, both proteins were underexpressed in the group-wise comparison. This suggests that there may be additional genetic changes during malignant transformation that alter the expression of SERPINA3 and SERPINB7 observed in the group-wise comparison. With the exception of PFKP, differentially expressed proteins were not exclusive to a specific HBEC stable cell line; most proteins were common to HBEC $EGFR^{L858R}$; $p53^{CT}$ (Figure S.3). This suggests that one or two genetic alterations may not significantly alter the HBEC secretome. It is also possible that the secretome of the pre-malignant state of my model is not discernibly different from the untransformed, normal state. If so, it would be more appropriate to identify changes in secreted proteins between the pre-malignant and malignant stages where a hypothetical patient would have invasive cancer; secretome profiles may be more discernible and result in more specific biomarkers.

A



B



C

Top 5 underexpressed proteins (by logFC)

Gene	logFC	adj.P.Val	Peptide count	Predicted secreted
INS	-4.07181	0.000227	2	Y
S100A2	-3.39714	0.000513	4	N
TF	-3.31143	0.000249	64	Y
IGLC2	-3.03245	0.000164	2	Y
HPX	-2.84123	0.000249	19	Y

Top 5 overexpressed proteins (by logFC)

Gene	logFC	adj.P.Val	Peptide count	Predicted secreted
APOH	2.048902	0.038475	2	Y
MDK	1.960715	0.027207	11	Y
RRBP1	1.923595	0.014399	7	N
NPC2	1.801741	0.045883	10	Y
SFPQ	1.525651	0.030639	2	N

Figure 3.4 Differential protein expression analysis of EGFR mutant NSCLC cell lines and HBEC stable cell lines (group-wise comparison).
(A) Schematic overview of the cell lines compared. (B) Volcano plot of differentially expressed proteins (adj p val < 0.05, minimum absolute $\log_2$ fold change > 0.6). The top 20 proteins sorted by p value are labeled. (C) Top 5 underexpressed and overexpressed proteins, sorted by $\log_2$ fold change.

GO biological process complete	Fold Enrichment	FDR
peptide antigen assembly with MHC class I protein complex	> 100	3.30E-02
↳immune system process	2.91	1.81E-05
↳cellular component organization	1.90	5.86E-04
↳cellular component organization or biogenesis	1.83	1.19E-03
↳cellular process	1.27	2.80E-04
↳MHC class I protein complex assembly	> 100	3.27E-02
sequestering of calcium ion	87.64	4.17E-02
↳localization	1.67	4.30E-02
↳regulation of biological quality	1.89	2.56E-02
↳biological regulation	1.42	9.12E-05
protein folding in endoplasmic reticulum	79.67	1.79E-04
↳protein folding	6.14	4.68E-02
positive regulation of blood coagulation	28.58	2.54E-02
↳positive regulation of coagulation	27.39	2.79E-02
↳regulation of coagulation	12.34	3.89E-02
↳regulation of multicellular organismal process	2.43	8.26E-04
↳regulation of biological process	1.49	1.81E-05
↳positive regulation of multicellular organismal process	3.22	4.46E-04
↳positive regulation of biological process	1.85	1.31E-04
↳positive regulation of hemostasis	28.58	2.56E-02
↳regulation of hemostasis	12.89	3.46E-02
↳positive regulation of response to wounding	12.89	3.48E-02
↳regulation of response to stress	2.99	4.76E-03
↳regulation of response to stimulus	2.24	9.08E-05
↳positive regulation of response to stimulus	3.26	1.47E-06
↳regulation of blood coagulation	13.28	3.19E-02
↳regulation of response to external stimulus	4.14	1.20E-04
T cell mediated immunity	22.67	4.09E-02
↳lymphocyte mediated immunity	9.92	1.70E-05
↳leukocyte mediated immunity	8.34	6.37E-05
↳immune effector process	6.36	8.33E-05
↳adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	9.45	2.18E-05
↳adaptive immune response	3.74	2.30E-02
↳immune response	2.93	1.37E-03
↳response to stimulus	1.77	6.42E-06
acute-phase response	20.87	8.57E-03
↳acute inflammatory response	16.43	6.33E-04
↳inflammatory response	5.18	8.84E-04
↳defense response	4.05	1.37E-06
↳response to stress	2.89	7.47E-08

Figure 3.5 Functional analysis of differentially expressed proteins between *EGFR* mutant NSCLC cell lines and HBEC stable cell lines (GO Biological Process, complete).
Terms are sorted by hierarchy (FDR < 0.05).

3.3.3 Differential protein expression analysis of normal and malignant states

As group comparisons may not capture the full scope of differentially expressed proteins, I performed DPE analysis by comparing each *EGFR* mutant NSCLC cell line individually to HBEC GFP;$p53^{wt}$ as a representative of the "normal" baseline state. Figure 3.6 illustrates differentially expressed proteins identified from each cell line. In contrast to the group-wise comparison performed in Figure 3.4, the number of differentially expressed proteins ranged from 119 to 316, compared to 91 previously. The difference in differentially expressed proteins suggests that comparing *EGFR* mutant NSCLC cell lines individually to the most basal state of transformation, HBEC GFP;$p53^{wt}$, may be more conducive to identifying more protein candidates for further analysis and validation. This may be due to additional genetic alterations unique to each NSCLC cell line that influence the secretome profile and may be heterogenous when comparing collectively as a group.

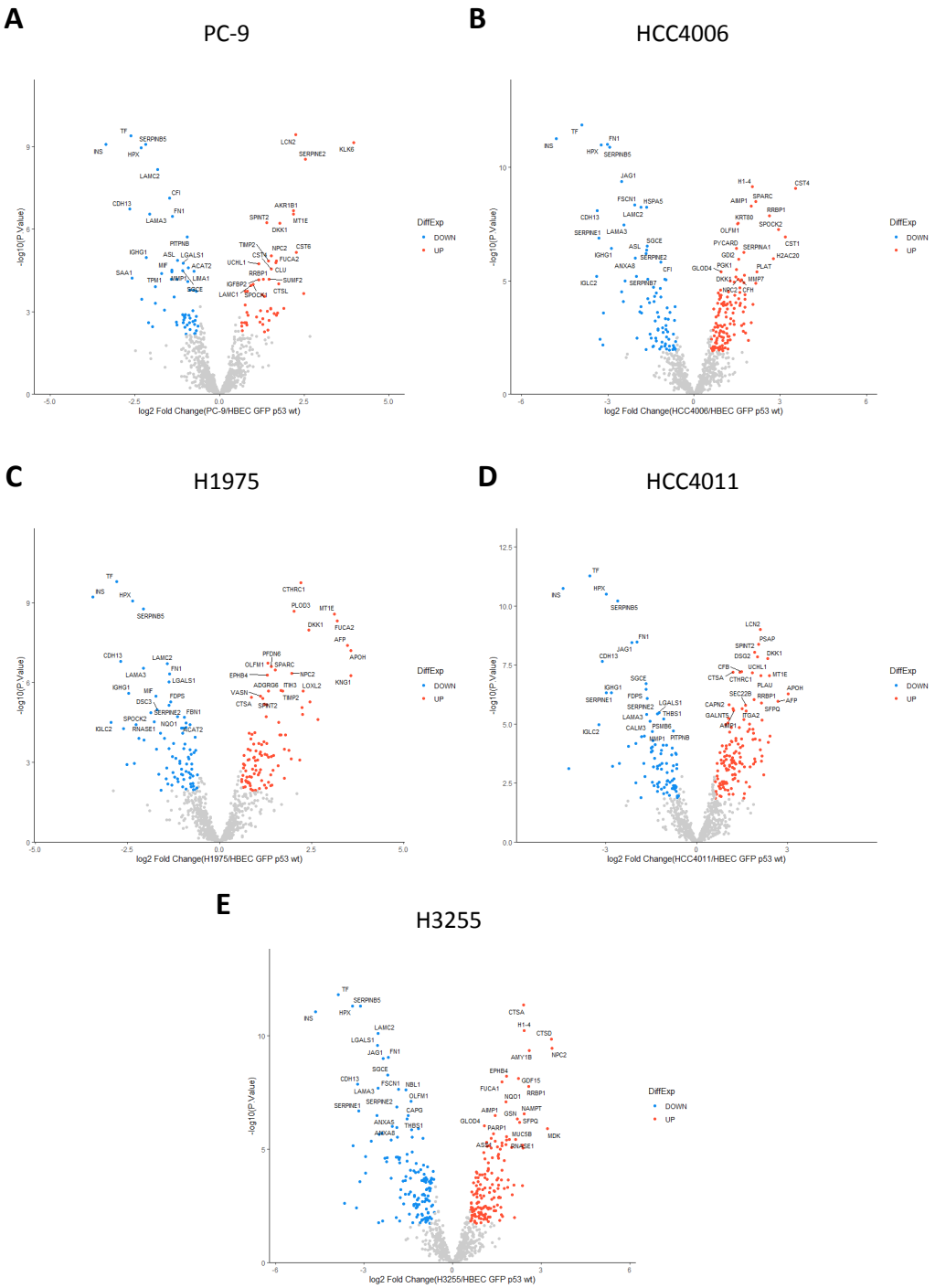


Figure 3.6 Volcano plots of differentially expressed proteins of each comparison made between *EGFR* mutant NSCLC cell line and HBEC GFP;p53^{wt}.
Adj p val < 0.05, minimum absolute $\log_2$ fold change > 0.6. The top 20 proteins sorted by p value are labeled. (A) PC-9. (B) HCC4006. (C) H1975. (D) HCC4011. (E) H3255.

3.3.4 Validation of differentially expressed proteins with *EGFR* mutant tumor patient cohort data

To validate my findings and identify protein candidates for further analysis, I performed filtering and differential gene expression (DGE) analysis of a subset of proteins from the DPE analysis with *EGFR* mutant NSCLC cell lines and HBEC GFP;$p53^{wt}$. Figure 3.7 illustrates the filtering and analysis process performed. In this analysis, I used transcriptomic data for validation as there was no applicable *EGFR* mutant NSCLC proteomic data available. I first identified differentially expressed proteins common to 3 or more *EGFR* mutant NSCLC cell lines, and then removed proteins with different $\log_2$ fold change directions to ensure consistency. I then performed DGE analysis on 130 proteins with Z-score normalized *EGFR* mutant LUAD tumor Affymetrix patient microarray data, generated from 199 primary LUAD tumors collected at Memorial Sloan Kettering Cancer Center; analysis was done comparing 39 *EGFR* mutant and 154 wild-type tumors (Chitale *et al.*, 2009). 22 genes were found to be differentially expressed between *EGFR* mutant and wild-type tumors, and further filtered for optimal probe mapping with the Jetset scoring method (Q. Li *et al.*, 2011). A total of 16 differentially expressed genes were found, confirming previous DPE analysis findings (Table 3.1). However, the change in relative abundance of the 16 protein candidates was not reflected in the HBEC stable cell lines expressing *EGFR L858R*, *p53* CT, or both, suggesting that their secretomes are not distinct from HBEC GFP;$p53^{wt}$ (Figure S.4).

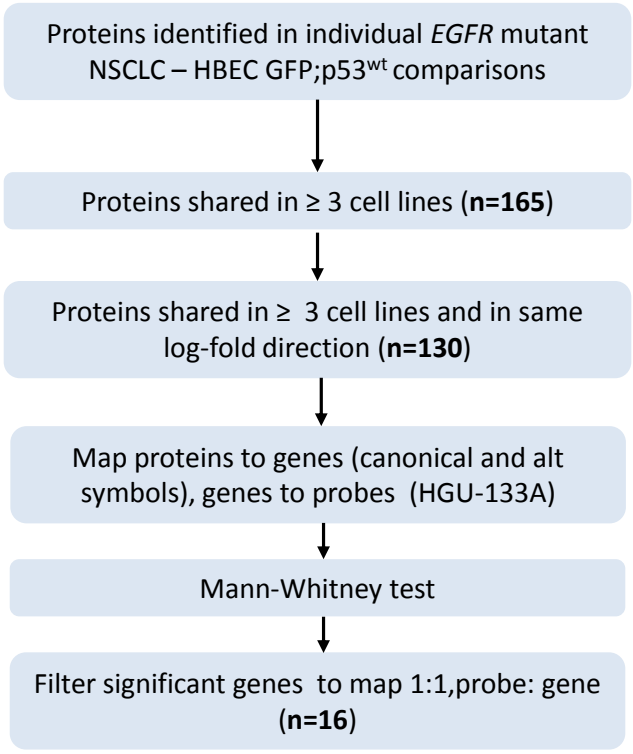


Figure 3.7 Protein candidate filtering pipeline.
Schematic describing the filtering process performed to identify protein candidates for further analysis and validation. Mann-Whitney test was performed on genes probe-wise, where each probe was treated as an individual gene and significant genes were determined with an adjusted p value < 0.05 (Benjamini-Hochberg adjustment). Differentially expressed genes were further filtered for optimal probe-to-gene mapping with the Jetset scoring method **(**Q. Li *et al.*, 2011).

Table 3.1 Protein candidates identified from filtering pipeline with patient microarray data.
Proteins are sorted by adjusted p value (Benjamini Hochberg-adjusted).

Accession	Gene	Peptides	Unique peptides	Quantified peptides	adj.P.Val
Q08380	LGALS3BP	23	23	20	9.5519E-05
P61916	NPC2	10	10	10	9.5519E-05
O43291	SPINT2	2	2	1	0.001714715
P06733	ENO1	30	28	27	0.003759408
P21741	MDK	12	12	11	0.003759408
P63000	RAC1	3	2	1	0.005590461
Q01813	PFKP	2	2	1	0.007683377
P31431	SDC4	7	7	5	0.012165019
Q99988	GDF15	6	6	4	0.020488018
P02749	APOH	2	2	2	0.02368962
P36952	SERPINB5	14	14	8	0.027201045
P10586	PTPRF	19	19	15	0.031704624
P07339	CTSD	20	20	20	0.034872449
O94907	DKK1	7	7	7	0.034872449
P03956	MMP1	2	2	2	0.035149978
Q9BY76	ANGPTL4	6	6	4	0.03699355

3.3.5 Survival analysis of validated secreted proteins in EGFR mutant tumor patient cohort data

I then examined if there was a survival difference with differential expression of the 16 protein candidates (Table 3.1) on a subset of *EGFR* mutant and wild-type patient tumor microarray data (NCBI GEO GSE 31210). I performed Kaplan-Meier analysis on 125 *EGFR* mutant and 68 *EGFR* wild-type tumors, comparing overall survival duration between patients with high and low expression of the genes that encode for the 16 proteins. Patients with *EGFR* mutant LUAD and high expression of *ENO1*, *PFKP*, *RAC1*, and *SPINT2* were found to have lower overall survival compared to those with low expression (Figure 3.8). Interestingly, these observations were not seen in *EGFR* wild-type tumors, indicating that the survival difference was specific to *EGFR* mutant LUAD progression. Another observation was that relative PFKP expression identified in Section 3.3.4 was in the negative log-fold direction, yet lower overall survival was associated with high *PKFP* expression. This may be due to translational changes or degradation that contributed to the discrepancy between gene expression and protein expression (Cox & Mann, 2007). Patients with high expression levels of *MDK* were also found to have overall lower survival, but this was not exclusive to *EGFR* mutant tumors (Figure 3.8). This is consistent with previous findings where elevated serum MDK was observed in NSCLC patients, regardless of histological and molecular subtype (Stern *et al.*, 2021).

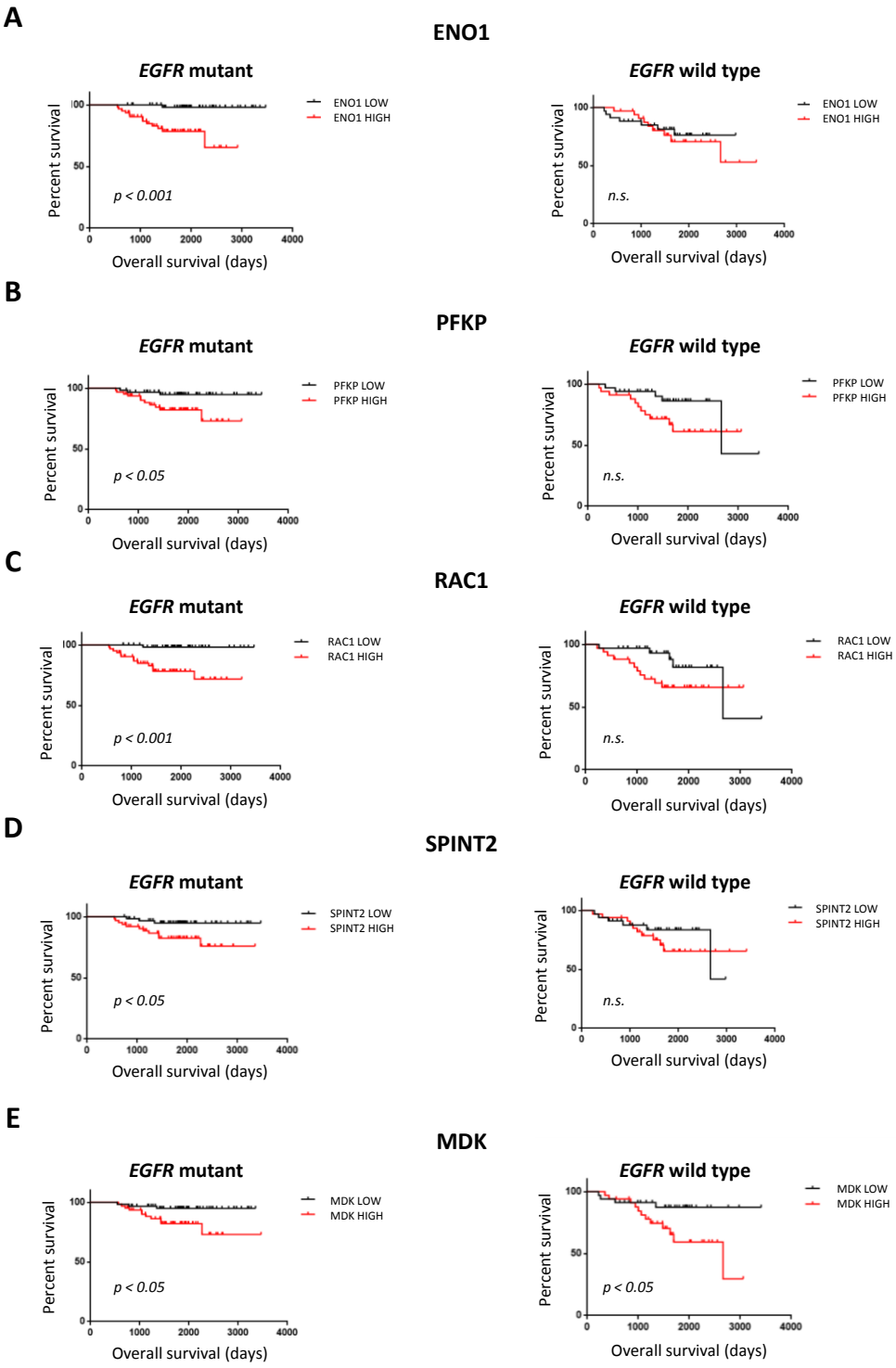


Figure 3.8 Kaplan-Meier plots of overall patient survival of *EGFR* mutant and wild type tumors, with selected protein candidate expression data (log rank, median split).

3.4 Discussion

The investigation of the secretome during LUAD progression is relatively unknown (Robinson *et al.*, 2019). I previously developed a method to profile secreted proteins during LUAD malignant transformation and applied it in this Chapter (Figure 3.1). Initial observations of the proteins identified indicated that this method could enrich for secreted proteins, including ones identified in previous studies, and that the quality of collected data accurately represented the distinct tissue origins of the cell lines used (Figure 3.2, Figure 3.3). To identify potential biomarker candidates, I initially performed DPE analysis of the pre-malignant and malignant states of my transformation model. 91 differentially expressed proteins were identified, where 2 of the top overexpressed proteins midkine (MDK) and intracellular cholesterol transporter 2 (NPC2) have been observed in patient samples, suggesting that these findings are biologically relevant (Figure 3.4) (Stern *et al.*, 2021; Taguchi *et al.*, 2011). However, $\log_2$ fold changes in the secretome may be more subtle than a value of 0.6, and more liberal fold change parameters may yield additional differentially expressed proteins. Interestingly, GO functional analysis revealed that most of the biological processes represented consisted of immune functions such as MHC class I protein complex assembly and T cell mediated immunity, and coagulation (Figure 3.5). This suggests that during LUAD transformation or progression, a tumor may secrete molecules such as cytokines to alter the microenvironment or its immunological phenotype. The overrepresentation of immune functions is correlated with previous findings where *EGFR* mutations were associated with poor responses to PD-1/PD-L1 immune checkpoint inhibitors in NSCLC patients and immunosuppression *in vivo* (Akbay *et al.*, 2013; Gainor *et al.*, 2016).

As I was interested in secreted protein changes during transformation, I also performed DPE analysis within the HBEC stable cell line group, comparing cell lines expressing *EGFR*

L858R, *TP53* CT, or both, to HBEC GFP; $p53^{wt}$, the normal state of my transformation model. Unfortunately, a total of 4 proteins were differentially expressed regardless of $\log_2$ fold change parameters set; most proteins were common among comparisons made (Figure S.3). This suggests that more genetic alterations may be needed to influence changes in secreted proteins. Moreover, further validation of the pre-malignant state is needed, such as performing anchorage-independent growth assays *in vitro* or observing tumor growth *in vivo* (Ahuja *et al.*, 2005; Sato *et al.*, 2013). Alternatively, the pre-malignant secretome may not be discernable from the normal state. An example of this is in urinary bladder cancer, where changes in the biomarker candidate pro-u-PA were not statistically different until late tumor stage (Lin *et al.*, 2006). Thus, profiling secretome from the malignant stage, the equivalent of late stage or invasive tumors, may be more suitable for biomarker identification.

As group comparisons may not resolve differentially expressed proteins unique to a specific tumor molecular profile, I performed DPE analysis on individual *EGFR* mutant NSCLC cell lines compared to HBEC GFP;$p53^{wt}$ (Figure 3.6). Given each comparison yielded 100 to 300 differentially expressed proteins, I sought to identify potential biomarker candidates by first filtering for proteins common to three or more *EGFR* mutant NSCLC cell line comparison made, confirming that $\log_2$ fold change direction was consistent, and validating differential expression with patient *EGFR* mutant tumor microarray data that included optimal probe to gene mapping for additional stringency (Figure 3.7) (Q. Li *et al.*, 2011). To maximize the number of protein candidates, proteins were not filtered for a "secreted" annotation; this resulted in a list of 16 potential biomarker candidates for further analysis (Table 3.1). To investigate their potential as prognostic or diagnostic biomarkers, I compared whether there was a difference in overall patient survival between high and low expression of the genes that encode for these proteins (Figure

3.8). 5 proteins were found to have significantly different overall patient survival, where patients with *EGFR* mutant tumors and high expression of *ENO1, PFKP, RAC1,* and *SPINT2* had lower overall survival compared to those with wild-type tumors. This exclusivity suggests that these proteins could be further validated for prognostic biomarker use, specifically in *EGFR* mutant LUAD. In contrast, high expression of *MDK* was associated with lower overall patient survival regardless of *EGFR* mutation status (Figure 3.8). This implies that it may be useful as a general NSCLC prognostic biomarker, and it has been explored independently and as part of an 8 serum biomarker panel (Birse *et al.*, 2015; Stern *et al.*, 2021).

Further literature analysis of the 5 proteins identified from the filtering pipeline and survival analysis was hampered by the limited information in their role in LUAD initiation, progression, and biomarker association. ENO1 is involved in pyruvate synthesis, and has been observed to promote tumor migration and metastasis *in vivo* through the hepatocyte growth factor signaling pathway (Hsiao *et al.*, 2013; H.-J. Li *et al.*, 2021). PFKP is involved in glycolysis, and its downregulation was demonstrated to decrease relative glucose metabolism, proliferation, and colony formation *in vitro*. Its expression was also correlated with tumor size and survival, suggesting potential as a prognostic biomarker (Shen *et al.*, 2020). RAC1 is a Rho family GTPase that has been shown to mediate EGFR-activated migration and growth *in vitro* and gefitinib-resistant tumor growth *in vivo*, and in a KRAS model of LUAD, was shown to be required for tumor formation *in vivo* (Kaneto *et al.*, 2014; Kissil *et al.*, 2007). SPINT2 is a serine protease inhibitor where reduced expression is involved in STYK1 driven proliferation, invasion, and migration *in vitro* and *in vivo* (Ma *et al.*, 2019). MDK is a heparin-binding growth factor observed to promote cell growth and survival *in vitro*, and tumor growth *in vivo* in an oncogenic KRAS LUAD and SCC cell lines (Hao *et al.*, 2013).

3.5 Conclusion

This Chapter summarizes the application of the serum-free mass spectrometry protocol on a model of LUAD malignant transformation to identify biomarker candidates in early NSCLC. This protocol was shown to enrich for secreted proteins and identify differentially expressed proteins in normal, pre-malignant, and malignant stages of transformation *in vitro*. When coupled with a filtering and clinical validation pipeline, I identified 4 *EGFR* mutant-specific and 1 *EGFR* mutant non-specific biomarker candidates that may be further validated for prognostic or diagnostic use. Overall, this pipeline consisting of mass spectrometry protocol and LUAD transformation model enables investigation of changes in the secretome during LUAD malignant transformation that may be applied for clinical use.

Chapter 4: Conclusion

4.1 Research summary

As lung cancer is detected in advanced and metastatic stages, earlier detection is vital to long-term patient survival; biomarkers have been explored as a complement to existing detection methods with none available for NSCLC (Duffy & O'Byrne, 2018; Herbst *et al.*, 2018). *EGFR* mutant LUAD is in specific need of biomarkers as this mutation is frequently observed in never smoker LUAD, where LDCT screening guidelines have yet to be established (Lam, 2019; Rudin *et al.*, 2009). A recent retrospective LDCT study on a 12,176 never smoker cohort subset diagnosed lung cancer at a detection rate of 0.45%, with most cases at stages 0 or 1 and greater overall survival compared to the ever smoker subset (Kang *et al.*, 2019). The rate of diagnosis was comparably less than the NSLT detection rate of 1.0% derived from ever smokers, however comparable with a previous never smoker study detection rate of 0.48% (Kang *et al.*, 2019; Sone *et al.*, 2007; The National Lung Screening Trial Research Team, 2011). These findings suggest potential detection benefits for never smoker LUAD, yet screening will benefit from the addition of biomarker-based detection. The cancer secretome is considered a valuable biomarker source, and the study of secreted protein changes during malignant transformation may yield candidates for prognostic and diagnostic use (Robinson *et al.*, 2019). To identify potential biomarkers, I established a stepwise LUAD model of *EGFR* malignant transformation and serum-free mass spectrometry protocol to profile the secretome. The protocol was confirmed to enrich for secreted proteins, while minimizing viability loss under nutrient deprivation. Through a filtering and validation pipeline, I identified 4 *EGFR* mutant-specific and 1 *EGFR* mutant non-specific biomarker candidates for future validation whose gene expression was associated with

differences in overall patient survival. This implies that ENO1, PFKP, RAC1, and SPINT2 may be possible biomarkers for *EGFR* mutant LUAD, while MDK may be a potential general LUAD biomarker. As secretome changes may not be discernible until late stage tumor progression, these candidates may be most applicable to patients with late stage or invasive tumors, as seen in urinary bladder cancer (Lin *et al.*, 2006). Given the differential expression of each protein with cell lines derived from late stage or metastatic LUAD, I hypothesize that the relative expression pattern of these proteins is consistent during malignant transformation. I further hypothesize that ENO1, PFKP, RAC1, and SPINT2 may be important in the progression of *EGFR* mutant LUAD. As differential protein expression may impact LUAD progression, it is important to investigate the potential functions of these proteins during malignant transformation in tandem with their potential biomarker utility. In this work, I hope to further understanding of these proteins as clinical LUAD biomarkers and their role in LUAD development.

4.2 Research limitations

Although this research identified potential biomarker candidates for future validation, there are notable limitations to the research performed. Firstly, the HBEC stages of my model were not validated *in vitro* or *in vivo* for changes in transformation capacity, such as anchorage-independent growth or colony formation *in vivo* (Ahuja *et al.*, 2005). This limited the accuracy of modeling stepwise transformation observed in clinical LUAD, and thus the accuracy of the secretome during the pre-malignant stage (Inamura, 2018). Secondly, while alterations in *EGFR* and *TP53* are observed in *EGFR* mutant LUAD progression there may be other genetic alterations necessary to fully promote malignant transformation; this has been observed previously with anchorage-independent growth (Sato *et al.*, 2006). Thirdly, in the effort to minimize the effect of serum-free proteins, the limited incubation period under serum-free

conditions may not fully quantify less abundant proteins (Brandi *et al.*, 2018). Finally, there are inherent limitations to mass spectrometry technology. Two limitations are mass spectrometer sensitivity and run reproducibility (Nilsson *et al.*, 2010). While commercial mass spectrometers can detect proteins at the femtomole or attomole range, they may not accurately capture all proteins present in a sample due to differing peptide ionization efficiencies that influence protein abundance (Ankney *et al.*, 2018; Nilsson *et al.*, 2010). This is further exacerbated with low abundance proteins, which this research sought to identify in further detail (Nilsson *et al.*, 2010). Mass spectrometers are also prone to run variability, where repeated analyses of a protein mixture may not identify the exact same proteins (Nilsson *et al.*, 2010). This is due to the nature of mass spectrometer MS/MS scans, which are performed with computer algorithms and are prone to undersampling, where a different subset of proteins in a sample may be used for identification thus resulting in run variation (Nilsson *et al.*, 2010; H. Wang *et al.*, 2010). Due to these two limitations, the full scope of the secretome may not have been fully or accurately quantified.

4.3 Future directions

This thesis initially identified potential secreted biomarker candidates during LUAD malignant transformation; however, further work is needed to validate these findings and investigate their potential role in LUAD progression. The first set of experiments to be performed will be to evaluate the extent of transformation in my HBEC stable cell lines. The laboratory will assess this by performing anchorage-independent growth assays *in vitro* and mouse xenograft growth assays *in vivo*, as done previously (Sato *et al.*, 2006). This will enable determination of additional genetic alterations needed to represent the pre-malignant stage of my model.

The laboratory will then validate protein expression and secretion *in vitro*. To do this, they will perform ELISA for the selected protein candidates on conditioned media and cell lysates previously collected in parallel. In theory, secreted proteins should be more abundant in conditioned media compared to the cell lysate. By performing an ELISA, this will confirm that these proteins are not due to cytoplasmic contamination during secretome sample collection or stress under nutrient conditions (Brandi *et al.*, 2018). If ELISA antibodies are unavailable, the laboratory will perform Western blot analysis on the selected protein candidates and confirm their secretion with a secretion inhibitor such as Brefeldin A or monensin, which inhibit transport between the ER and Golgi apparatus or within the Golgi respectively(Chardin & McCormick, 1999; Nylander & Kalies, 1999). However, subtle differences in expression may not be resolvable by Western blot and loading controls are limited to non-secreted housekeeping genes (Collins *et al.*, 2015; X.-D. Wang *et al.*, 2019).

The laboratory will also validate the protein expression changes of ENO1, PFKP, RAC1, and SPINT2 in clinical patient samples to further assess relative changes during LUAD progression, and their potential as clinical biomarkers. This will be done by performing ELISA on collected BC Cancer Research Institute patient BAL samples. BAL composition includes immune cells such as macrophages and neutrophils, and proteins from epithelial, immune, and malignant tumor cells, and is highly specific for lung cancer diagnosis; these features make it a suitable medium to assess potential patterns of protein expression during LUAD progression (Carvalho *et al.*, 2017; Hmmier *et al.*, 2017; Stanzel, 2013). If these proteins show a consistent pattern of abundance change, such as a decrease from the stages of AAH to invasive carcinoma, this may suggest further investigation for biomarker use. However, patient samples per clinicopathological stage may be limited, thus restricting the scope of analysis. Further, BAL is

noted to have a low sensitivity range, where studies have identified LUAD at 11.6% and 29% (Bezel *et al.*, 2016; Labbé *et al.*, 2015). If BAL samples are limited, blood serum would be used as an alternative as it is easier to collect; however, highly abundant proteins such as albumin may make validation difficult due to the resulting high dynamic range (Hmmier *et al.*, 2017).

Finally, as differential protein expression of these candidates may affect LUAD progression, the laboratory will perform functional validation to study their potential role. As we are interested in *EGFR* mutant LUAD biomarker applications, ENO1, PFKP, RAC1, and SPINT2 will be further investigated for their function in LUAD progression. This will be done by generating stable overexpression or CRISPR knock-out HBEC cell lines, in tandem with *EGFR* and *TP53* alterations, and confirmed with Western blot. Potential roles for these proteins will be investigated with assays such as proliferation and migration assays, as some have been implicated in tumor growth and invasion in cancer cell lines (Kaneto *et al.*, 2014; H.-J. Li *et al.*, 2021; Ma *et al.*, 2019; Shen *et al.*, 2020). Transformation capacity is also an aspect of LUAD progression; the laboratory will also perform anchorage-independent growth assays of the stable HBEC cell lines to assess the influence of these proteins in transformation.

A potential avenue for exploration would be further optimization of the secretome collection method. Recent studies have reported the use of proximity biotinylation, where secreted proteins are biotinylated in the ER lumen, cytosol, or lumen-oriented membrane to capture conventional and unconventional secreted proteins (Liu *et al.*, 2021; Wei *et al.*, 2021). This method would circumvent the limitation of serum starvation conditions for MS/MS analysis; however, cytoplasmic protein contamination is still possible during sample collection (Xue *et al.*, 2008).

As the goal of this research was to identify biomarker candidates for lung adenocarcinoma detection, successfully validated biomarker candidates will require further investigation to identify a suitable screening population. Current screening standards for LDCT lung cancer screening are recommended for adults aged 50 to 80 years who currently smoke and have a 20 pack year history or previous smokers that have quit within 15 years (US Preventive Services Task Force *et al.*, 2021). As suitable screening standards for non-smokers are unclear, further investigation would best be suited to individuals that are recommended to undergo LDCT screening (Lam, 2019). There are two areas where the identified biomarker candidates can be implemented. The first area is as a complementary approach in tandem with LDCT. LDCT is still prone to false positives with criteria that are based on nodule size and volume, and a complementary approach with biomarkers can validate findings (Seijo *et al.*, 2019). The second applicable area would be as a post-LDCT risk assessment to determine the need for further diagnostic tests such as biopsy (Oudkerk *et al.*, 2021; Seijo *et al.*, 2019). LDCT screening can identify indeterminate pulmonary nodules, which are often benign yet would benefit from biomarker panel validation to confirm whether benign or malignant (Massion & Walker, 2014). This would be especially useful for nodules found during annual screening, which tend to be too small to biopsy (Seijo *et al.*, 2019).

Overall, this thesis provides support in the characterization of secreted proteins during *EGFR*-driven malignant transformation in LUAD and future avenues for validation and investigation of the biomarker candidates identified.

References

Ahmadiankia, N. (2020). In vitro and in vivo studies of cancer cell behavior under nutrient deprivation. *Cell Biology International*, *44*(8), 1588–1597. https://doi.org/10.1002/cbin.11368

Ahuja, D., Sáenz-Robles, M. T., & Pipas, J. M. (2005). SV40 large T antigen targets multiple cellular pathways to elicit cellular transformation. *Oncogene*, *24*(52), 7729–7745. https://doi.org/10.1038/sj.onc.1209046

Akbay, E. A., Koyama, S., Carretero, J., Altabef, A., Tchaicha, J. H., Christensen, C. L., Mikse, O. R., Cherniack, A. D., Beauchamp, E. M., Pugh, T. J., Wilkerson, M. D., Fecci, P. E., Butaney, M., Reibel, J. B., Soucheray, M., Cohoon, T. J., Janne, P. A., Meyerson, M., Hayes, D. N., … Wong, K.-K. (2013). Activation of the PD-1 Pathway Contributes to Immune Escape in EGFR-Driven Lung Tumors. *Cancer Discovery*, *3*(12), 1355–1363. https://doi.org/10.1158/2159-8290.CD-13-0310

Alexander, M., Kim, S. Y., & Cheng, H. (2020). Update 2020: Management of Non-Small Cell Lung Cancer. *Lung*, *198*(6), 897–907. https://doi.org/10.1007/s00408-020-00407-5

Ankney, J. A., Muneer, A., & Chen, X. (2018). Relative and Absolute Quantitation in Mass Spectrometry–Based Proteomics. *Annual Review of Analytical Chemistry*, *11*(1), 49–77. https://doi.org/10.1146/annurev-anchem-061516-045357

Arbour, K. C., & Riely, G. J. (2019). Systemic Therapy for Locally Advanced and Metastatic Non–Small Cell Lung Cancer: A Review. *JAMA*, *322*(8), 764. https://doi.org/10.1001/jama.2019.11058

Bade, B. C., & Dela Cruz, C. S. (2020). Lung Cancer 2020. *Clinics in Chest Medicine*, *41*(1), 1–24. https://doi.org/10.1016/j.ccm.2019.10.001

Basisty, N., Kale, A., Jeon, O. H., Kuehnemann, C., Payne, T., Rao, C., Holtz, A., Shah, S., Sharma, V., Ferrucci, L., Campisi, J., & Schilling, B. (2020). A proteomic atlas of senescence-associated secretomes for aging biomarker development. *PLOS Biology*, *18*(1), e3000599. https://doi.org/10.1371/journal.pbio.3000599

Bezel, P., Tischler, V., Robinson, C., Baumueller, S., Bode-Lesniewska, B., Kohler, M., Freitag, L., & Franzen, D. (2016). Diagnostic Value of Bronchoalveolar Lavage for Diagnosis of Suspected Peripheral Lung Cancer. *Clinical Lung Cancer*, *17*(5), e151–e156. https://doi.org/10.1016/j.cllc.2015.12.012

Birse, C. E., Lagier, R. J., FitzHugh, W., Pass, H. I., Rom, W. N., Edell, E. S., Bungum, A. O., Maldonado, F., Jett, J. R., Mesri, M., Sult, E., Joseloff, E., Li, A., Heidbrink, J., Dhariwal, G., Danis, C., Tomic, J. L., Bruce, R. J., Moore, P. A., … Ruben, S. M. (2015). Blood-based lung cancer biomarkers identified through proteomic discovery in cancer tissues, cell lines and conditioned medium. *Clinical Proteomics*, *12*(1), 18. https://doi.org/10.1186/s12014-015-9090-9

Bosse, K., Haneder, S., Arlt, C., Ihling, C. H., Seufferlein, T., & Sinz, A. (2016). Mass spectrometry-based secretome analysis of non-small cell lung cancer cell lines. *PROTEOMICS*, *16*(21), 2801–2814. https://doi.org/10.1002/pmic.201600297

Bost, J. P., Saher, O., Hagey, D., Mamand, D. R., Liang, X., Zheng, W., Corso, G., Gustafsson, O., Görgens, A., Smith, C. E., Zain, R., El Andaloussi, S., & Gupta, D. (2022). Growth Media Conditions Influence the Secretion Route and Release Levels of Engineered Extracellular

Vesicles. *Advanced Healthcare Materials*, *11*(5), 2101658. https://doi.org/10.1002/adhm.202101658
Böttger, F., Schaaij-Visser, T. B., de Reus, I., Piersma, S. R., Pham, T. V., Nagel, R., Brakenhoff, R. H., Thunnissen, E., Smit, E. F., & Jimenez, C. R. (2019). Proteome analysis of non-small cell lung cancer cell line secretomes and patient sputum reveals biofluid biomarker candidates for cisplatin response prediction. *Journal of Proteomics*, *196*, 106–119. https://doi.org/10.1016/j.jprot.2019.01.018
Brady, J. J., Chuang, C.-H., Greenside, P. G., Rogers, Z. N., Murray, C. W., Caswell, D. R., Hartmann, U., Connolly, A. J., Sweet-Cordero, E. A., Kundaje, A., & Winslow, M. M. (2016). An Arntl2-Driven Secretome Enables Lung Adenocarcinoma Metastatic Self-Sufficiency. *Cancer Cell*, *29*(5), 697–710. https://doi.org/10.1016/j.ccell.2016.03.003
Brandi, J., Manfredi, M., Speziali, G., Gosetti, F., Marengo, E., & Cecconi, D. (2018). Proteomic approaches to decipher cancer cell secretome. *Seminars in Cell & Developmental Biology*, *78*, 93–101. https://doi.org/10.1016/j.semcdb.2017.06.030
Brenes, A., Hukelmann, J., Bensaddek, D., & Lamond, A. I. (2019). Multibatch TMT Reveals False Positives, Batch Effects and Missing Values. *Molecular & Cellular Proteomics*, *18*(10), 1967–1980. https://doi.org/10.1074/mcp.RA119.001472
Brown, K. J., Seol, H., Pillai, D. K., Sankoorikal, B.-J., Formolo, C. A., Mac, J., Edwards, N. J., Rose, M. C., & Hathout, Y. (2013). The human secretome atlas initiative: Implications in health and disease conditions. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1834*(11), 2454–2461. https://doi.org/10.1016/j.bbapap.2013.04.007
Canadian Cancer Statistics Advisory Committee, Canadian Cancer Society, Public Health Agency of Canada, & Statistics Canada. (2021). *Canadian Cancer Statistics 2021*. Canadian Cancer Society. http://cancer.ca/Canadian-Cancer-Statistics-2021-EN
Carlson, M. (2017a). *Hgu133a.db*. Bioconductor. https://doi.org/10.18129/B9.BIOC.HGU133A.DB
Carlson, M. (2017b). *Hgu133plus2.db*. Bioconductor. https://doi.org/10.18129/B9.BIOC.HGU133PLUS2.DB
Carvalho, A. S., Cuco, C. M., Lavareda, C., Miguel, F., Ventura, M., Almeida, S., Pinto, P., de Abreu, T. T., Rodrigues, L. V., Seixas, S., Bárbara, C., Azkargorta, M., Elortza, F., Semedo, J., Field, J. K., Mota, L., & Matthiesen, R. (2017). Bronchoalveolar Lavage Proteomics in Patients with Suspected Lung Cancer. *Scientific Reports*, *7*(1), 42190. https://doi.org/10.1038/srep42190
Cell Signaling Technology. (2022). *EGF Receptor (L858R Mutant Specific) (43B2) Rabbit mAb product page*. EGF Receptor (L858R Mutant Specific) (43B2) Rabbit MAb. https://www.cellsignal.com/products/primary-antibodies/egf-receptor-l858r-mutant-specific-43b2-rabbit-mab/3197
Chardin, P., & McCormick, F. (1999). Brefeldin A. *Cell*, *97*(2), 153–155. https://doi.org/10.1016/S0092-8674(00)80724-2
Cheng, L., & Hill, A. F. (2022). Therapeutically harnessing extracellular vesicles. *Nature Reviews Drug Discovery*, *21*(5), 379–399. https://doi.org/10.1038/s41573-022-00410-w
Chevallier, M., Borgeaud, M., Addeo, A., & Friedlaender, A. (2021). Oncogenic driver mutations in non-small cell lung cancer: Past, present and future. *World Journal of Clinical Oncology*, *12*(4), 217–237. https://doi.org/10.5306/wjco.v12.i4.217
Chitale, D., Gong, Y., Taylor, B. S., Broderick, S., Brennan, C., Somwar, R., Golas, B., Wang, L., Motoi, N., Szoke, J., Reinersman, J. M., Major, J., Sander, C., Seshan, V. E., Zakowski, M.

F., Rusch, V., Pao, W., Gerald, W., & Ladanyi, M. (2009). An integrated genomic analysis of lung cancer reveals loss of DUSP4 in EGFR-mutant tumors. *Oncogene*, *28*(31), 2773–2783. https://doi.org/10.1038/onc.2009.135
Cohen, M. J., Chirico, W. J., & Lipke, P. N. (2020). Through the back door: Unconventional protein secretion. *The Cell Surface*, *6*, 100045. https://doi.org/10.1016/j.tcsw.2020.100045
Collins, M. A., An, J., Peller, D., & Bowser, R. (2015). Total protein is an effective loading control for cerebrospinal fluid western blots. *Journal of Neuroscience Methods*, *251*, 72–82. https://doi.org/10.1016/j.jneumeth.2015.05.011
Cox, J., & Mann, M. (2007). Is Proteomics the New Genomics? *Cell*, *130*(3), 395–398. https://doi.org/10.1016/j.cell.2007.07.032
Dela Cruz, C. S., Tanoue, L. T., & Matthay, R. A. (2011). Lung Cancer: Epidemiology, Etiology, and Prevention. *Clinics in Chest Medicine*, *32*(4), 605–644. https://doi.org/10.1016/j.ccm.2011.09.001
Dimou, E., & Nickel, W. (2018). Unconventional mechanisms of eukaryotic protein secretion. *Current Biology*, *28*(8), R406–R410. https://doi.org/10.1016/j.cub.2017.11.074
Dowling, P., & Clynes, M. (2011). Conditioned media from cell lines: A complementary model to clinical specimens for the discovery of disease-specific biomarkers. *PROTEOMICS*, *11*(4), 794–804. https://doi.org/10.1002/pmic.201000530
Du, Z., & Lovly, C. M. (2018). Mechanisms of receptor tyrosine kinase activation in cancer. *Molecular Cancer*, *17*(1), 58. https://doi.org/10.1186/s12943-018-0782-4
Duffy, M. J., & O'Byrne, K. (2018). Tissue and Blood Biomarkers in Lung Cancer: A Review. In *Advances in Clinical Chemistry* (Vol. 86, pp. 1–21). Elsevier. https://doi.org/10.1016/bs.acc.2018.05.001
Dupont, N., Jiang, S., Pilli, M., Ornatowski, W., Bhattacharya, D., & Deretic, V. (2011). Autophagy-based unconventional secretory pathway for extracellular delivery of IL-1β: Autophagy-based unconventional secretory pathway. *The EMBO Journal*, *30*(23), 4701–4711. https://doi.org/10.1038/emboj.2011.398
Foggetti, G., Li, C., Cai, H., Hellyer, J. A., Lin, W.-Y., Ayeni, D., Hastings, K., Choi, J., Wurtz, A., Andrejka, L., Maghini, D. G., Rashleigh, N., Levy, S., Homer, R., Gettinger, S. N., Diehn, M., Wakelee, H. A., Petrov, D. A., Winslow, M. M., & Politi, K. (2021). Genetic Determinants of EGFR-Driven Lung Cancer Growth and Therapeutic Response *In Vivo*. *Cancer Discovery*, *11*(7), 1736–1753. https://doi.org/10.1158/2159-8290.CD-20-1385
Frampton, G. M., Ali, S. M., Rosenzweig, M., Chmielecki, J., Lu, X., Bauer, T. M., Akimov, M., Bufill, J. A., Lee, C., Jentz, D., Hoover, R., Ou, S.-H. I., Salgia, R., Brennan, T., Chalmers, Z. R., Jaeger, S., Huang, A., Elvin, J. A., Erlich, R., … Miller, V. A. (2015). Activation of MET via Diverse Exon 14 Splicing Alterations Occurs in Multiple Tumor Types and Confers Clinical Sensitivity to MET Inhibitors. *Cancer Discovery*, *5*(8), 850–859. https://doi.org/10.1158/2159-8290.CD-15-0285
Fulda, S., Gorman, A. M., Hori, O., & Samali, A. (2010). Cellular Stress Responses: Cell Survival and Cell Death. *International Journal of Cell Biology*, *2010*, 1–23. https://doi.org/10.1155/2010/214074
Gainor, J. F., Shaw, A. T., Sequist, L. V., Fu, X., Azzoli, C. G., Piotrowska, Z., Huynh, T. G., Zhao, L., Fulton, L., Schultz, K. R., Howe, E., Farago, A. F., Sullivan, R. J., Stone, J. R., Digumarthy, S., Moran, T., Hata, A. N., Yagi, Y., Yeap, B. Y., … Mino-Kenudson, M. (2016). *EGFR* Mutations and *ALK* Rearrangements Are Associated with Low Response Rates to PD-1

Pathway Blockade in Non–Small Cell Lung Cancer: A Retrospective Analysis. *Clinical Cancer Research*, *22*(18), 4585–4593. https://doi.org/10.1158/1078-0432.CCR-15-3101
Gazdar, A. F., & Minna, J. D. (2008). Deregulated *EGFR* Signaling during Lung Cancer Progression: Mutations, Amplicons, and Autocrine Loops: Fig. 1. *Cancer Prevention Research*, *1*(3), 156–160. https://doi.org/10.1158/1940-6207.CAPR-08-0080
Gierada, D. S., Black, W. C., Chiles, C., Pinsky, P. F., & Yankelevitz, D. F. (2020). Low-Dose CT Screening for Lung Cancer: Evidence from 2 Decades of Study. *Radiology: Imaging Cancer*, *2*(2), e190058. https://doi.org/10.1148/rycan.2020190058
Gridelli, C., Rossi, A., Carbone, D. P., Guarize, J., Karachaliou, N., Mok, T., Petrella, F., Spaggiari, L., & Rosell, R. (2015). Non-small-cell lung cancer. *Nature Reviews Disease Primers*, *1*(1), 15009. https://doi.org/10.1038/nrdp.2015.9
Han, X., Na, T., Wu, T., & Yuan, B.-Z. (2020). Human lung epithelial BEAS-2B cells exhibit characteristics of mesenchymal stem cells. *PLOS ONE*, *15*(1), e0227174. https://doi.org/10.1371/journal.pone.0227174
Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. *Cancer Discovery*, *12*(1), 31–46. https://doi.org/10.1158/2159-8290.CD-21-1059
Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of Cancer: The Next Generation. *Cell*, *144*(5), 646–674. https://doi.org/10.1016/j.cell.2011.02.013
Hao, H., Maeda, Y., Fukazawa, T., Yamatsuji, T., Takaoka, M., Bao, X.-H., Matsuoka, J., Okui, T., Shimo, T., Takigawa, N., Tomono, Y., Nakajima, M., Fink-Baldauf, I. M., Nelson, S., Seibel, W., Papoian, R., Whitsett, J. A., & Naomoto, Y. (2013). Inhibition of the Growth Factor MDK/Midkine by a Novel Small Molecule Compound to Treat Non-Small Cell Lung Cancer. *PLoS ONE*, *8*(8), e71093. https://doi.org/10.1371/journal.pone.0071093
Hayashida, K., Bartlett, A. H., Chen, Y., & Park, P. W. (2010). Molecular and Cellular Mechanisms of Ectodomain Shedding. *The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology*, *293*(6), 925–937. https://doi.org/10.1002/ar.20757
Herbst, R. S., Heymach, J. V., & Lippman, S. M. (2008). Lung Cancer. *New England Journal of Medicine*, *359*(13), 1367–1380. https://doi.org/10.1056/NEJMra0802714
Herbst, R. S., Morgensztern, D., & Boshoff, C. (2018). The biology and management of non-small cell lung cancer. *Nature*, *553*(7689), 446–454. https://doi.org/10.1038/nature25183
Hirsch, F. R., Suda, K., Wiens, J., & Bunn, P. A. (2016). New and emerging targeted treatments in advanced non-small-cell lung cancer. *The Lancet*, *388*(10048), 1012–1024. https://doi.org/10.1016/S0140-6736(16)31473-8
Hmmier, A., O'Brien, M. E., Lynch, V., Clynes, M., Morgan, R., & Dowling, P. (2017). Proteomic analysis of bronchoalveolar lavage fluid (BALF) from lung cancer patients using label-free mass spectrometry. *BBA Clinical*, *7*, 97–104. https://doi.org/10.1016/j.bbacli.2017.03.001
Hsiao, K.-C., Shih, N.-Y., Fang, H.-L., Huang, T.-S., Kuo, C.-C., Chu, P.-Y., Hung, Y.-M., Chou, S.-W., Yang, Y.-Y., Chang, G.-C., & Liu, K.-J. (2013). Surface α-Enolase Promotes Extracellular Matrix Degradation and Tumor Metastasis and Represents a New Therapeutic Target. *PLoS ONE*, *8*(7), e69354. https://doi.org/10.1371/journal.pone.0069354
Hu, R., Huffman, K. E., Chu, M., Zhang, Y., Minna, J. D., & Yu, Y. (2016). Quantitative Secretomic Analysis Identifies Extracellular Protein Factors That Modulate the Metastatic Phenotype of Non-Small Cell Lung Cancer. *Journal of Proteome Research*, *15*(2), 477–486. https://doi.org/10.1021/acs.jproteome.5b00819

Hynds, R. E., Frese, K. K., Pearce, D. R., Grönroos, E., Dive, C., & Swanton, C. (2021). Progress towards non-small-cell lung cancer models that represent clinical evolutionary trajectories. *Open Biology*, *11*(1), 200247. https://doi.org/10.1098/rsob.200247
Imielinski, M., Berger, A. H., Hammerman, P. S., Hernandez, B., Pugh, T. J., Hodis, E., Cho, J., Suh, J., Capelletti, M., Sivachenko, A., Sougnez, C., Auclair, D., Lawrence, M. S., Stojanov, P., Cibulskis, K., Choi, K., de Waal, L., Sharifnia, T., Brooks, A., … Meyerson, M. (2012). Mapping the Hallmarks of Lung Adenocarcinoma with Massively Parallel Sequencing. *Cell*, *150*(6), 1107–1120. https://doi.org/10.1016/j.cell.2012.08.029
Inamura, K. (2018). Clinicopathological Characteristics and Mutations Driving Development of Early Lung Adenocarcinoma: Tumor Initiation and Progression. *International Journal of Molecular Sciences*, *19*(4), 1259. https://doi.org/10.3390/ijms19041259
Inoue, Y., Nikolic, A., Farnsworth, D., Liu, A., Ladanyi, M., Somwar, R., Gallo, M., & Lockwood, W. W. (2020). *Extracellular signal-regulated kinase mediates chromatin rewiring and lineage transformation in lung cancer* [Preprint]. Cancer Biology. https://doi.org/10.1101/2020.11.12.368522
Izbicka, E., Streeper, R. T., Michalek, J. E., Louden, C. L., Iii, A. D., & Campos, D. R. (2012). Plasma Biomarkers Distinguish Non-small Cell Lung Cancer from Asthma and Differ in Men and Women. *CANCER GENOMICS*, 9.
Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065), 20150202. https://doi.org/10.1098/rsta.2015.0202
Jordan, E. J., Kim, H. R., Arcila, M. E., Barron, D., Chakravarty, D., Gao, J., Chang, M. T., Ni, A., Kundra, R., Jonsson, P., Jayakumaran, G., Gao, S. P., Johnsen, H. C., Hanrahan, A. J., Zehir, A., Rekhtman, N., Ginsberg, M. S., Li, B. T., Yu, H. A., … Riely, G. J. (2017). Prospective Comprehensive Molecular Characterization of Lung Adenocarcinomas for Efficient Patient Matching to Approved and Emerging Therapies. *Cancer Discovery*, *7*(6), 596–609. https://doi.org/10.1158/2159-8290.CD-16-1337
Jung, J., Kim, J., Roh, S. H., Jun, I., Sampson, R. D., Gee, H. Y., Choi, J. Y., & Lee, M. G. (2016). The HSP70 co-chaperone DNAJC14 targets misfolded pendrin for unconventional protein secretion. *Nature Communications*, *7*(1), 11386. https://doi.org/10.1038/ncomms11386
Kadara, H., Choi, M., Zhang, J., Parra, E. R., Rodriguez-Canales, J., Gaffney, S. G., Zhao, Z., Behrens, C., Fujimoto, J., Chow, C., Yoo, Y., Kalhor, N., Moran, C., Rimm, D., Swisher, S., Gibbons, D. L., Heymach, J., Kaftan, E., Townsend, J. P., … Herbst, R. S. (2017). Whole-exome sequencing and immune profiling of early-stage lung adenocarcinoma with fully annotated clinical follow-up. *Annals of Oncology*, *28*(1), 75–82. https://doi.org/10.1093/annonc/mdw436
Kaneto, N., Yokoyama, S., Hayakawa, Y., Kato, S., Sakurai, H., & Saiki, I. (2014). RAC 1 inhibition as a therapeutic target for gefitinib-resistant non-small-cell lung cancer. *Cancer Science*, *105*(7), 788–794. https://doi.org/10.1111/cas.12425
Kang, H.-R., Cho, J. Y., Lee, S. H., Lee, Y. J., Park, J. S., Cho, Y.-J., Yoon, H. I., Lee, K. W., Lee, J. H., & Lee, C.-T. (2019). Role of Low-Dose Computerized Tomography in Lung Cancer Screening among Never-Smokers. *Journal of Thoracic Oncology*, *14*(3), 436–444. https://doi.org/10.1016/j.jtho.2018.11.002
Katz, M., Amit, I., & Yarden, Y. (2007). Regulation of MAPKs by growth factors and receptor tyrosine kinases. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, *1773*(8), 1161–1176. https://doi.org/10.1016/j.bbamcr.2007.01.002

Kent, K. D., & Bomser, J. A. (2003). Bovine pituitary extract provides remarkable protection against oxidative stress in human prostate epithelial cells. *In Vitro Cellular & Developmental Biology-Animal*, *39*(8), 388–394.
Kerpedjieva, S. S., Kim, D. S., Barbeau, D. J., & Tamama, K. (2012). EGFR Ligands Drive Multipotential Stromal Cells to Produce Multiple Growth Factors and Cytokines via Early Growth Response-1. *Stem Cells and Development*, *21*(13), 2541–2551. https://doi.org/10.1089/scd.2011.0711
Kim, J. E., Koo, K. H., Kim, Y. H., Sohn, J., & Park, Y. G. (2008). Identification of potential lung cancer biomarkers using an in vitro carcinogenesis model. *Experimental and Molecular Medicine*, *40*(6), 709. https://doi.org/10.3858/emm.2008.40.6.709
Kim, J., Gee, H. Y., & Lee, M. G. (2018). Unconventional protein secretion – new insights into the pathogenesis and therapeutic targets of human diseases. *Journal of Cell Science*, *131*(12), jcs213686. https://doi.org/10.1242/jcs.213686
Kissil, J. L., Walmsley, M. J., Hanlon, L., Haigis, K. M., Bender Kim, C. F., Sweet-Cordero, A., Eckman, M. S., Tuveson, D. A., Capobianco, A. J., Tybulewicz, V. L. J., & Jacks, T. (2007). Requirement for Rac1 in a K-ras–Induced Lung Cancer in the Mouse. *Cancer Research*, *67*(17), 8089–8094. https://doi.org/10.1158/0008-5472.CAN-07-2300
Kojima, K., Konopleva, M., McQueen, T., O'Brien, S., Plunkett, W., & Andreeff, M. (2006). Mdm2 inhibitor Nutlin-3a induces p53-mediated apoptosis by transcription-dependent and transcription-independent mechanisms and may overcome Atm-mediated resistance to fludarabine in chronic lymphocytic leukemia. *Blood*, *108*(3), 993–1000. https://doi.org/10.1182/blood-2005-12-5148
Kubo, T., Yamamoto, H., Lockwood, W. W., Valencia, I., Soh, J., Peyton, M., Jida, M., Otani, H., Fujii, T., Ouchida, M., Takigawa, N., Kiura, K., Shimizu, K., Date, H., Minna, J. D., Varella-Garcia, M., Lam, W. L., Gazdar, A. F., & Toyooka, S. (2009). *MET* gene amplification or *EGFR* mutation activate MET in lung cancers untreated with EGFR tyrosine kinase inhibitors. *International Journal of Cancer*, *124*(8), 1778–1784. https://doi.org/10.1002/ijc.24150
Labbé, C., Beaudoin, S., Martel, S., Delage, A., Joubert, P., Drapeau, C., & Provencher, S. (2015). Diagnostic yield of non-guided flexible bronchoscopy for peripheral pulmonary neoplasia: Bronchoscopy & peripheral lung cancer. *Thoracic Cancer*, *6*(4), 517–523. https://doi.org/10.1111/1759-7714.12223
Lam, S. (2019). Lung Cancer Screening in Never-Smokers. *Journal of Thoracic Oncology*, *14*(3), 336–337. https://doi.org/10.1016/j.jtho.2018.12.019
Lamberti, G., Andrini, E., Sisi, M., Rizzo, A., Parisi, C., Di Federico, A., Gelsomino, F., & Ardizzoni, A. (2020). Beyond EGFR, ALK and ROS1: Current evidence and future perspectives on newly targetable oncogenic drivers in lung adenocarcinoma. *Critical Reviews in Oncology/Hematology*, *156*, 103119. https://doi.org/10.1016/j.critrevonc.2020.103119
Larsen, J. E., & Minna, J. D. (2011). Molecular Biology of Lung Cancer: Clinical Implications. *Clinics in Chest Medicine*, *32*(4), 703–740. https://doi.org/10.1016/j.ccm.2011.08.003
Lemjabbar-Alaoui, H., Hassan, O. U., Yang, Y.-W., & Buchanan, P. (2015). Lung cancer: Biology and treatment options. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, *1856*(2), 189–210. https://doi.org/10.1016/j.bbcan.2015.08.002
Lemmon, M. A., Schlessinger, J., & Ferguson, K. M. (2014). The EGFR Family: Not So Prototypical Receptor Tyrosine Kinases. *Cold Spring Harbor Perspectives in Biology*, *6*(4), a020768–a020768. https://doi.org/10.1101/cshperspect.a020768

Li, H.-J., Ke, F.-Y., Lin, C.-C., Lu, M.-Y., Kuo, Y.-H., Wang, Y.-P., Liang, K.-H., Lin, S.-C., Chang, Y.-H., Chen, H.-Y., Yang, P.-C., & Wu, H.-C. (2021). ENO1 Promotes Lung Cancer Metastasis via HGFR and WNT Signaling–Driven Epithelial-to-Mesenchymal Transition. *Cancer Research*, *81*(15), 4094–4109. https://doi.org/10.1158/0008-5472.CAN-20-3543
Li, Q., Birkbak, N. J., Gyorffy, B., Szallasi, Z., & Eklund, A. C. (2011). Jetset: Selecting the optimal microarray probe set to represent a gene. *BMC Bioinformatics*, *12*(1), 474. https://doi.org/10.1186/1471-2105-12-474
Lin, C.-Y., Tsui, K.-H., Yu, C.-C., Yeh, C.-W., Chang, P.-L., & Yung, B. Y.-M. (2006). Searching cell-secreted proteomes for potential urinary bladder tumor markers. *PROTEOMICS*, *6*(15), 4381–4389. https://doi.org/10.1002/pmic.200600066
Liu, J., Jang, J. Y., Pirooznia, M., Liu, S., & Finkel, T. (2021). The secretome mouse provides a genetic platform to delineate tissue-specific in vivo secretion. *Proceedings of the National Academy of Sciences*, *118*(3), e2005134118. https://doi.org/10.1073/pnas.2005134118
López-Otín, C., & Bond, J. S. (2008). Proteases: Multifunctional Enzymes in Life and Disease. *Journal of Biological Chemistry*, *283*(45), 30433–30437. https://doi.org/10.1074/jbc.R800035200
Ma, Z., Liu, D., Li, W., Di, S., Zhang, Z., Zhang, J., Xu, L., Guo, K., Zhu, Y., Han, J., Li, X., & Yan, X. (2019). STYK1 promotes tumor growth and metastasis by reducing SPINT2/HAI-2 expression in non-small cell lung cancer. *Cell Death & Disease*, *10*(6), 435. https://doi.org/10.1038/s41419-019-1659-1
Makridakis, M., & Vlahou, A. (2010). Secretome proteomics for discovery of cancer biomarkers. *Journal of Proteomics*, *73*(12), 2291–2305. https://doi.org/10.1016/j.jprot.2010.07.001
Manser, R. L., Irving, L. B., Byrnes, G., Abramson, M. J., Stone, C. A., & Campbell, D. A. (2003). Screening for lung cancer: A systematic review and meta-analysis of controlled trials. *Thorax*, *58*(9), 784–789. https://doi.org/10.1136/thorax.58.9.784
Massion, P. P., & Walker, R. C. (2014). Indeterminate Pulmonary Nodules: Risk for Having or for Developing Lung Cancer? *Cancer Prevention Research*, *7*(12), 1173–1178. https://doi.org/10.1158/1940-6207.CAPR-14-0364
Mbeunkui, F., Fodstad, O., & Pannell, L. K. (2006). Secretory Protein Enrichment and Analysis: An Optimized Approach Applied on Cancer Cell Lines Using 2D LC−MS/MS. *Journal of Proteome Research*, *5*(4), 899–906. https://doi.org/10.1021/pr050375p
Meldolesi, J. (2018). Exosomes and Ectosomes in Intercellular Communication. *Current Biology*, *28*(8), R435–R444. https://doi.org/10.1016/j.cub.2018.01.059
Meldolesi, J. (2022). Unconventional Protein Secretion Dependent on Two Extracellular Vesicles: Exosomes and Ectosomes. *Frontiers in Cell and Developmental Biology*, *10*, 877344. https://doi.org/10.3389/fcell.2022.877344
Méndez, O., & Villanueva, J. (2015). Challenges and opportunities for cell line secretomes in cancer proteomics. *PROTEOMICS - Clinical Applications*, *9*(3–4), 348–357. https://doi.org/10.1002/prca.201400131
Mukherjee, P., & Mani, S. (2013). Methodologies to decipher the cell secretome. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1834*(11), 2226–2232. https://doi.org/10.1016/j.bbapap.2013.01.022
Nicholson, A. G., Tsao, M. S., Beasley, M. B., Borczuk, A. C., Brambilla, E., Cooper, W. A., Dacic, S., Jain, D., Kerr, K. M., Lantuejoul, S., Noguchi, M., Papotti, M., Rekhtman, N., Scagliotti, G., van Schil, P., Sholl, L., Yatabe, Y., Yoshida, A., & Travis, W. D. (2022). The

2021 WHO Classification of Lung Tumors: Impact of Advances Since 2015. *Journal of Thoracic Oncology*, *17*(3), 362–387. https://doi.org/10.1016/j.jtho.2021.11.003
Nilsson, T., Mann, M., Aebersold, R., Yates, J. R., Bairoch, A., & Bergeron, J. J. M. (2010). Mass spectrometry in high-throughput proteomics: Ready for the big time. *Nature Methods*, *7*(9), 681–685. https://doi.org/10.1038/nmeth0910-681
Nooreldeen, R., & Bach, H. (2021). Current and Future Development in Lung Cancer Diagnosis. *International Journal of Molecular Sciences*, *22*(16), 8661. https://doi.org/10.3390/ijms22168661
Nylander, S., & Kalies, I. (1999). *Brefeldin A, but not monensin, completely blocks CD69 expression on mouse lymphocytes: Efficacy of inhibitors of protein secretion in protocols for intracellular cytokine staining by flow cytometry*. 8.
Öhlund, D., Lundin, C., Ardnor, B., Öman, M., Naredi, P., & Sund, M. (2009). Type IV collagen is a tumour stroma-derived biomarker for pancreas cancer. *British Journal of Cancer*, *101*(1), 91–97. https://doi.org/10.1038/sj.bjc.6605107
Omenn, G. S., States, D. J., Adamski, M., Blackwell, T. W., Menon, R., Hermjakob, H., Apweiler, R., Haab, B. B., Simpson, R. J., Eddes, J. S., Kapp, E. A., Moritz, R. L., Chan, D. W., Rai, A. J., Admon, A., Aebersold, R., Eng, J., Hancock, W. S., Hefta, S. A., … Hanash, S. M. (2005). Overview of the HUPO Plasma Proteome Project: Results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *PROTEOMICS*, *5*(13), 3226–3245. https://doi.org/10.1002/pmic.200500358
Oudkerk, M., Liu, S., Heuvelmans, M. A., Walter, J. E., & Field, J. K. (2021). Lung cancer LDCT screening and mortality reduction—Evidence, pitfalls and future perspectives. *Nature Reviews Clinical Oncology*, *18*(3), 135–151. https://doi.org/10.1038/s41571-020-00432-6
Owji, H., Nezafat, N., Negahdaripour, M., Hajiebrahimi, A., & Ghasemi, Y. (2018). A comprehensive review of signal peptides: Structure, roles, and applications. *European Journal of Cell Biology*, *97*(6), 422–441. https://doi.org/10.1016/j.ejcb.2018.06.003
Pikor, L. A., Lockwood, W. W., Thu, K. L., Vucic, E. A., Chari, R., Gazdar, A. F., Lam, S., & Lam, W. L. (2013). YEATS4 Is a Novel Oncogene Amplified in Non–Small Cell Lung Cancer That Regulates the p53 Pathway. *Cancer Research*, *73*(24), 7301–7312. https://doi.org/10.1158/0008-5472.CAN-13-1897
Plubell, D. L., Wilmarth, P. A., Zhao, Y., Fenton, A. M., Minnier, J., Reddy, A. P., Klimek, J., Yang, X., David, L. L., & Pamir, N. (2017). Extended Multiplexing of Tandem Mass Tags (TMT) Labeling Reveals Age and High Fat Diet Specific Proteome Changes in Mouse Epididymal Adipose Tissue. *Molecular & Cellular Proteomics*, *16*(5), 873–890. https://doi.org/10.1074/mcp.M116.065524
Prabhakar, B., Shende, P., & Augustine, S. (2018). Current trends and emerging diagnostic techniques for lung cancer. *Biomedicine & Pharmacotherapy*, *106*, 1586–1599. https://doi.org/10.1016/j.biopha.2018.07.145
Rabouille, C. (2017). Pathways of Unconventional Protein Secretion. *Trends in Cell Biology*, *27*(3), 230–240. https://doi.org/10.1016/j.tcb.2016.11.007
Ramirez, R. D., Sheridan, S., Girard, L., Sato, M., Kim, Y., Pollack, J., Peyton, M., Zou, Y., Kurie, J. M., DiMaio, J. M., Milchgrub, S., Smith, A. L., Souza, R. F., Gilbey, L., Zhang, X., Gandia, K., Vaughan, M. B., Wright, W. E., Gazdar, A. F., … Minna, J. D. (2004).

Immortalization of Human Bronchial Epithelial Cells in the Absence of Viral Oncoproteins. *Cancer Research*, *64*(24), 9027–9034. https://doi.org/10.1158/0008-5472.CAN-04-3703
Rappsilber, J., Ishihama, Y., & Mann, M. (2003). Stop and Go Extraction Tips for Matrix-Assisted Laser Desorption/Ionization, Nanoelectrospray, and LC/MS Sample Pretreatment in Proteomics. *Analytical Chemistry*, *75*(3), 663–670. https://doi.org/10.1021/ac026117i
Reddel, R. R., Ke, Y., Gerwin, B. I., McMenamin, M. G., Lechner, J. F., Su, R. T., Brash, D. E., Park, J.-B., Rhim, J. S., & Harris, C. C. (1988). Transformation of Human Bronchial Epithelial Cells by Infection with SV40 or Adenovirus-12 SV40 Hybrid Virus, or Transfection via Strontium Phosphate Coprecipitation with a Plasmid Containing SV40 Early Region Genes. *Cancer Research*, *48*(7), 1904–1909.
Reddel, R. R., Salghetti, S. E., Willey, J. C., Ohnuki, Y., Ke, Y., Gerwin, B. I., Lechner, J. F., & Harris, C. C. (1993). Development of Tumorigenicity in Simian Virus 40-immortalized Human Bronchial Epithelial Cell Lines1. *Cancer Research*, *53*(5), 985–991.
Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *43*(7), e47–e47. https://doi.org/10.1093/nar/gkv007
Roberts, P. J., & Der, C. J. (2007). Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer. *Oncogene*, *26*(22), 3291–3310. https://doi.org/10.1038/sj.onc.1210422
Robinson, J. L., Feizi, A., Uhlén, M., & Nielsen, J. (2019). A Systematic Investigation of the Malignant Functions and Diagnostic Potential of the Cancer Secretome. *Cell Reports*, *26*(10), 2622-2635.e5. https://doi.org/10.1016/j.celrep.2019.02.025
Rudin, C. M., Avila-Tang, E., Harris, C. C., Herman, J. G., Hirsch, F. R., Pao, W., Schwartz, A. G., Vahakangas, K. H., & Samet, J. M. (2009). Lung Cancer in Never Smokers: Molecular Profiles and Therapeutic Implications. *Clinical Cancer Research*, *15*(18), 5646–5661. https://doi.org/10.1158/1078-0432.CCR-09-0377
Rudin, C. M., Brambilla, E., Faivre-Finn, C., & Sage, J. (2021). Small-cell lung cancer. *Nature Reviews Disease Primers*, *7*(1), 3. https://doi.org/10.1038/s41572-020-00235-0
Sasai, K., Sukezane, T., Yanagita, E., Nakagawa, H., Hotta, A., Itoh, T., & Akagi, T. (2011). Oncogene-Mediated Human Lung Epithelial Cell Transformation Produces Adenocarcinoma Phenotypes <i>in Vivo</I>. *Cancer Research*, *71*(7), 2541–2549. https://doi.org/10.1158/0008-5472.CAN-10-2221
Sato, M., Larsen, J. E., Lee, W., Sun, H., Shames, D. S., Dalvi, M. P., Ramirez, R. D., Tang, H., DiMaio, J. M., Gao, B., Xie, Y., Wistuba, I. I., Gazdar, A. F., Shay, J. W., & Minna, J. D. (2013). Human Lung Epithelial Cells Progressed to Malignancy through Specific Oncogenic Manipulations. *Molecular Cancer Research*, *11*(6), 638–650. https://doi.org/10.1158/1541-7786.MCR-12-0634-T
Sato, M., Shay, J. W., & Minna, J. D. (2020). Immortalized normal human lung epithelial cell models for studying lung cancer biology. *Respiratory Investigation*, *58*(5), 344–354. https://doi.org/10.1016/j.resinv.2020.04.005
Sato, M., Vaughan, M. B., Girard, L., Peyton, M., Lee, W., Shames, D. S., Ramirez, R. D., Sunaga, N., Gazdar, A. F., Shay, J. W., & Minna, J. D. (2006). Multiple Oncogenic Changes (*K-RAS* ^{V12} , p53 Knockdown, *Mutant EGFRs, p16* Bypass, Telomerase) Are Not Sufficient to Confer a Full Malignant Phenotype on Human Bronchial Epithelial Cells. *Cancer Research*, *66*(4), 2116–2128. https://doi.org/10.1158/0008-5472.CAN-05-2521

Seijo, L. M., Peled, N., Ajona, D., Boeri, M., Field, J. K., Sozzi, G., Pio, R., Zulueta, J. J., Spira, A., Massion, P. P., Mazzone, P. J., & Montuenga, L. M. (2019). Biomarkers in Lung Cancer Screening: Achievements, Promises, and Challenges. *Journal of Thoracic Oncology*, *14*(3), 343–357. https://doi.org/10.1016/j.jtho.2018.11.023
Semenova, E. A., Nagel, R., & Berns, A. (2015). Origins, genetic landscape, and emerging therapies of small cell lung cancer. *Genes & Development*, *29*(14), 1447–1462. https://doi.org/10.1101/gad.263145.115
Severino, V., Farina, A., & Chambery, A. (2013). Analysis of Secreted Proteins. In M. Zhou & T. Veenstra (Eds.), *Proteomics for Biomarker Discovery* (Vol. 1002, pp. 37–60). Humana Press. https://doi.org/10.1007/978-1-62703-360-2_4
Shen, J., Jin, Z., Lv, H., Jin, K., Jonas, K., Zhu, C., & Chen, B. (2020). PFKP is highly expressed in lung cancer and regulates glucose metabolism. *Cellular Oncology*, *43*(4), 617–629. https://doi.org/10.1007/s13402-020-00508-6
Shin, J., Rhim, J., Kwon, Y., Choi, S. Y., Shin, S., Ha, C.-W., & Lee, C. (2019). Comparative analysis of differentially secreted proteins in serum-free and serum-containing media by using BONCAT and pulsed SILAC. *Scientific Reports*, *9*(1), 3096. https://doi.org/10.1038/s41598-019-39650-z
Siegelin, M. D., & Borczuk, A. C. (2014). Epidermal growth factor receptor mutations in lung adenocarcinoma. *Laboratory Investigation*, *94*(2), 129–137. https://doi.org/10.1038/labinvest.2013.147
Skoulidis, F., & Heymach, J. V. (2019). Co-occurring genomic alterations in non-small-cell lung cancer biology and therapy. *Nature Reviews Cancer*, *19*(9), 495–509. https://doi.org/10.1038/s41568-019-0179-8
Soh, J., Toyooka, S., Ichihara, S., Asano, H., Kobayashi, N., Suehisa, H., Otani, H., Yamamoto, H., Ichimura, K., Kiura, K., Gazdar, A. F., & Date, H. (2008). Sequential Molecular Changes during Multistage Pathogenesis of Small Peripheral Adenocarcinomas of the Lung. *Journal of Thoracic Oncology*, *3*(4), 340–347. https://doi.org/10.1097/JTO.0b013e318168d20a
Sone, S., Nakayama, T., Honda, T., Tsushima, K., Li, F., Haniuda, M., Takahashi, Y., Suzuki, T., Yamanda, T., Kondo, R., Hanaoka, T., Takayama, F., Kubo, K., & Fushimi, H. (2007). Long-term follow-up study of a population-based 1996–1998 mass screening programme for lung cancer using mobile low-dose spiral computed tomography. *Lung Cancer*, *58*(3), 329–341. https://doi.org/10.1016/j.lungcan.2007.06.022
Sorkin, A., & Goh, L. K. (2009). Endocytosis and intracellular trafficking of ErbBs. *Experimental Cell Research*, *315*(4), 683–696. https://doi.org/10.1016/j.yexcr.2008.07.029
Sozzi, G., & Boeri, M. (2014). Potential biomarkers for lung cancer screening. *Translational Lung Cancer Research*, *3*(3), 10.
Stanzel, F. (2013). Bronchoalveolar Lavage. In A. Ernst & F. J. Herth (Eds.), *Principles and Practice of Interventional Pulmonology* (pp. 165–176). Springer New York. https://doi.org/10.1007/978-1-4614-4292-9_16
Steringer, J. P., Müller, H.-M., & Nickel, W. (2015). Unconventional Secretion of Fibroblast Growth Factor 2—A Novel Type of Protein Translocation across Membranes? *Journal of Molecular Biology*, *427*(6), 1202–1210. https://doi.org/10.1016/j.jmb.2014.07.012
Stern, L., Mueller, E., Bellon, E., Reeh, M., Grotelueschen, R., Guengoer, C., Melling, N., Goetz, M., Perez, D. R., Izbicki, J. R., Rawnaq-Möllers, T., & Ghadban, T. (2021). Serum

midkine as non-invasive biomarker for detection and prognosis of non-small cell lung cancer. *Scientific Reports*, *11*(1), 14616. https://doi.org/10.1038/s41598-021-94272-8
Sun, S., Schiller, J. H., & Gazdar, A. F. (2007). Lung cancer in never smokers—A different disease. *Nature Reviews Cancer*, *7*(10), 778–790. https://doi.org/10.1038/nrc2190
Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, *71*(3), 209–249. https://doi.org/10.3322/caac.21660
Taguchi, A., Politi, K., Pitteri, S. J., Lockwood, W. W., Faça, V. M., Kelly-Spratt, K., Wong, C.-H., Zhang, Q., Chin, A., Park, K.-S., Goodman, G., Gazdar, A. F., Sage, J., Dinulescu, D. M., Kucherlapati, R., DePinho, R. A., Kemp, C. J., Varmus, H. E., & Hanash, S. M. (2011). Lung Cancer Signatures in Plasma Based on Proteome Profiling of Mouse Tumor Models. *Cancer Cell*, *20*(3), 289–299. https://doi.org/10.1016/j.ccr.2011.08.007
Tammemagi, M. C., & Lam, S. (2014). Screening for lung cancer using low dose computed tomography. *BMJ*, *348*(may27 7), g2253–g2253. https://doi.org/10.1136/bmj.g2253
The National Lung Screening Trial Research Team. (2011). Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *New England Journal of Medicine*, *365*(5), 395–409. https://doi.org/10.1056/NEJMoa1102873
Travis, W. D. (2012). Update on small cell carcinoma and its differentiation from squamous cell carcinoma and other non-small cell carcinomas. *Modern Pathology*, *25*(S1), S18–S30. https://doi.org/10.1038/modpathol.2011.150
Travis, W. D., Brambilla, E., Nicholson, A. G., Yatabe, Y., Austin, J. H. M., Beasley, M. B., Chirieac, Lucian. R., Dacic, S., Duhig, E., Flieder, D. B., Geisinger, K., Hirsch, F. R., Ishikawa, Y., Kerr, K. M., Noguchi, M., Pelosi, G., Powell, C. A., Tsao, M. S., & Wistuba, I. (2015). The 2015 World Health Organization Classification of Lung Tumors. *Journal of Thoracic Oncology*, *10*(9), 1243–1260. https://doi.org/10.1097/JTO.0000000000000630
Travis, W. D., Brambilla, E., Noguchi, M., Nicholson, A. G., Geisinger, K. R., Yatabe, Y., Beer, D. G., Powell, C. A., Riely, G. J., Van Schil, P. E., Garg, K., Austin, J. H. M., Asamura, H., Rusch, V. W., Hirsch, F. R., Scagliotti, G., Mitsudomi, T., Huber, R. M., Ishikawa, Y., … Yankelewitz, D. (2011). International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society International Multidisciplinary Classification of Lung Adenocarcinoma. *Journal of Thoracic Oncology*, *6*(2), 244–285. https://doi.org/10.1097/JTO.0b013e318206a221
Turk, B., Turk, D., & Turk, V. (2012). Protease signalling: The cutting edge: Protease signalling. *The EMBO Journal*, *31*(7), 1630–1643. https://doi.org/10.1038/emboj.2012.42
Uhlén, M., Karlsson, M. J., Hober, A., Svensson, A.-S., Scheffel, J., Kotol, D., Zhong, W., Tebani, A., Strandberg, L., Edfors, F., Sjöstedt, E., Mulder, J., Mardinoglu, A., Berling, A., Ekblad, S., Dannemeyer, M., Kanje, S., Rockberg, J., Lundqvist, M., … Sivertsson, Å. (2019). The human secretome. *Science Signaling*, *12*(609), eaaz0274. https://doi.org/10.1126/scisignal.aaz0274
UniProt Consortium. (2022). *UniProtKB evidence page*. UniProtKB Evidence Page. https://www.uniprot.org/help/evidences
US Preventive Services Task Force, Krist, A. H., Davidson, K. W., Mangione, C. M., Barry, M. J., Cabana, M., Caughey, A. B., Davis, E. M., Donahue, K. E., Doubeni, C. A., Kubik, M., Landefeld, C. S., Li, L., Ogedegbe, G., Owens, D. K., Pbert, L., Silverstein, M., Stevermer, J.,

Tseng, C.-W., & Wong, J. B. (2021). Screening for Lung Cancer: US Preventive Services Task Force Recommendation Statement. *JAMA*, *325*(10), 962. https://doi.org/10.1001/jama.2021.1117
van Niel, G., D'Angelo, G., & Raposo, G. (2018). Shedding light on the cell biology of extracellular vesicles. *Nature Reviews Molecular Cell Biology*, *19*(4), 213–228. https://doi.org/10.1038/nrm.2017.125
von Heijne, G. (1990). The signal peptide. *The Journal of Membrane Biology*, *115*(3), 195–201. https://doi.org/10.1007/BF01868635
Wang, H., Chang-Wong, T., Tang, H.-Y., & Speicher, D. W. (2010). Comparison of Extensive Protein Fractionation and Repetitive LC-MS/MS Analyses on Depth of Analysis for Complex Proteomes. *Journal of Proteome Research*, *9*(2), 1032–1040. https://doi.org/10.1021/pr900927y
Wang, X.-D., Hu, R., Ding, Q., Savage, T. K., Huffman, K. E., Williams, N., Cobb, M. H., Minna, J. D., Johnson, J. E., & Yu, Y. (2019). Subtype-specific secretomic characterization of pulmonary neuroendocrine tumor cells. *Nature Communications*, *10*(1), 3201. https://doi.org/10.1038/s41467-019-11153-5
Wei, W., Riley, N. M., Yang, A. C., Kim, J. T., Terrell, S. M., Li, V. L., Garcia-Contreras, M., Bertozzi, C. R., & Long, J. Z. (2021). Cell type-selective secretome profiling in vivo. *Nature Chemical Biology*, *17*(3), 326–334. https://doi.org/10.1038/s41589-020-00698-y
Wickham, H. (2016). *Ggplot2*. Springer International Publishing. https://doi.org/10.1007/978-3-319-24277-4
Xue, H., Lu, B., & Lai, M. (2008). The cancer secretome: A reservoir of biomarkers. *Journal of Translational Medicine*, *6*(1), 52. https://doi.org/10.1186/1479-5876-6-52
Yan, Q., Jiang, L., Liu, M., Yu, D., Zhang, Y., Li, Y., Fang, S., Li, Y., Zhu, Y.-H., Yuan, Y.-F., & Guan, X.-Y. (2017). ANGPTL1 Interacts with Integrin α1β1 to Suppress HCC Angiogenesis and Metastasis by Inhibiting JAK2/STAT3 Signaling. *Cancer Research*, *77*(21), 5831–5845. https://doi.org/10.1158/0008-5472.CAN-17-0579
Zhang, M., Kenny, S. J., Ge, L., Xu, K., & Schekman, R. (2015). Translocation of interleukin-1β into a vesicle intermediate in autophagy-mediated secretion. *ELife*, *4*, e11205. https://doi.org/10.7554/eLife.11205
Zhao, Y., Aguilar, A., Bernard, D., & Wang, S. (2015). Small-Molecule Inhibitors of the MDM2–p53 Protein–Protein Interaction (MDM2 Inhibitors) in Clinical Trials for Cancer Treatment: Miniperspective. *Journal of Medicinal Chemistry*, *58*(3), 1038–1052. https://doi.org/10.1021/jm501092z
Zheng, M. (2016). Classification and Pathology of Lung Cancer. *Surgical Oncology Clinics of North America*, *25*(3), 447–468. https://doi.org/10.1016/j.soc.2016.02.003
Ziegler, Y. S., Moresco, J. J., Yates, J. R., & Nardulli, A. M. (2016). Integration of Breast Cancer Secretomes with Clinical Data Elucidates Potential Serum Markers for Disease Detection, Diagnosis, and Prognosis. *PLOS ONE*, *11*(6), e0158296. https://doi.org/10.1371/journal.pone.0158296
Zou, D., Ma, L., Yu, J., & Zhang, Z. (2015). Biological Databases for Human Research. *Genomics, Proteomics & Bioinformatics*, *13*(1), 55–63. https://doi.org/10.1016/j.gpb.2015.01.006

Appendix

This section consists of supplemental figures and methods described in the main text.

A.1 Supplemental figures

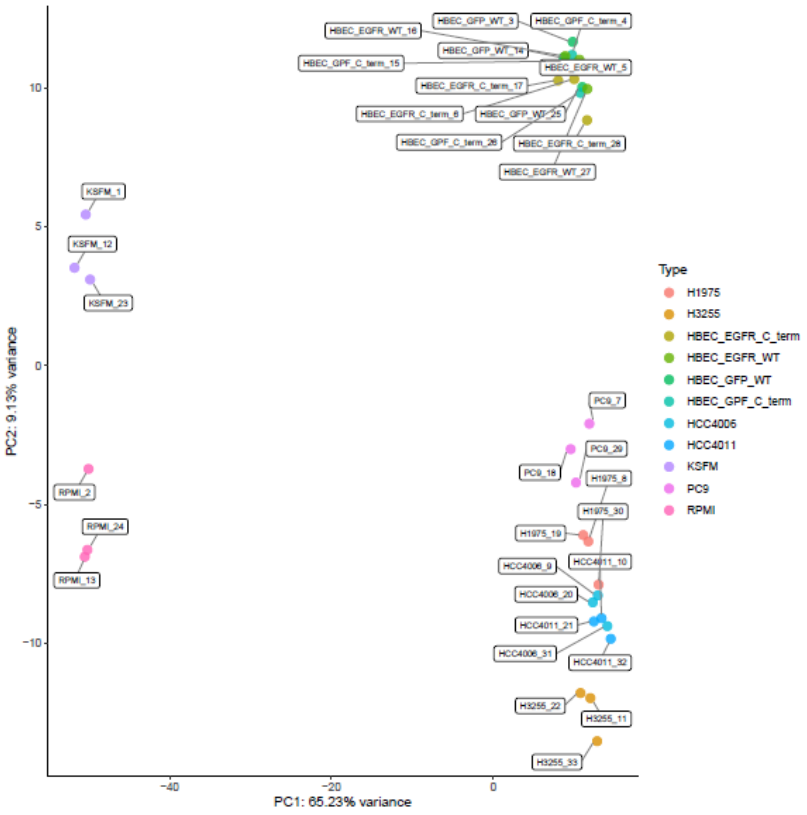


Figure S.1 PCA of identified proteins (including media background controls).

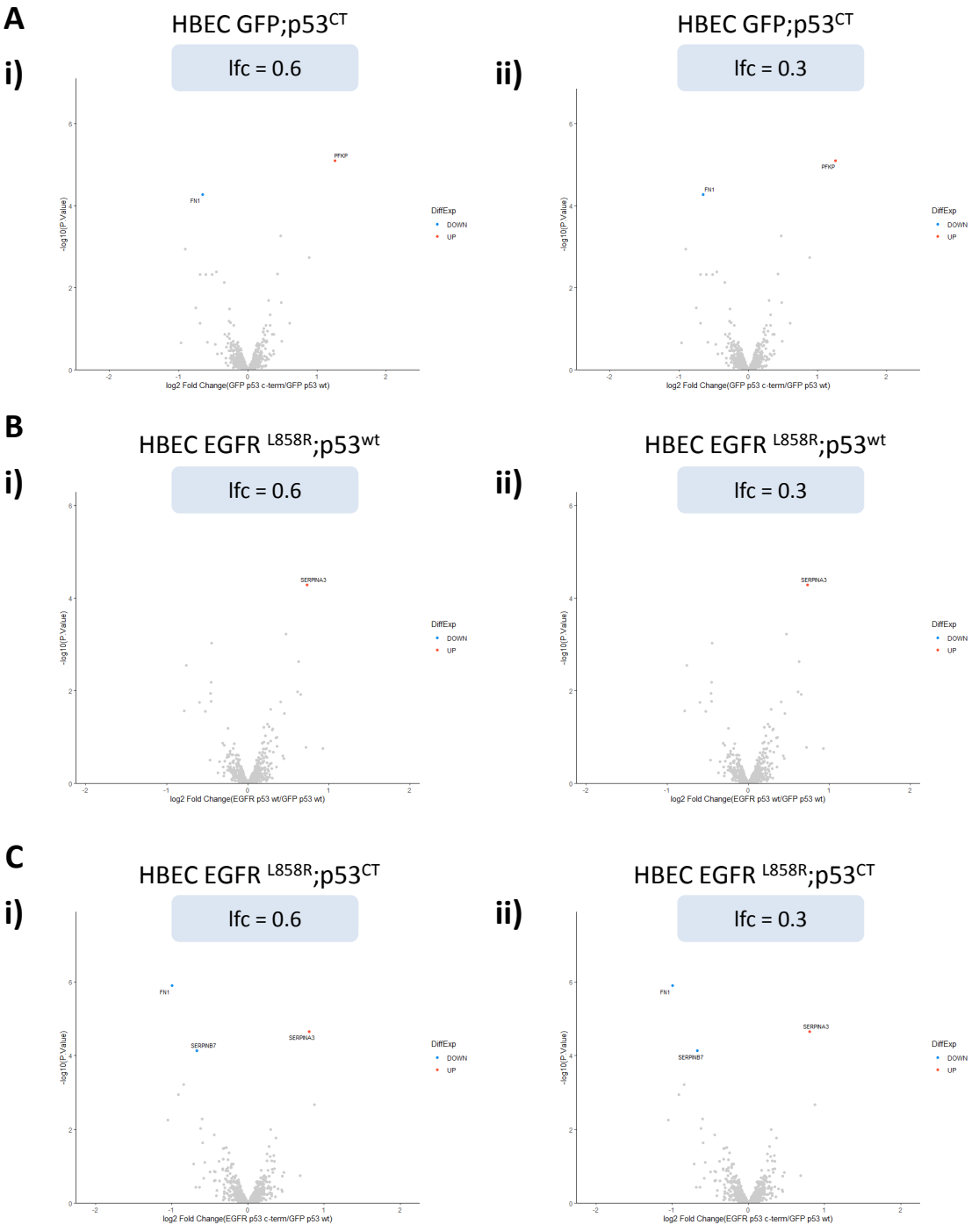


Figure S.2 Volcano plots of differentially expressed proteins of each intragroup comparison made between HBEC cell lines and HBEC GFP;p53^{wt}.
Adj p val < 0.05, variable $\log_2$ fold change (LFC). The top 20 proteins sorted by p value are labeled. (A) HBEC GFP;p53^{CT} (i) LFC > 0.6. (ii) LFC > 0.3. (B) HBEC EGFR^{L858R};p53^{wt} (i) LFC > 0.6. (ii) LFC > 0.3. (B) HBEC EGFR^{L858R};p53^{CT}(i) LFC > 0.6. (ii) LFC > 0.3.

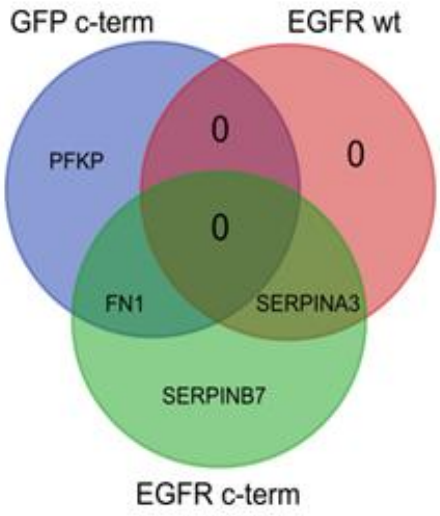


Figure S.3 Venn diagram of shared differentially expressed proteins between HBEC intragroup comparisons. Results are consistent regardless of LFC parameters. "EGFR" denotes $EGFR^{L858R}$, "wt" denotes $p53^{wt}$, and "c-term" denotes $p53^{CT}$.

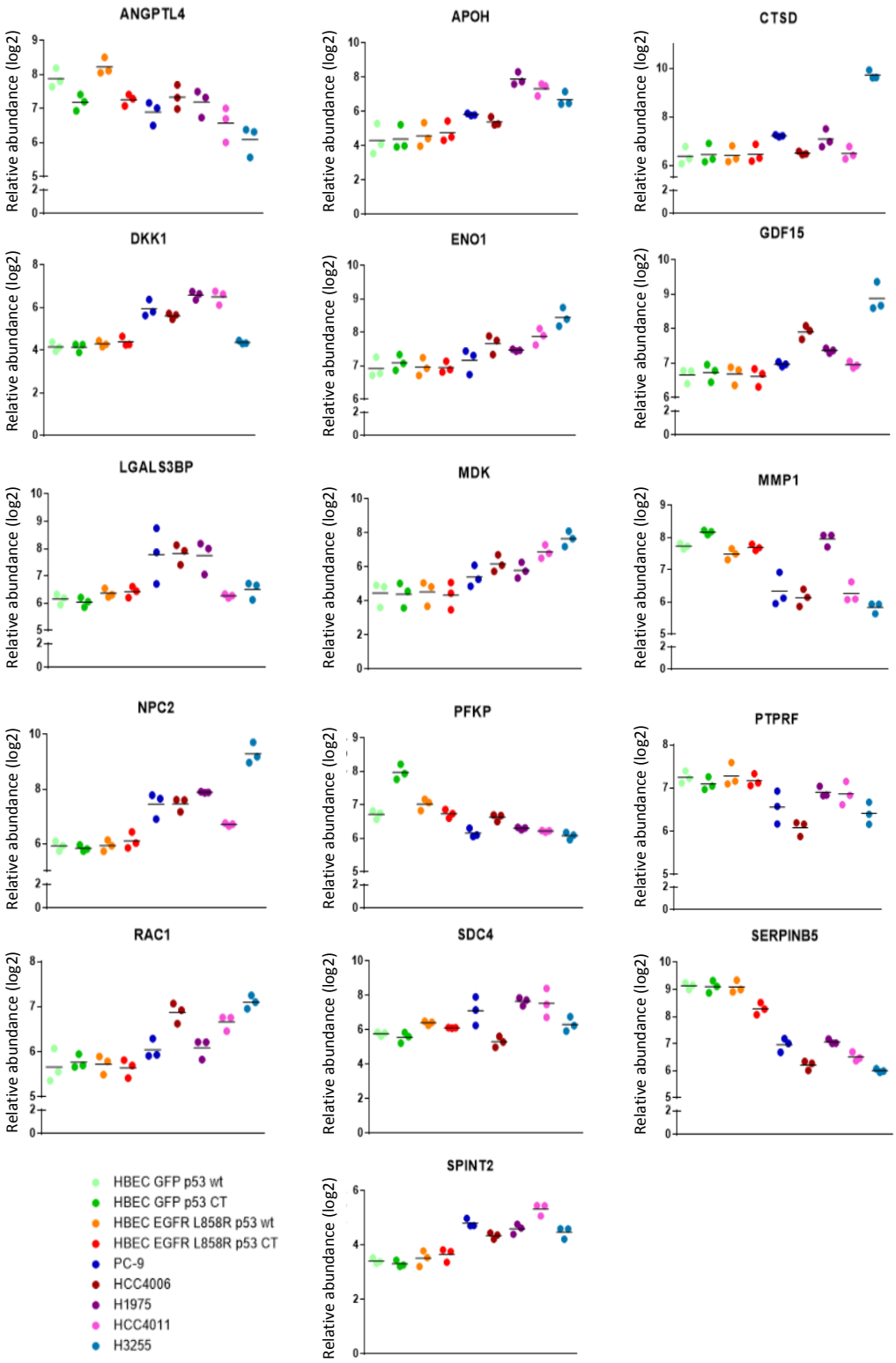


Figure S.4 TMT log_2 relative abundance plots of validated differentially expressed proteins from Table 3.1.

A.2 Supplemental methods

A

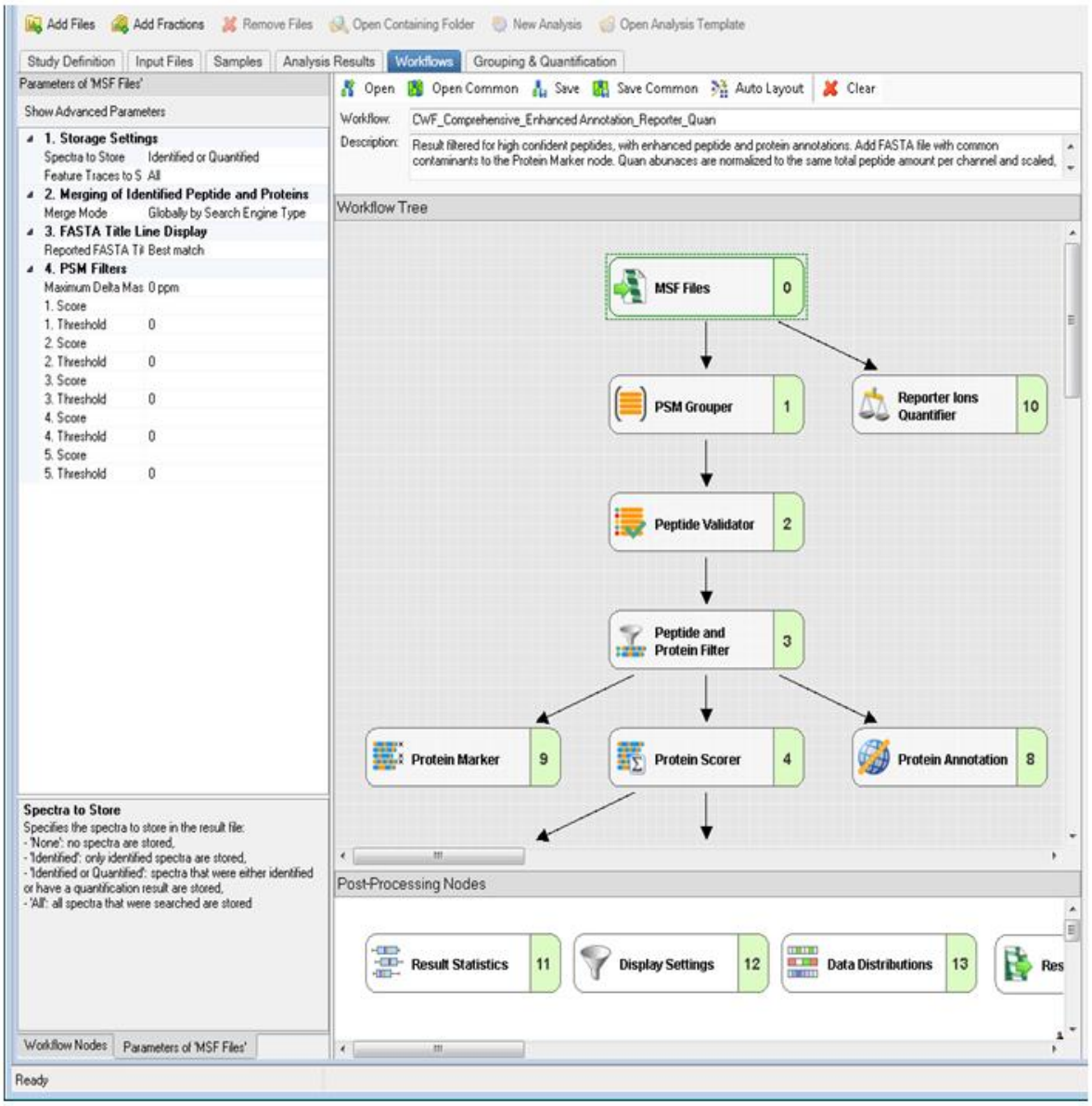
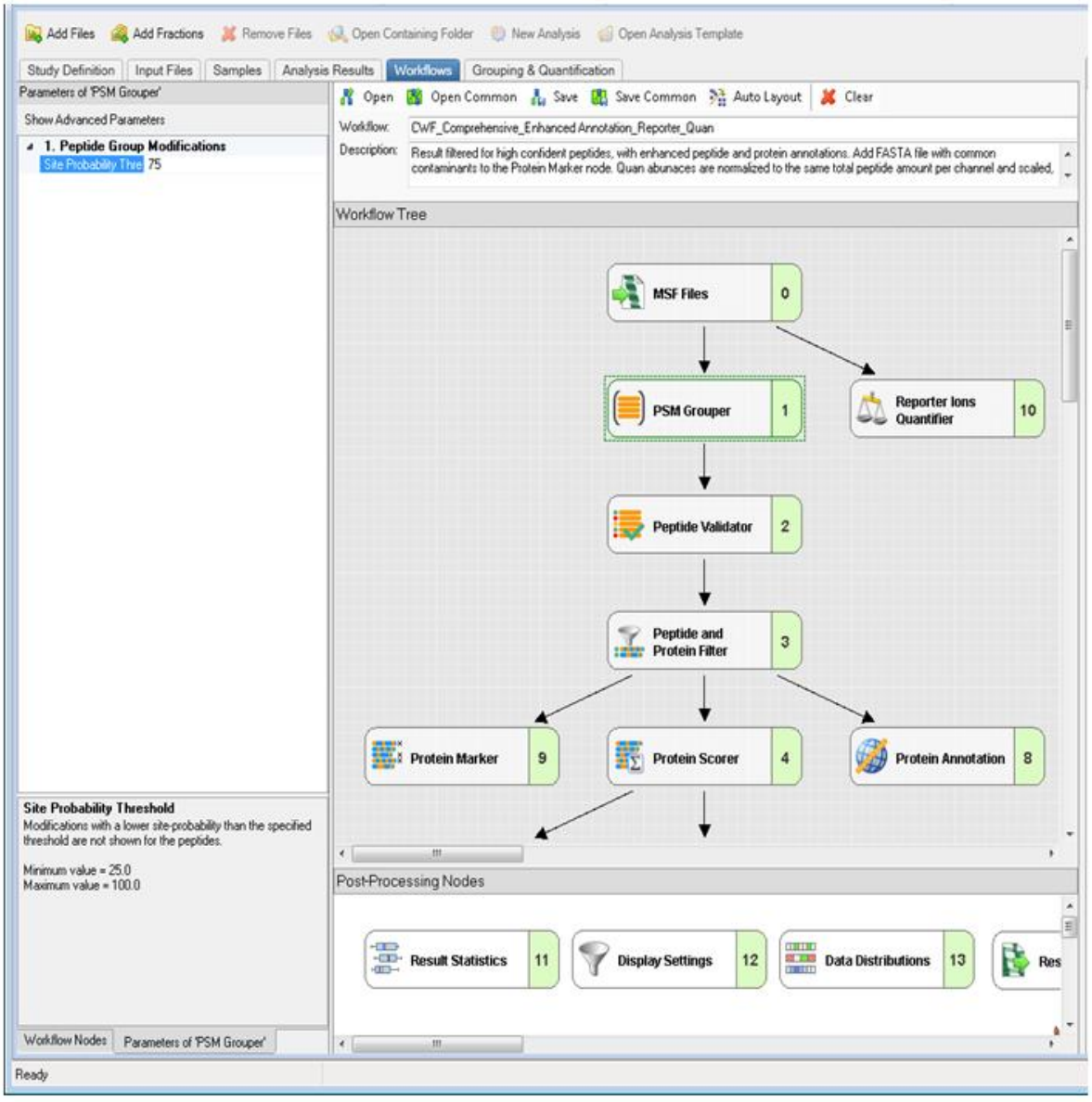
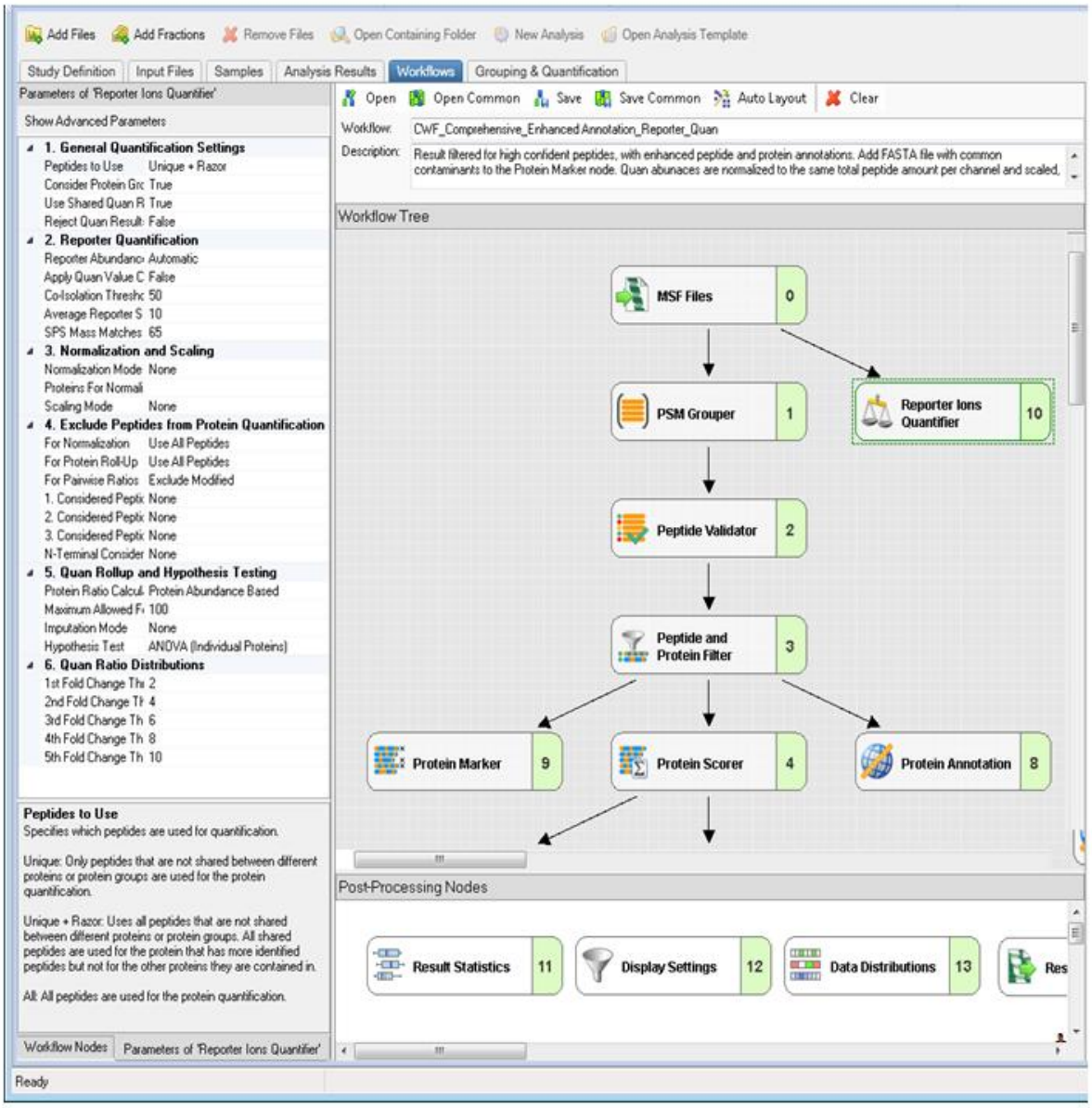


Figure S.5 Proteome Discoverer MS/MS search settings.
(A) MSF Files parameters. (B) PSM Grouper parameters. (C) Reporter Ions Quantifier parameters. (D) Peptide Validator parameters. (E) Peptide and Protein Filter parameters. (F) Protein Marker parameters. (G) Protein Annotation Marker parameters. (H) Protein Grouping parameters. (I) Protein FDR Validator parameters. (J) Peptide in Protein Annotation parameters. (K) Reporter Ions Quantifier tolerance parameters. (L) Spectrum Selector parameters. (M) Sequest HT parameters. (N) Sequest HT parameters, continued. (O) Percolator parameters.

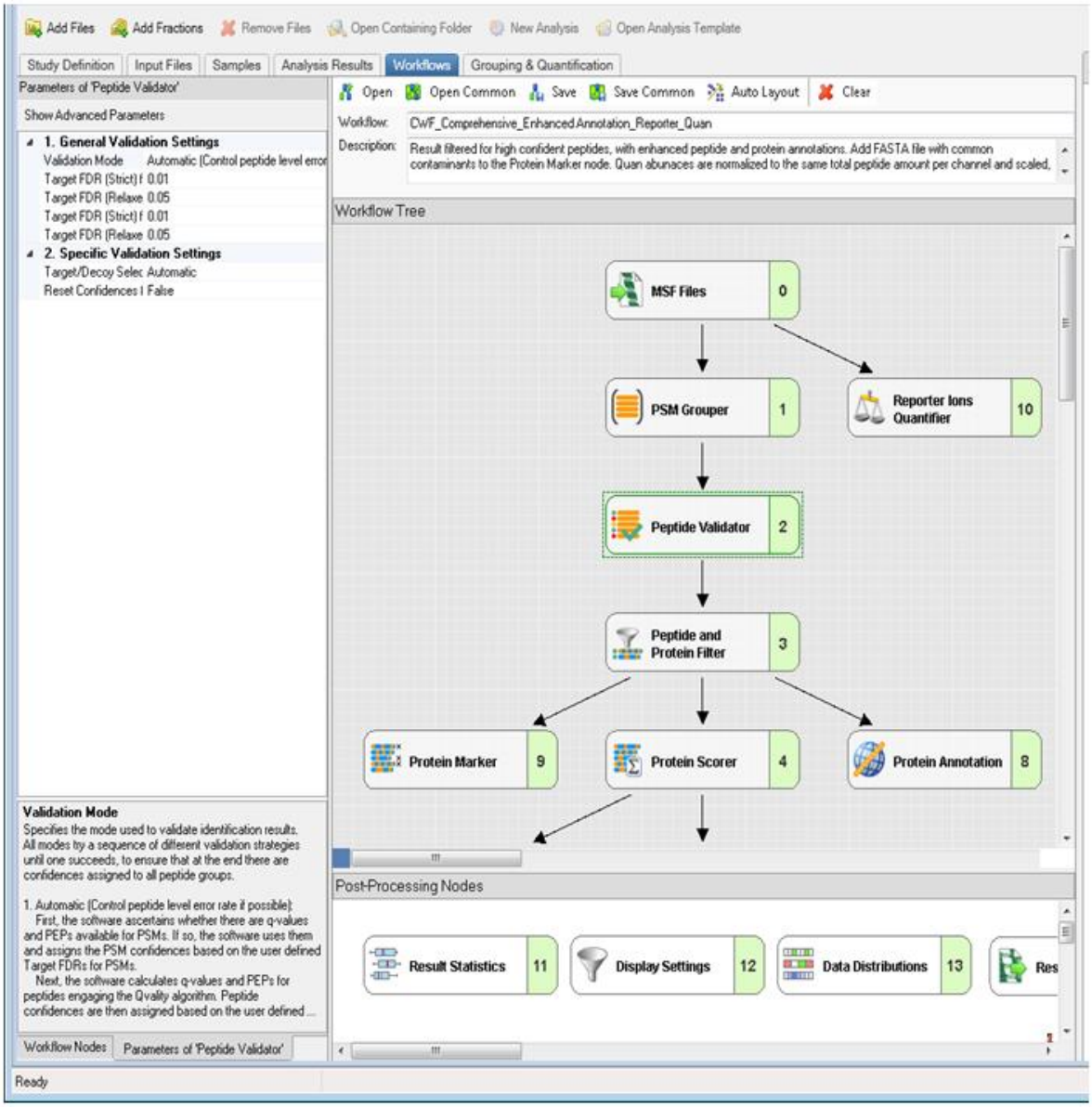
B


Add Files
Add Fractions
Remove Files
Open Containing Folder
New Analysis
Open Analysis Template
Study Definition
Input Files
Samples
Analysis Results
Workflows
Grouping & Quantification
Parameters of 'PSM Grouper'
Show Advanced Parameters
1. Peptide Group Modifications
Site Probability Thre 75
Open
Open Common
Save
Save Common
Auto Layout
Clear
Workflow:
CWF_Comprehensive_Enhanced Annotation_Reporter_Quan
Description:
Result filtered for high confident peptides, with enhanced peptide and protein annotations. Add FASTA file with common contaminants to the Protein Marker node. Quan abunaces are normalized to the same total peptide amount per channel and scaled.
Workflow Tree
MSF Files 0
PSM Grouper 1
Reporter Ions Quantifier 10
Peptide Validator 2
Peptide and Protein Filter 3
Protein Marker 9
Protein Scorer 4
Protein Annotation 8
Post-Processing Nodes
Result Statistics 11
Display Settings 12
Data Distributions 13
Site Probability Threshold
Modifications with a lower site-probability than the specified threshold are not shown for the peptides.
Minimum value = 25.0
Maximum value = 100.0
Workflow Nodes
Parameters of 'PSM Grouper'
Ready

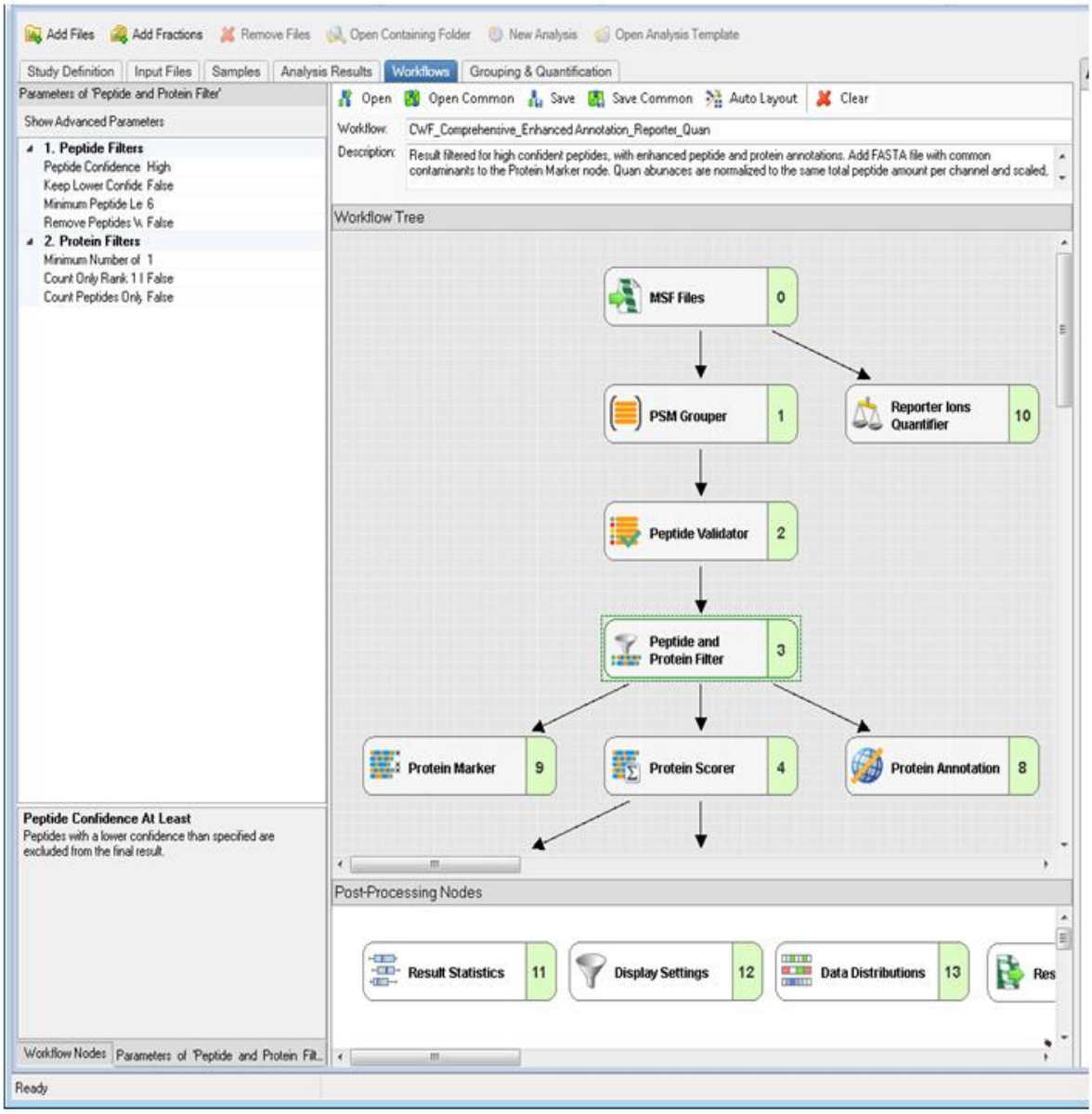
C


Add Files
Add Fractions
Remove Files
Open Containing Folder
New Analysis
Open Analysis Template
Study Definition
Input Files
Samples
Analysis Results
Workflows
Grouping & Quantification
Parameters of 'Reporter Ions Quantifier'
Show Advanced Parameters
1. General Quantification Settings
Peptides to Use Unique + Razor
Consider Protein Grc True
Use Shared Quan R True
Reject Quan Result: False
2. Reporter Quantification
Reporter Abundanc Automatic
Apply Quan Value C False
Co-Isolation Threshc 50
Average Reporter S 10
SPS Mass Matches 65
3. Normalization and Scaling
Normalization Mode None
Proteins For Normali
Scaling Mode None
4. Exclude Peptides from Protein Quantification
For Normalization Use All Peptides
For Protein Roll-Up Use All Peptides
For Pairwise Ratios Exclude Modified
1. Considered Peptic None
2. Considered Peptic None
3. Considered Peptic None
N-Terminal Consider None
5. Quan Rollup and Hypothesis Testing
Protein Ratio Calcul Protein Abundance Based
Maximum Allowed F 100
Imputation Mode None
Hypothesis Test ANOVA (Individual Proteins)
6. Quan Ratio Distributions
1st Fold Change Thr 2
2nd Fold Change Th 4
3rd Fold Change Th 6
4th Fold Change Th 8
5th Fold Change Th 10
Peptides to Use
Specifies which peptides are used for quantification.
Unique: Only peptides that are not shared between different proteins or protein groups are used for the protein quantification.
Unique + Razor: Uses all peptides that are not shared between different proteins or protein groups. All shared peptides are used for the protein that has more identified peptides but not for the other proteins they are contained in.
All: All peptides are used for the protein quantification.
Workflow Nodes
Parameters of 'Reporter Ions Quantifier'
Open
Open Common
Save
Save Common
Auto Layout
Clear
Workflow:
CWF_Comprehensive_Enhanced Annotation_Reporter_Quan
Description:
Result filtered for high confident peptides, with enhanced peptide and protein annotations. Add FASTA file with common contaminants to the Protein Marker node. Quan abunaces are normalized to the same total peptide amount per channel and scaled.
Workflow Tree
MSF Files 0
PSM Grouper 1
Reporter Ions Quantifier 10
Peptide Validator 2
Peptide and Protein Filter 3
Protein Marker 9
Protein Scorer 4
Protein Annotation 8
Post-Processing Nodes
Result Statistics 11
Display Settings 12
Data Distributions 13
Ready

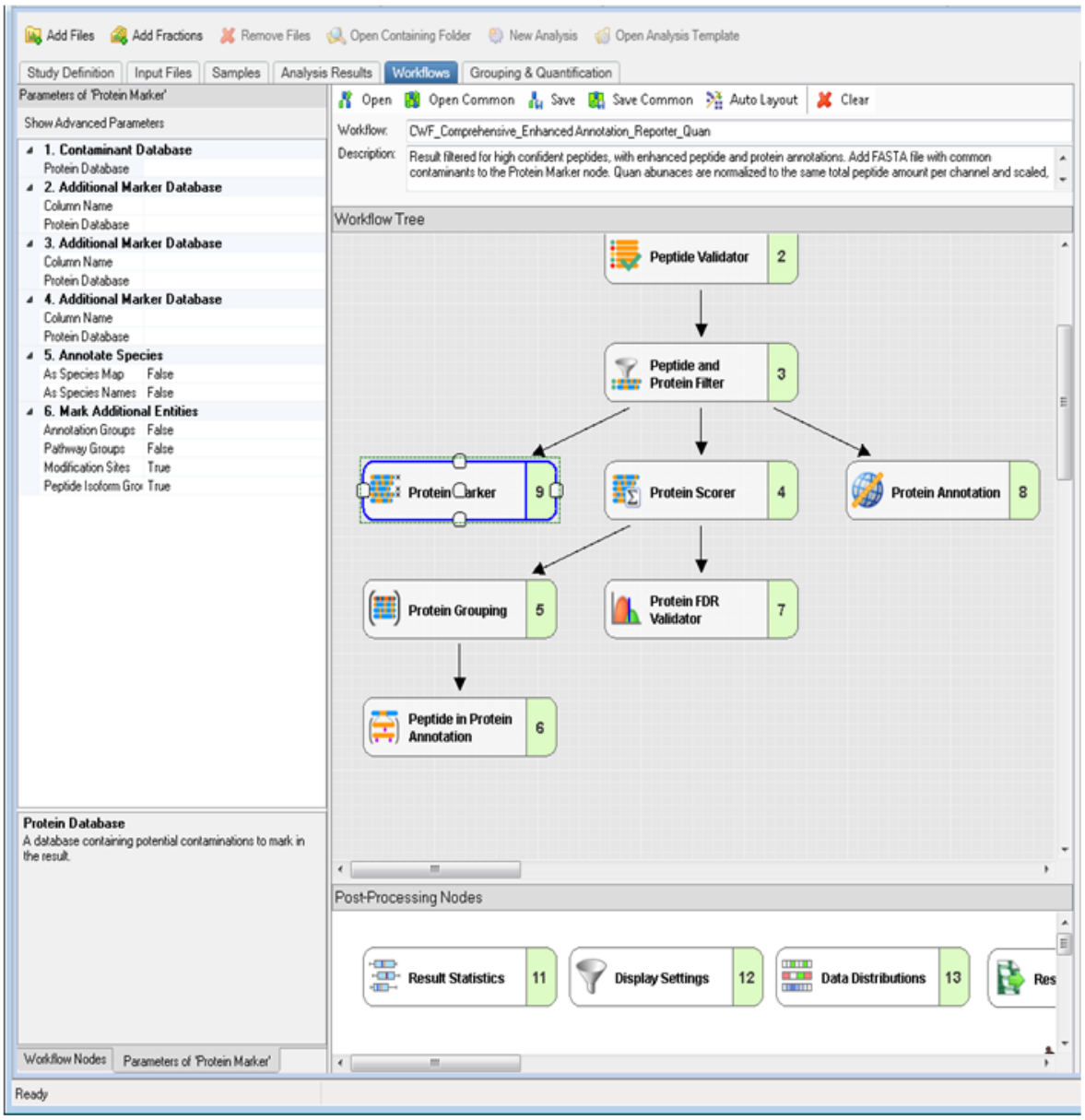
D


Add Files
Add Fractions
Remove Files
Open Containing Folder
New Analysis
Open Analysis Template
Study Definition
Input Files
Samples
Analysis Results
Workflows
Grouping & Quantification
Parameters of 'Peptide Validator'
Show Advanced Parameters
1. General Validation Settings
Validation Mode Automatic (Control peptide level error
Target FDR (Strict) f 0.01
Target FDR (Relaxe 0.05
Target FDR (Strict) f 0.01
Target FDR (Relaxe 0.05
2. Specific Validation Settings
Target/Decoy Selec Automatic
Reset Confidences I False
Validation Mode
Specifies the mode used to validate identification results. All modes try a sequence of different validation strategies until one succeeds, to ensure that at the end there are confidences assigned to all peptide groups.
1. Automatic (Control peptide level error rate if possible):
First, the software ascertains whether there are q-values and PEPs available for PSMs. If so, the software uses them and assigns the PSM confidences based on the user defined Target FDRs for PSMs.
Next, the software calculates q-values and PEPs for peptides engaging the Qvality algorithm. Peptide confidences are then assigned based on the user defined ...
Workflow Nodes
Parameters of 'Peptide Validator'
Open
Open Common
Save
Save Common
Auto Layout
Clear
Workflow:
CWF_Comprehensive_Enhanced Annotation_Reporter_Quan
Description:
Result filtered for high confident peptides, with enhanced peptide and protein annotations. Add FASTA file with common contaminants to the Protein Marker node. Quan abunaces are normalized to the same total peptide amount per channel and scaled,
Workflow Tree
MSF Files 0
PSM Grouper 1
Reporter Ions Quantifier 10
Peptide Validator 2
Peptide and Protein Filter 3
Protein Marker 9
Protein Scorer 4
Protein Annotation 8
Post-Processing Nodes
Result Statistics 11
Display Settings 12
Data Distributions 13
Res
Ready

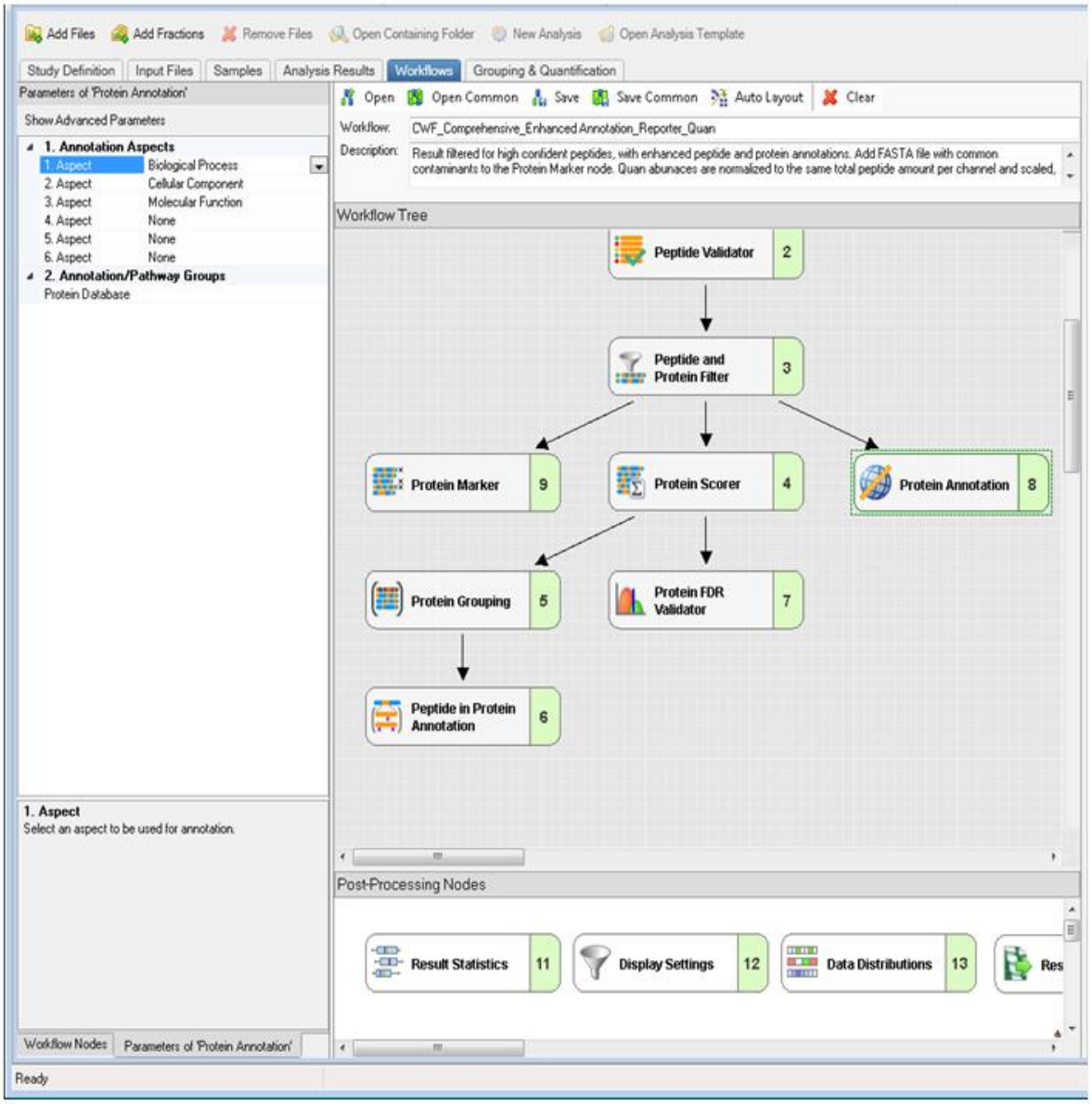
E


Add Files
Add Fractions
Remove Files
Open Containing Folder
New Analysis
Open Analysis Template
Study Definition
Input Files
Samples
Analysis Results
Workflows
Grouping & Quantification
Parameters of 'Peptide and Protein Filter'
Show Advanced Parameters
1. Peptide Filters
Peptide Confidence High
Keep Lower Confide False
Minimum Peptide Le 6
Remove Peptides \v. False
2. Protein Filters
Minimum Number of 1
Count Only Rank 1 I False
Count Peptides Only False
Peptide Confidence At Least
Peptides with a lower confidence than specified are excluded from the final result.
Workflow Nodes
Parameters of 'Peptide and Protein Filt...
Open
Open Common
Save
Save Common
Auto Layout
Clear
Workflow:
CWF_Comprehensive_Enhanced Annotation_Reporter_Quan
Description:
Result filtered for high confident peptides, with enhanced peptide and protein annotations. Add FASTA file with common contaminants to the Protein Marker node. Quan abunaces are normalized to the same total peptide amount per channel and scaled.
Workflow Tree
MSF Files 0
PSM Grouper 1
Reporter Ions Quantifier 10
Peptide Validator 2
Peptide and Protein Filter 3
Protein Marker 9
Protein Scorer 4
Protein Annotation 8
Post-Processing Nodes
Result Statistics 11
Display Settings 12
Data Distributions 13
Ready

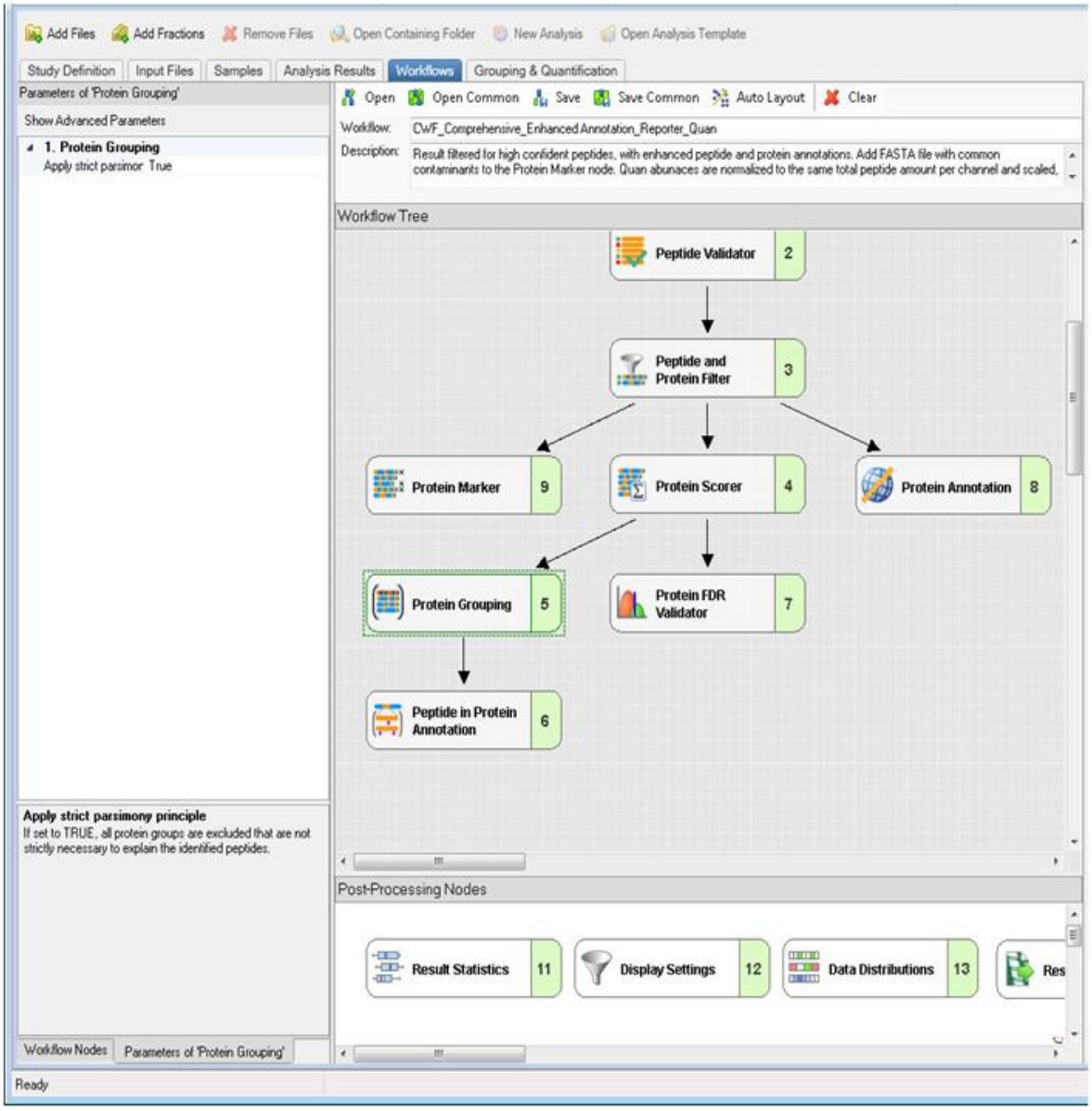
F


Add Files
Add Fractions
Remove Files
Open Containing Folder
New Analysis
Open Analysis Template
Study Definition
Input Files
Samples
Analysis Results
Workflows
Grouping & Quantification
Parameters of 'Protein Marker'
Show Advanced Parameters
1. Contaminant Database
Protein Database
2. Additional Marker Database
Column Name
Protein Database
3. Additional Marker Database
Column Name
Protein Database
4. Additional Marker Database
Column Name
Protein Database
5. Annotate Species
As Species Map False
As Species Names False
6. Mark Additional Entities
Annotation Groups False
Pathway Groups False
Modification Sites True
Peptide Isoform Gro True
Protein Database
A database containing potential contaminations to mark in the result.
Workflow Nodes
Parameters of 'Protein Marker'
Open
Open Common
Save
Save Common
Auto Layout
Clear
Workflow:
CWF_Comprehensive_Enhanced Annotation_Reporter_Quan
Description:
Result filtered for high confident peptides, with enhanced peptide and protein annotations. Add FASTA file with common contaminants to the Protein Marker node. Quan abunaces are normalized to the same total peptide amount per channel and scaled.
Workflow Tree
Peptide Validator 2
Peptide and Protein Filter 3
Protein Marker 9
Protein Scorer 4
Protein Annotation 8
Protein Grouping 5
Protein FDR Validator 7
Peptide in Protein Annotation 6
Post-Processing Nodes
Result Statistics 11
Display Settings 12
Data Distributions 13
Ready

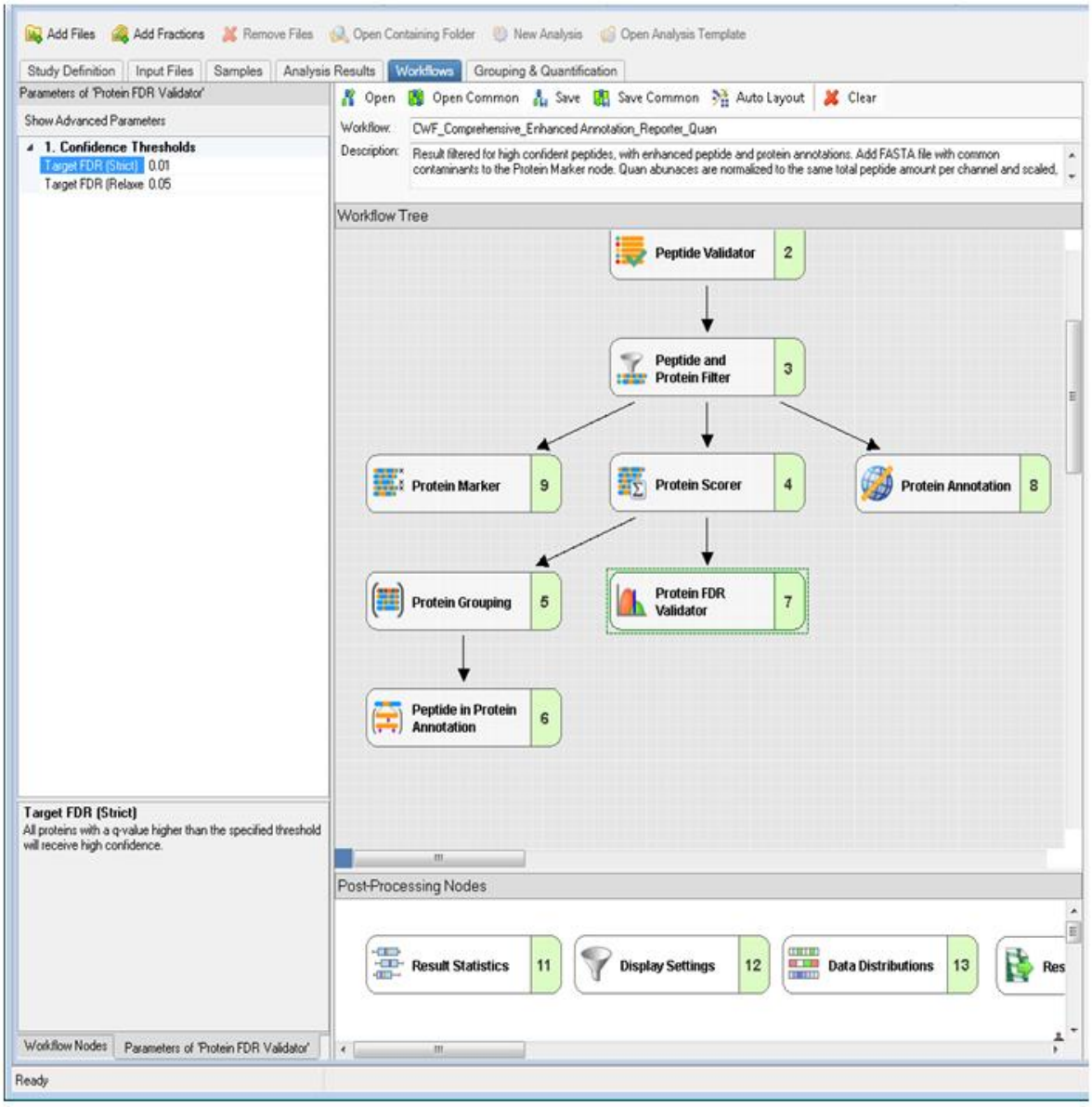
G


Add Files
Add Fractions
Remove Files
Open Containing Folder
New Analysis
Open Analysis Template
Study Definition
Input Files
Samples
Analysis Results
Workflows
Grouping & Quantification
Parameters of 'Protein Annotation'
Show Advanced Parameters
1. Annotation Aspects
1. Aspect Biological Process
2. Aspect Cellular Component
3. Aspect Molecular Function
4. Aspect None
5. Aspect None
6. Aspect None
2. Annotation/Pathway Groups
Protein Database
1. Aspect
Select an aspect to be used for annotation.
Workflow Nodes
Parameters of 'Protein Annotation'
Open
Open Common
Save
Save Common
Auto Layout
Clear
Workflow:
CWF_Comprehensive_Enhanced Annotation_Reporter_Quan
Description:
Result filtered for high confident peptides, with enhanced peptide and protein annotations. Add FASTA file with common contaminants to the Protein Marker node. Quan abunaces are normalized to the same total peptide amount per channel and scaled.
Workflow Tree
Peptide Validator 2
Peptide and Protein Filter 3
Protein Marker 9
Protein Scorer 4
Protein Annotation 8
Protein Grouping 5
Protein FDR Validator 7
Peptide in Protein Annotation 6
Post-Processing Nodes
Result Statistics 11
Display Settings 12
Data Distributions 13
Ready

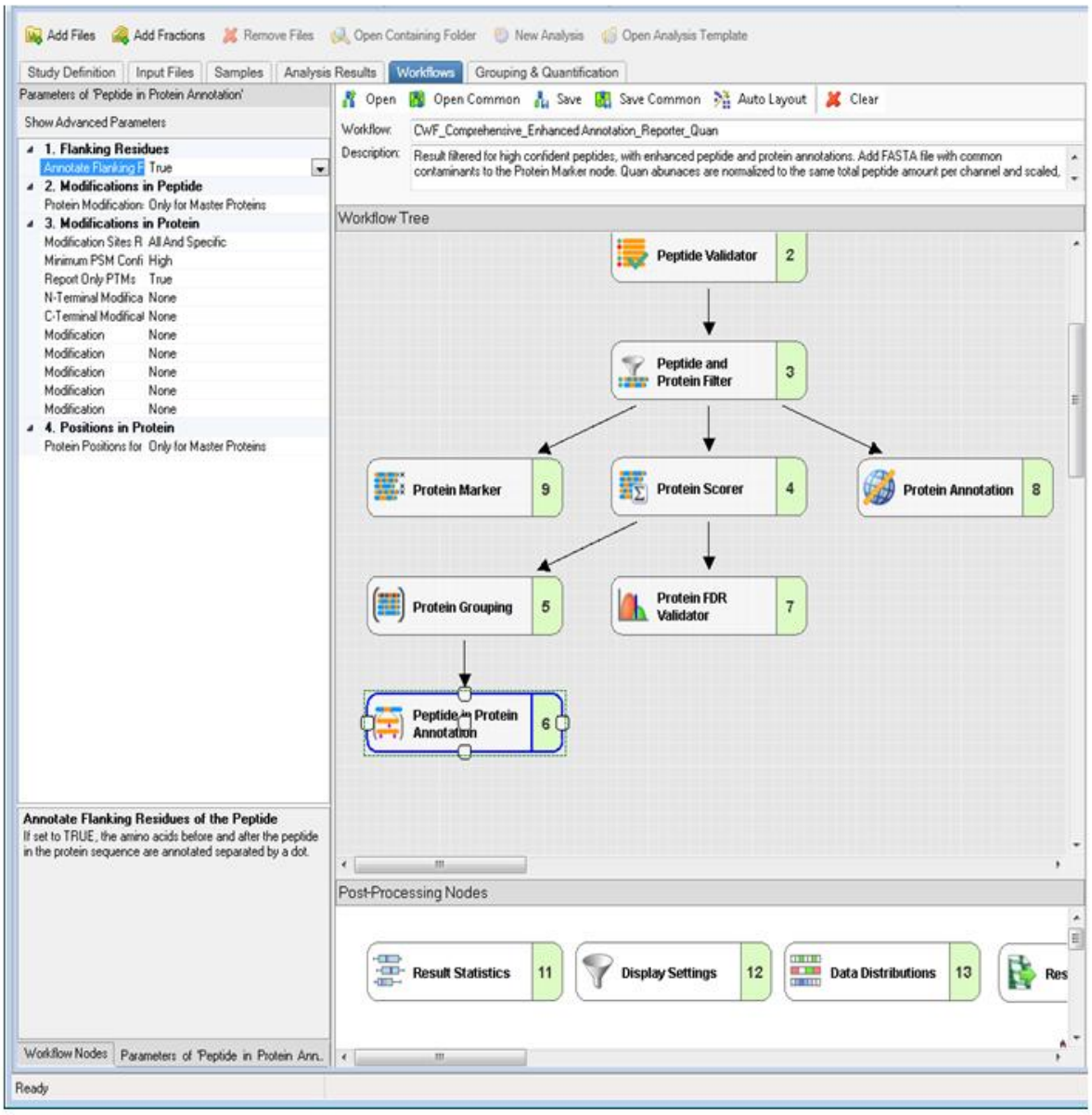
H


Add Files
Add Fractions
Remove Files
Open Containing Folder
New Analysis
Open Analysis Template
Study Definition
Input Files
Samples
Analysis Results
Workflows
Grouping & Quantification
Parameters of 'Protein Grouping'
Show Advanced Parameters
1. Protein Grouping
Apply strict parsimor True
Open
Open Common
Save
Save Common
Auto Layout
Clear
Workflow:
CWF_Comprehensive_Enhanced Annotation_Reporter_Quan
Description:
Result filtered for high confident peptides, with enhanced peptide and protein annotations. Add FASTA file with common contaminants to the Protein Marker node. Quan abunaces are normalized to the same total peptide amount per channel and scaled.
Workflow Tree
Peptide Validator 2
Peptide and Protein Filter 3
Protein Marker 9
Protein Scorer 4
Protein Annotation 8
Protein Grouping 5
Protein FDR Validator 7
Peptide in Protein Annotation 6
Apply strict parsimony principle
If set to TRUE, all protein groups are excluded that are not strictly necessary to explain the identified peptides.
Post-Processing Nodes
Result Statistics 11
Display Settings 12
Data Distributions 13
Workflow Nodes
Parameters of 'Protein Grouping'
Ready

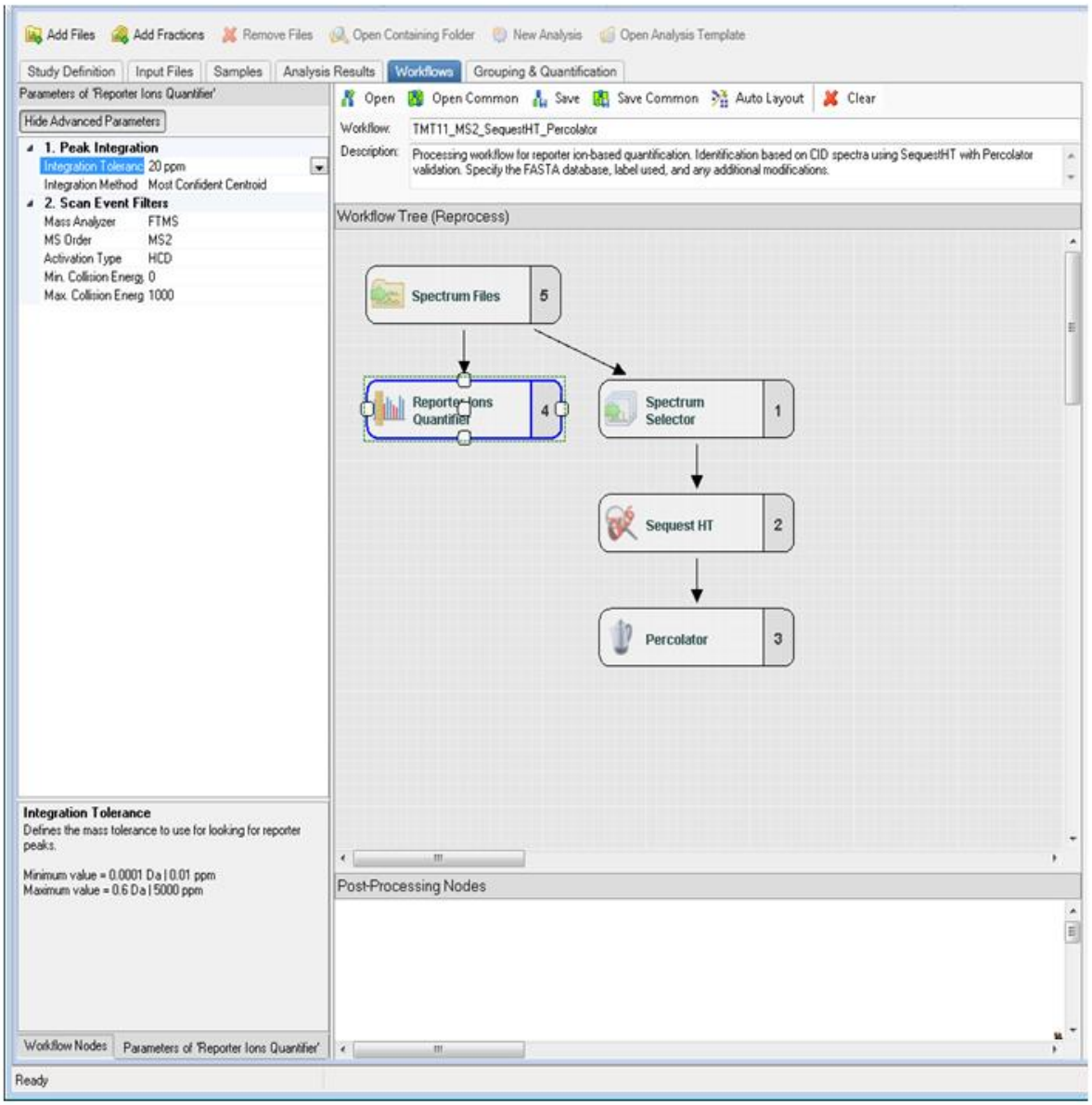
I


Add Files
Add Fractions
Remove Files
Open Containing Folder
New Analysis
Open Analysis Template
Study Definition
Input Files
Samples
Analysis Results
Workflows
Grouping & Quantification
Parameters of 'Protein FDR Validator'
Show Advanced Parameters
1. Confidence Thresholds
Target FDR (Strict) 0.01
Target FDR (Relaxe 0.05
Open
Open Common
Save
Save Common
Auto Layout
Clear
Workflow: CWF_Comprehensive_Enhanced Annotation_Reporter_Quan
Description: Result filtered for high confident peptides, with enhanced peptide and protein annotations. Add FASTA file with common contaminants to the Protein Marker node. Quan abunaces are normalized to the same total peptide amount per channel and scaled.
Workflow Tree
Peptide Validator 2
Peptide and Protein Filter 3
Protein Marker 9
Protein Scorer 4
Protein Annotation 8
Protein Grouping 5
Protein FDR Validator 7
Peptide in Protein Annotation 6
Target FDR (Strict)
All proteins with a q-value higher than the specified threshold will receive high confidence.
Post-Processing Nodes
Result Statistics 11
Display Settings 12
Data Distributions 13
Workflow Nodes
Parameters of 'Protein FDR Validator'
Ready

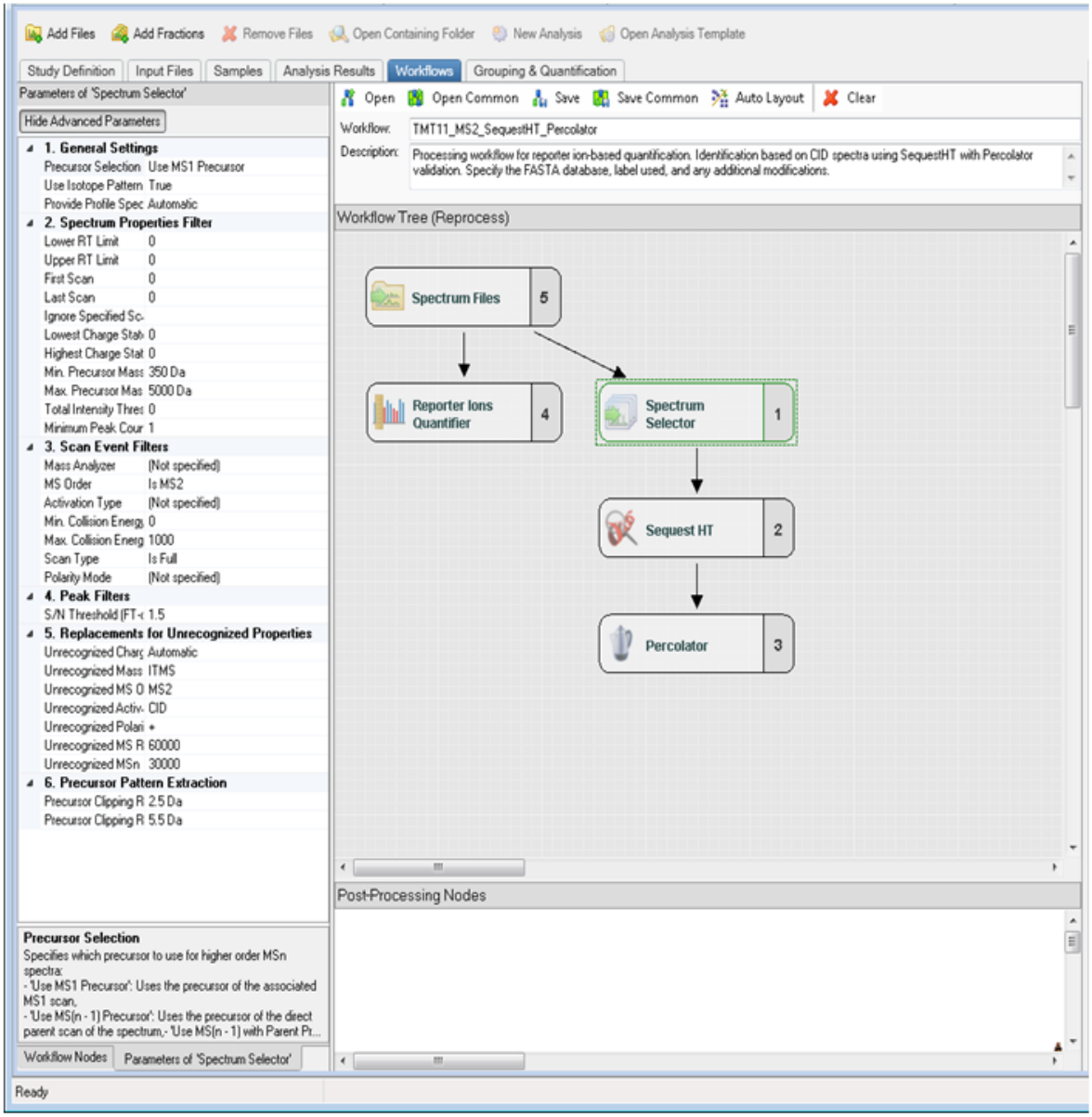
J


Add Files
Add Fractions
Remove Files
Open Containing Folder
New Analysis
Open Analysis Template
Study Definition
Input Files
Samples
Analysis Results
Workflows
Grouping & Quantification
Parameters of 'Peptide in Protein Annotation'
Show Advanced Parameters
1. Flanking Residues
Annotate Flanking F True
2. Modifications in Peptide
Protein Modification: Only for Master Proteins
3. Modifications in Protein
Modification Sites R All And Specific
Minimum PSM Confi High
Report Only PTMs True
N-Terminal Modifica None
C-Terminal Modifica None
Modification None
Modification None
Modification None
Modification None
Modification None
4. Positions in Protein
Protein Positions for Only for Master Proteins
Annotate Flanking Residues of the Peptide
If set to TRUE, the amino acids before and after the peptide in the protein sequence are annotated separated by a dot.
Workflow Nodes
Parameters of 'Peptide in Protein Ann..
Open
Open Common
Save
Save Common
Auto Layout
Clear
Workflow:
CWF_Comprehensive_Enhanced Annotation_Reporter_Quan
Description:
Result filtered for high confident peptides, with enhanced peptide and protein annotations. Add FASTA file with common contaminants to the Protein Marker node. Quan abunaces are normalized to the same total peptide amount per channel and scaled.
Workflow Tree
Peptide Validator 2
Peptide and Protein Filter 3
Protein Marker 9
Protein Scorer 4
Protein Annotation 8
Protein Grouping 5
Protein FDR Validator 7
Peptide in Protein Annotation 6
Post-Processing Nodes
Result Statistics 11
Display Settings 12
Data Distributions 13
Ready

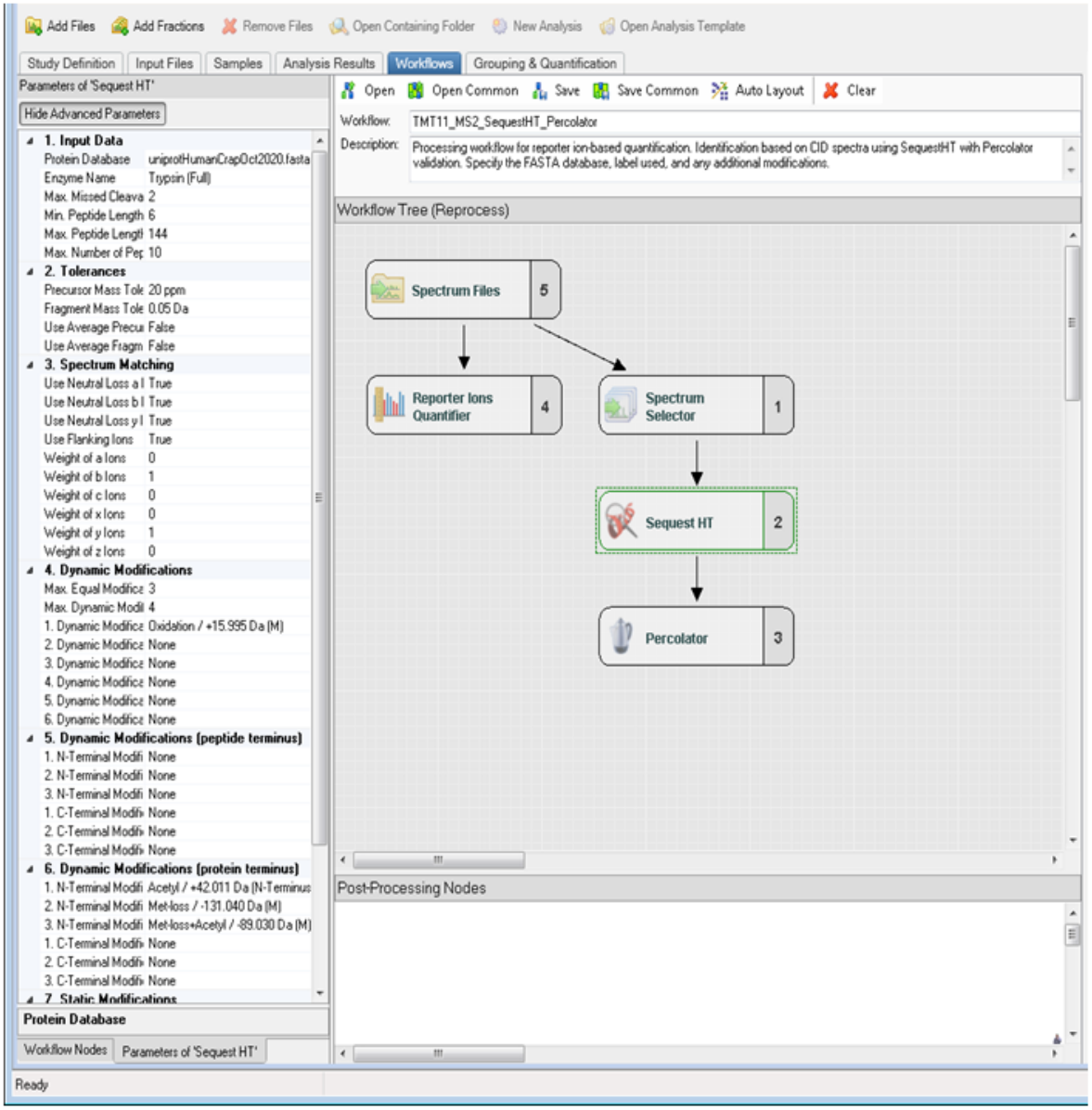
K


Add Files
Add Fractions
Remove Files
Open Containing Folder
New Analysis
Open Analysis Template
Study Definition
Input Files
Samples
Analysis Results
Workflows
Grouping & Quantification
Parameters of 'Reporter Ions Quantifier'
Hide Advanced Parameters
1. Peak Integration
Integration Tolerance 20 ppm
Integration Method Most Confident Centroid
2. Scan Event Filters
Mass Analyzer FTMS
MS Order MS2
Activation Type HCD
Min. Collision Energy 0
Max. Collision Energy 1000
Integration Tolerance
Defines the mass tolerance to use for looking for reporter peaks.
Minimum value = 0.0001 Da | 0.01 ppm
Maximum value = 0.6 Da | 5000 ppm
Workflow Nodes
Parameters of 'Reporter Ions Quantifier'
Open
Open Common
Save
Save Common
Auto Layout
Clear
Workflow:
TMT11_MS2_SequestHT_Percolator
Description:
Processing workflow for reporter ion-based quantification. Identification based on CID spectra using SequestHT with Percolator validation. Specify the FASTA database, label used, and any additional modifications.
Workflow Tree (Reprocess)
Spectrum Files
5
Reporter Ions Quantifier
4
Spectrum Selector
1
Sequest HT
2
Percolator
3
Post-Processing Nodes
Ready

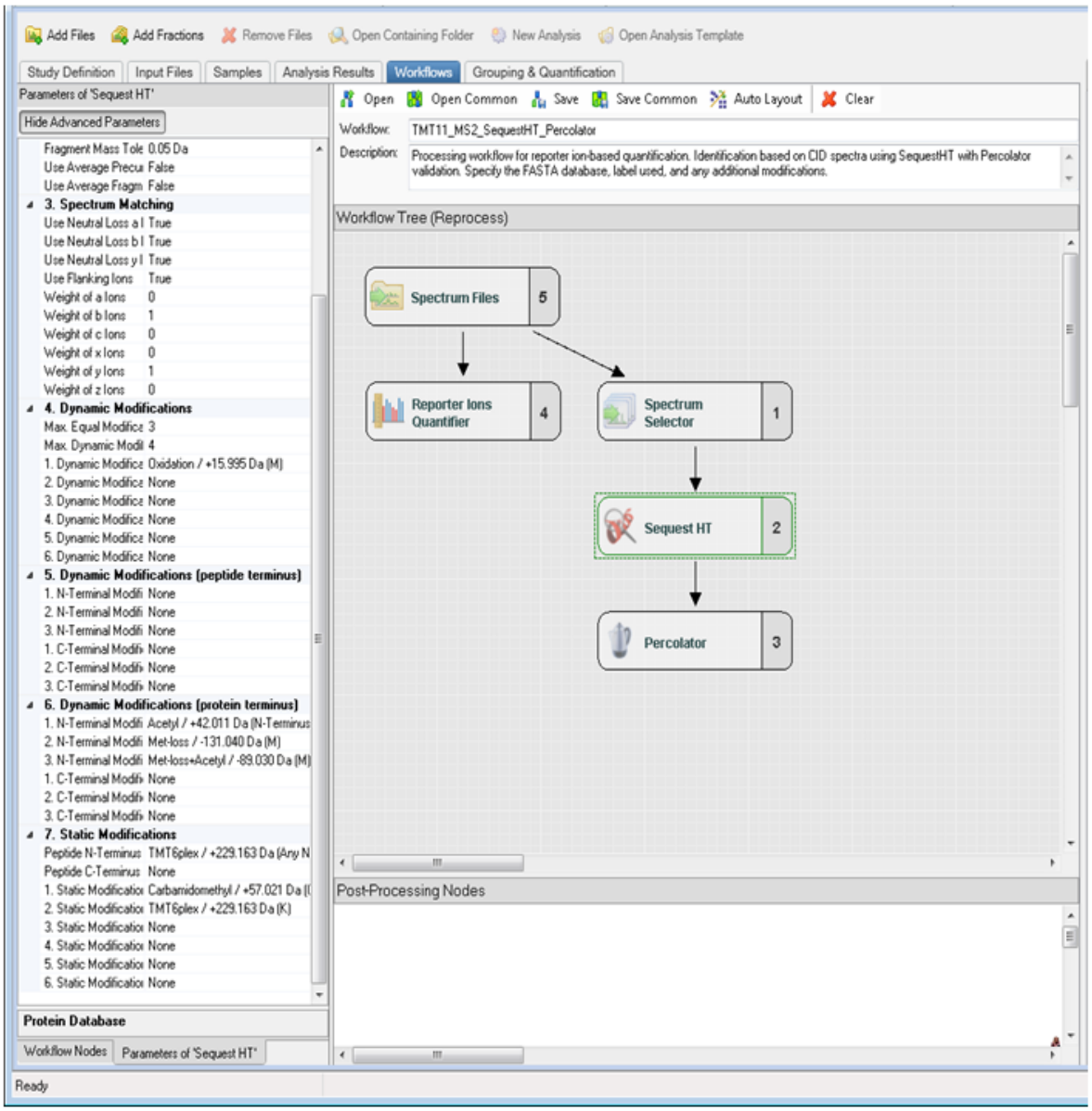
L


Add Files
Add Fractions
Remove Files
Open Containing Folder
New Analysis
Open Analysis Template
Study Definition
Input Files
Samples
Analysis Results
Workflows
Grouping & Quantification
Parameters of 'Spectrum Selector'
Hide Advanced Parameters
1. General Settings
Precursor Selection Use MS1 Precursor
Use Isotope Pattern True
Provide Profile Spec Automatic
2. Spectrum Properties Filter
Lower RT Limit 0
Upper RT Limit 0
First Scan 0
Last Scan 0
Ignore Specified Sc
Lowest Charge Stat 0
Highest Charge Stat 0
Min. Precursor Mass 350 Da
Max. Precursor Mas 5000 Da
Total Intensity Thres 0
Minimum Peak Cour 1
3. Scan Event Filters
Mass Analyzer (Not specified)
MS Order Is MS2
Activation Type (Not specified)
Min. Collision Energ 0
Max. Collision Energ 1000
Scan Type Is Full
Polarity Mode (Not specified)
4. Peak Filters
S/N Threshold (FT-(1.5
5. Replacements for Unrecognized Properties
Unrecognized Charg Automatic
Unrecognized Mass ITMS
Unrecognized MS O MS2
Unrecognized Activ CID
Unrecognized Polari +
Unrecognized MS R 60000
Unrecognized MSn 30000
6. Precursor Pattern Extraction
Precursor Clipping R 2.5 Da
Precursor Clipping R 5.5 Da
Precursor Selection
Specifies which precursor to use for higher order MSn spectra:
- 'Use MS1 Precursor': Uses the precursor of the associated MS1 scan,
- 'Use MS(n - 1) Precursor': Uses the precursor of the direct parent scan of the spectrum,- 'Use MS(n - 1) with Parent Pr...
Workflow Nodes
Parameters of 'Spectrum Selector'
Open
Open Common
Save
Save Common
Auto Layout
Clear
Workflow: TMT11_MS2_SequestHT_Percolator
Description: Processing workflow for reporter ion-based quantification. Identification based on CID spectra using SequestHT with Percolator validation. Specify the FASTA database, label used, and any additional modifications.
Workflow Tree (Reprocess)
Spectrum Files 5
Reporter Ions Quantifier 4
Spectrum Selector 1
Sequest HT 2
Percolator 3
Post-Processing Nodes
Ready

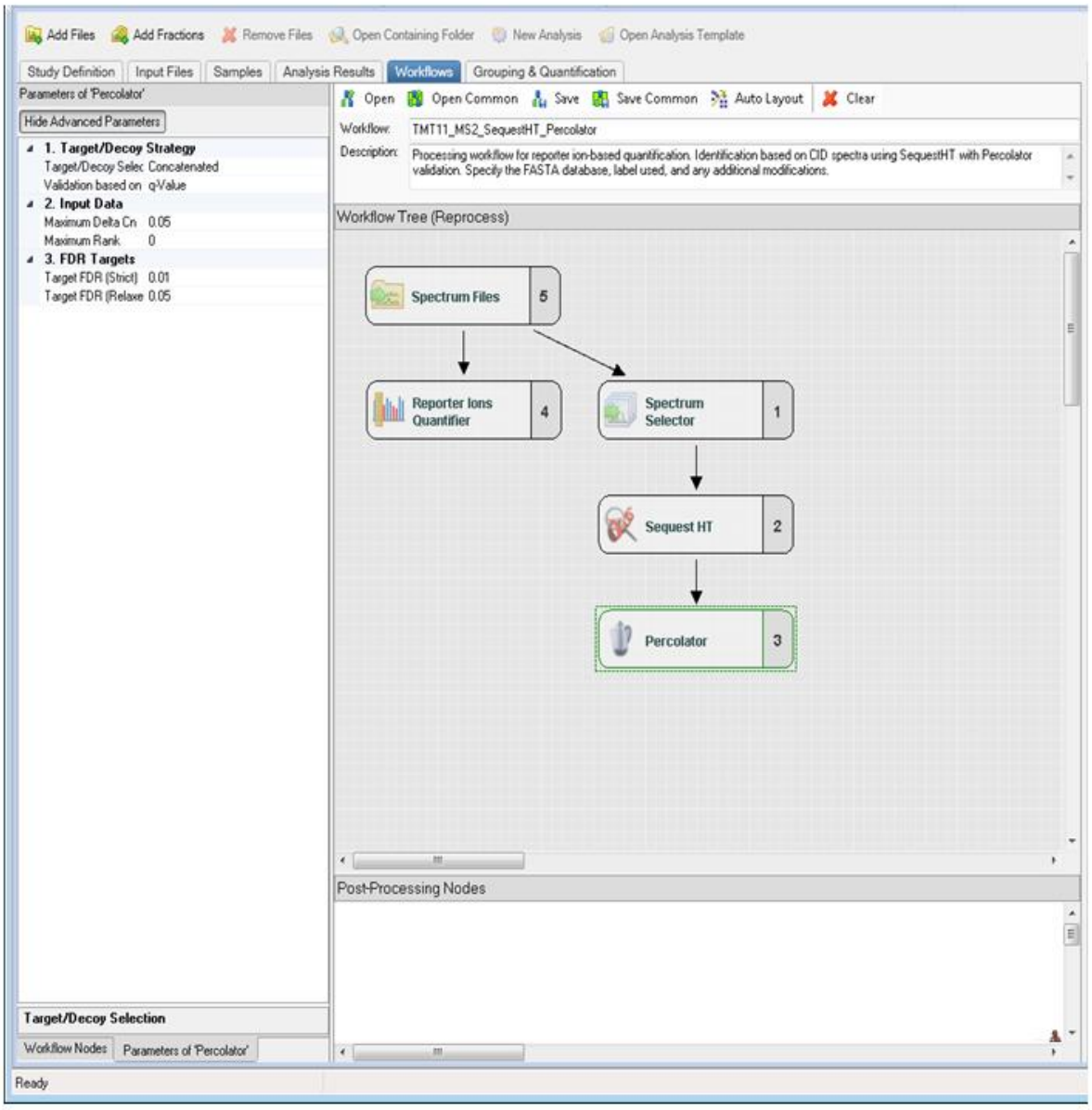
M


Add Files
Add Fractions
Remove Files
Open Containing Folder
New Analysis
Open Analysis Template
Study Definition
Input Files
Samples
Analysis Results
Workflows
Grouping & Quantification
Parameters of 'Sequest HT'
Hide Advanced Parameters
1. Input Data
Protein Database uniprotHumanCrapOct2020.fasta
Enzyme Name Trypsin (Full)
Max. Missed Cleava 2
Min. Peptide Length 6
Max. Peptide Lengtl 144
Max. Number of Pep 10
2. Tolerances
Precursor Mass Tole 20 ppm
Fragment Mass Tole 0.05 Da
Use Average Precur False
Use Average Fragm False
3. Spectrum Matching
Use Neutral Loss a I True
Use Neutral Loss b I True
Use Neutral Loss y I True
Use Flanking Ions True
Weight of a Ions 0
Weight of b Ions 1
Weight of c Ions 0
Weight of x Ions 0
Weight of y Ions 1
Weight of z Ions 0
4. Dynamic Modifications
Max. Equal Modifica 3
Max. Dynamic Modi 4
1. Dynamic Modifica Oxidation / +15.995 Da (M)
2. Dynamic Modifica None
3. Dynamic Modifica None
4. Dynamic Modifica None
5. Dynamic Modifica None
6. Dynamic Modifica None
5. Dynamic Modifications (peptide terminus)
1. N-Terminal Modifi None
2. N-Terminal Modifi None
3. N-Terminal Modifi None
1. C-Terminal Modifi None
2. C-Terminal Modifi None
3. C-Terminal Modifi None
6. Dynamic Modifications (protein terminus)
1. N-Terminal Modifi Acetyl / +42.011 Da (N-Terminus
2. N-Terminal Modifi Met-loss / -131.040 Da (M)
3. N-Terminal Modifi Met-loss+Acetyl / -89.030 Da (M)
1. C-Terminal Modifi None
2. C-Terminal Modifi None
3. C-Terminal Modifi None
7. Static Modifications
Protein Database
Workflow Nodes
Parameters of 'Sequest HT'
Open
Open Common
Save
Save Common
Auto Layout
Clear
Workflow: TMT11_MS2_SequestHT_Percolator
Description: Processing workflow for reporter ion-based quantification. Identification based on CID spectra using SequestHT with Percolator validation. Specify the FASTA database, label used, and any additional modifications.
Workflow Tree (Reprocess)
Spectrum Files 5
Reporter Ions Quantifier 4
Spectrum Selector 1
Sequest HT 2
Percolator 3
Post-Processing Nodes
Ready

N


Add Files
Add Fractions
Remove Files
Open Containing Folder
New Analysis
Open Analysis Template
Study Definition
Input Files
Samples
Analysis Results
Workflows
Grouping & Quantification
Parameters of 'Sequest HT'
Hide Advanced Parameters
Fragment Mass Tole 0.05 Da
Use Average Precu False
Use Average Fragm False
3. Spectrum Matching
Use Neutral Loss a I True
Use Neutral Loss b I True
Use Neutral Loss y I True
Use Flanking Ions True
Weight of a Ions 0
Weight of b Ions 1
Weight of c Ions 0
Weight of x Ions 0
Weight of y Ions 1
Weight of z Ions 0
4. Dynamic Modifications
Max. Equal Modifica 3
Max. Dynamic Modif 4
1. Dynamic Modifica Oxidation / +15.995 Da (M)
2. Dynamic Modifica None
3. Dynamic Modifica None
4. Dynamic Modifica None
5. Dynamic Modifica None
6. Dynamic Modifica None
5. Dynamic Modifications (peptide terminus)
1. N-Terminal Modifi None
2. N-Terminal Modifi None
3. N-Terminal Modifi None
1. C-Terminal Modifi None
2. C-Terminal Modifi None
3. C-Terminal Modifi None
6. Dynamic Modifications (protein terminus)
1. N-Terminal Modifi Acetyl / +42.011 Da (N-Terminus
2. N-Terminal Modifi Met-loss / -131.040 Da (M)
3. N-Terminal Modifi Met-loss+Acetyl / -89.030 Da (M)
1. C-Terminal Modifi None
2. C-Terminal Modifi None
3. C-Terminal Modifi None
7. Static Modifications
Peptide N-Terminus TMT6plex / +229.163 Da (Any N
Peptide C-Terminus None
1. Static Modificatio Carbamidomethyl / +57.021 Da (
2. Static Modificatio TMT6plex / +229.163 Da (K)
3. Static Modificatio None
4. Static Modificatio None
5. Static Modificatio None
6. Static Modificatio None
Protein Database
Workflow Nodes
Parameters of 'Sequest HT'
Open
Open Common
Save
Save Common
Auto Layout
Clear
Workflow:
TMT11_MS2_SequestHT_Percolator
Description:
Processing workflow for reporter ion-based quantification. Identification based on CID spectra using SequestHT with Percolator validation. Specify the FASTA database, label used, and any additional modifications.
Workflow Tree (Reprocess)
Spectrum Files
5
Reporter Ions Quantifier
4
Spectrum Selector
1
Sequest HT
2
Percolator
3
Post-Processing Nodes
Ready

o


Add Files
Add Fractions
Remove Files
Open Containing Folder
New Analysis
Open Analysis Template
Study Definition
Input Files
Samples
Analysis Results
Workflows
Grouping & Quantification
Parameters of 'Percolator'
Hide Advanced Parameters
1. Target/Decoy Strategy
Target/Decoy Selec Concatenated
Validation based on q-Value
2. Input Data
Maximum Delta Cn 0.05
Maximum Rank 0
3. FDR Targets
Target FDR (Strict) 0.01
Target FDR (Relaxe 0.05
Target/Decoy Selection
Workflow Nodes
Parameters of 'Percolator'
Open
Open Common
Save
Save Common
Auto Layout
Clear
Workflow:
TMT11_MS2_SequestHT_Percolator
Description:
Processing workflow for reporter ion-based quantification. Identification based on CID spectra using SequestHT with Percolator validation. Specify the FASTA database, label used, and any additional modifications.
Workflow Tree (Reprocess)
Spectrum Files
5
Reporter Ions Quantifier
4
Spectrum Selector
1
Sequest HT
2
Percolator
3
Post-Processing Nodes
Ready