Application of Proteomics to Identify Proteins Involved in Pancreatic Cancer and to Improve Biotherapeutic Production in Chinese Hamster Ovary Cells

A thesis submitted for the degree of Ph.D.

Orla Coleman B.Sc. (Hons)

January 2019

The research described in this thesis was conducted under the supervision of

Dr. Paula Meleady and Prof. Martin Clynes.

National Institute for Cellular Biotechnology,

School of Biotechnology,

Dublin City University

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Ph.D. is entirely my own work, that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: _____ (Candidate) ID No.: 14212020

Date: _____

This thesis is dedicated to my loving family

Acknowledgements

I would like to thank my two supervisors Dr. Paula Meleady and Prof. Martin Clynes. Paula, you have guided me through the last 4 years and taught me how to be an independent researcher. You have supported me in every aspect, from my PhD worries to my personal life. I can never thank you enough for giving me this opportunity, the proteomics group would not exist without you. I sincerely hope to remain in contact with you in the future.

Martin, always a friendly smile and a silly joke to be had no matter what problems we had to face. I will never forget the email from you the evening after my PhD interview- you told me I am unlikely to get the position, but you offered me RA work for a few months to get the lab experience I needed. Thankfully Paula took a risk and did choose me, however your kindness and desire to help others that was so obvious that day has never dulled. Even in retirement you are always thinking of your students, with 3am emails and seminar talks you gently push everybody to reach their full potential. Thank you for everything you have done for me, I am forever grateful.

To the senior researchers who without this work would not have been possible. Fiona and Sandra, the backbone of the pancreatic cancer research group. You both have been a huge help to me over the past 4 years and been there to answer all my questions. The PDX program is a huge asset especially to my PhD, it would not be the success that it is without both of your hard work and diligence. Fiona, thank you particularly for sharing your office with me although I wasn't there too often your chats, book recommendations and fashion advice was always appreciated.

Mick, what can I say. My love of mass spectrometry and all things proteomics is all thanks to you. I have thoroughly enjoyed being your apprentice – I have learned to plumb a HPLC, sing the alphabet quickly, share my food, and never leave spaces in the sequence table. I know we will remain friends- can't wait to graduate together!

To all my office and lunch time friends past and present. The originals, Shane, Gemma, Andrew, Prashant, Edel and Mark you paved the way and made me feel so welcome when I started. To my day ones Kevin and Alan, we started together and now we will graduate together. You both have been like brothers to me. We have each gone through the highest highs and the lowest lows of our PhD's but it was always made better having each other to talk to. To Laura, my gal pal! I won't get sappy but you have been my best friend here, every gym session, lunchtime and pharmacy walk provided the perfect break from the PhD. I will miss you but I know we will remain friends.

Finally, thanks to my family, friends and Jamie. You have all been such a great support to me and I wouldn't have been able to undertake the PhD without you.

Overview of publications presented in this thesis:

Publication title	**Journal and status**	**Authors**	**Contribution**
Proteomic strategies in the search for novel pancreatic cancer biomarkers and drug targets: recent advances and clinical impact	*Expert Review of Proteomics* Published March 2016 Volume 13, No.4, pp 383-394.	Orla Coleman, Henry, M., McVey, G., Clynes, M., Moriarty, M. and Meleady, P.	First author, investigation and manuscript preparation.
A comparative quantitative LC-MS/MS profiling analysis of human pancreatic adenocarcinoma, adjacent-normal tissue, and patient-derived tumour xenografts	*Proteomes* Published November 2018 Volume 6, issue 4, article number 45	Orla Coleman, Henry, M., O'Neill, F., Roche, S., Swan, N., Boyle, L., Murphy, J., Meiller, J., Conlon, N.T., Geoghegan, J., Conlon, K., Lynch, V., Straubinger, N.L., Straubinger, R.M. McVey, G., Moriarty, M., Meleady, P. and Clynes, M.	First author, investigation, methodology, and manuscript preparation

Filter-Aided Sample Preparation (FASP) for improved proteome analysis of recombinant Chinese hamster ovary cells	*Methods in Molecular Biology* Published May 2017 Volume 1603, pp 187-194.	Orla Coleman, Henry, M., Clynes, M. and Meleady, P.	First author, investigation, methodology, and manuscript preparation
Depletion of endogenous miRNA-378-3p increases peak cell density of CHO DP12 cells and is correlated with elevated levels of ubiquitin carboxyl-terminal hydrolase 14	*Journal of Biotechnology* Published October 2018 Volume 288, pp 30-40	Costello, A., Orla Coleman, Lao, N.T., Henry, M., Meleady, P., Barron N. and Clynes, M.	Joint first author, investigation, and methodology
A proteomic profiling dataset of recombinant Chinese hamster ovary cells showing enhanced cellular growth following miR-378 depletion	*Data in Brief* Published December 2018 Volume 21, pp 2679-2688	Orla Coleman, Costello, A., Henry, M., Lao, N.T., Barron, N., Clynes, M. and Meleady, P.	Joint first author, investigation, methodology, and manuscript preparation

Increased growth rate and productivity following stable depletion of miR-7 in a mAb producing CHO cell line causes an increase in proteins associated with the Akt pathway and ribosome biogenesis	*Journal of Proteomics* Accepted January 2019	Orla Coleman, Suda, S., Meiller, J., Henry, M., Riedl, M., Barron, N., Meleady, P. and Clynes, M.	First author, investigation, methodology, and manuscript preparation

The role of this candidate stated here is accurate.

Primary supervisor	Signed:	Date:
Secondary supervisor	Signed:	Date:
Student	Signed:	Date:

Publications not included in this thesis:

1. **Process-relevant concentrations of the leachable bDtBPP impact negatively on CHO cell production characteristics.** Kelly PS, McSweeney S, Coleman O, Carillo S, Henry M, Chandran D, Kellett A, Bones J, Clynes M, Meleady P, Barron N. *Biotechnology Progress*. (2016). Volume 32(6) pp 1547-1558.

2. **Phosphopeptide enrichment and LC-MS/MS analysis to study the phosphoproteome of recombinant Chinese Hamster Ovary cells**. Henry M, Coleman O, Prashant, Clynes M, Meleady P. *Methods in Molecular Biology* (2017). Volume 1603 pp 195-208.

3. **Differential Phosphoproteomic Analysis of Recombinant Chinese Hamster Ovary Cells Following Temperature Shift.** Henry M, Power M, Kaushik P, Coleman O, Clynes M, Meleady P. *Journal of Proteome Research* (2017). Volume 16(7) pp 2339-2358.

Achievements & Awards

Appointed secretary for the Irish Mass Spectrometry Society 2018

ESACT UK travel grant awarded 2018

Royal society of chemistry travel grant to travel to Portugal to present at ICAP 2017

Irish Mass Spectrometry best student presentation 2016

Presentations

Internationally: AACR Boston USA 2018, PMF Manchester UK 2018, ESACT UK Leeds UK 2018, ICAP Portugal 2017 and ProteoMMX Chester UK 2016

Nationally: HUPO 2017, IMSS 2016 and 2017, SoBT 2016 and 2017

Table of contents

Abbreviations

ACN	Acetonitrile
ADC	Antibody-drug conjugate
ANOVA	Analysis of variance
ANT	Adjacent-normal tissue
BSA	Bovine serum albumin
CA	Cancer antigen
CHCA	α-Cyano-4-hydroxycinnamic acid
CHO	Chinese hamster ovary
CRIGR	Cricetulus griesus
CID	Collision induced dissociation
CE	Capillary electrophoresis
CV	Coefficient of variance
DDA	Data dependant acquisition
DIA	Data independent acquisition
DTT	Dithiothreitol
ECL	Enhanced chemiluminescence
ELISA	Enzyme-linked immunosorbent assay
ETD	Electron transfer dissociation
ESI	Electrospray ionisation
FA	Formic acid
FBS	Fetal bovine serum
FDA	Food and drug administration
FDR	False-discvoery rate
GO	Gene ontology
HCD	Higher energy C-trap dissociation
HCl	Hydrochloric acid
HPLC	High performance liquid chromatography

IPA	Ingenuity pathway analysis
KCl	Potassium chloride
KEGG	Kyoto encyclopedia of genes and genomes
iCAT	Isotope coded affinity tag
LC-MS	Liquid chromatography mass spectrometry
LTQ	Linear trap quadrupole
MAb	Monoclonal antibody
MALDI	Matrix assisted laser desorption ionisation
MiR	microRNA
MS	Mass spectrometer
m/z	Mass-to-charge ratio
NC	Negative control
NCBI	National Center for Biotechnology Information
OS	Overall survival
PBS	Phosphate buffered saline
PCA	Principal components analysis
PDAC	Pancreatic ductal adenocarcinoma
PDX	Patient-derived xenograft
PTM	Post-translational modification
PPM	Parts per million
PSM	Peptide spectrum match
qPCR	Quantitative polymerase chain reaction
SILAC	Stable isotope label by amino acids in culture
SWATH	Sequential window acquisition of all theoretical mass spectra
TFA	Trifluroacetic acid
TOF	Time-of-flight
TMT	Tandem mass tags

List of figures

List of tables

Abstract

Title: Application of proteomics to identify proteins involved in pancreatic cancer and to improve biotherapeutic production in Chinese Hamster Ovary cells

Author: Orla Coleman

This thesis focuses on the application of advanced proteomics and mass spectrometry to pancreatic cancer research and the biology of the Chinese hamster ovary (CHO) cell. Understanding these diverse cell phenotypes is important for the treatment of pancreatic cancer (PDAC) and to improve the efficiency of production of biotherapeutics. Sample preparation, mass spectrometry (using two sensitive, highly accurate and powerful machines which identified thousands of proteins per sample) and bioinformatics analysis are the key components undertaken in this thesis for these studies.

Proteomic analysis of PDAC using primary tumours, adjacent-normal tissues, PDX F1 and PDX F2 tumours was carried out. Using subcellular protein enrichment during sample preparation in the PDAC studies allowed the identification of membrane-associated proteins which may have potential as novel targeted therapeutic drugs for PDAC treatment. Bioinformatic analysis in the PDAC study identified proteins overexpressed in the primary tumour and tumour-cell associated proteins in the PDX models.

Proteomic analysis of recombinant CHO cells following microRNA engineering of miR-378 and miR-7 which improved culture phenotypes was carried out. Subcellular protein enrichment of the CHO cells enabled deeper proteomic coverage. Bioinformatic analysis in the CHO proteomic studies identified roles for Usp14, the Akt pathway and ribosome biogenesis in improved CHO cell phenotypes related to growth and productivity.

The scientific output from this thesis resulted in six publications to date. The proteomic data presented here provide a deeper understanding into the biology of both pancreatic cancer and CHO cells.

1 Chapter One: Proteomics and its applications throughout this thesis

1.1 An overview of proteomics and mass spectrometry

Proteomics is the large scale study of proteins from any organism, system or biological material at any given time. The proteome is a broad term that encompasses all of the proteins in an organ, cell or biological sample. Proteins are synthesised by translating the information encoded in RNA to a polypeptide chain which takes a specific three dimensional structure. Proteins are the principal functional molecules within a cell and are therefore more likely to reflect the phenotype of any living material. Since the completion of the human genome project in 2003, genomic studies have escalated; however, (i) a genome for an organism is predominantly static (ii) mRNAs are not always translated even when transcribed and (iii) mRNA analysis does not always correlate with protein expression. Although mRNA analysis is very powerful in gathering complete information of gene expression, proteomic analysis, although less comprehensive, can be more informative, reflecting the actual biological condition of a species, organ, or cell line. Nevertheless, modern proteomics would not have excelled to today's point without the genetic information gathered by sequencing projects. Proteomics presents some distinct challenges over its genomic counterpart in that protein cannot be amplified, proteins are present at a large range of concentrations, and that the proteome is constantly adapting to the cellular environment and inner state.

Whole protein analysis, referred to as "top-down" proteomics, is the mass analysis of an intact protein or its isoforms. Combining protein chemistry and modern mass spectrometry (MS) has paved the way for structural analysis of proteins and protein complexes as these techniques require less protein than the conventional structural techniques like NMR and X-Ray crystallography (Artigues et al., 2016). Protein characterization is often required in the biopharmaceutical industry to ensure bio-products such as monoclonal antibodies are structurally defined and reproduced identically between batch productions. More commonly proteomics refers to the

analysis of a proteome, a large set of proteins expressed at a given time, which is enabled by “bottom-up” proteomics.

Bottom-up proteomics is the common method to identify proteins through partial characterization of their amino acid sequence by a proteolytic enzyme digestion of the intact proteins prior to mass spectrometry analysis. By comparing the masses of the proteolytic peptides or their mass spectra with those predicted from a theoretically digested sequence database, peptides can be identified and multiple peptide sequences assembled into a protein identification. Bottom-up proteomics gives a better front-end separation of the peptides compared to top-down whole protein separation. The major disadvantage of bottom-up proteomics is the limited sequence coverage attained by identified peptides. Bottom-up proteomics, which is the approach used in this thesis, is performed by digestion of a crude protein extract with the resulting peptides separated by liquid chromatography prior to tandem mass spectrometry (LC-MS/MS).

Proteomics relies on three basic technological cornerstones which include (i) a method to extract and separate complex protein or peptide mixtures, (ii) mass spectrometry to acquire the data necessary to identify individual proteins, and (iii) bioinformatics to analyse and assemble the MS data.

1.1.1 Sample preparation

Protein extraction is the first step of sample preparation for proteomics. Many protein extraction methods exist including detergent-based lysis, sonication and mechanical disruption. The most appropriate method of protein extraction is determined upon careful consideration of sample source, type, physical properties, abundance, complexity and cellular location of the proteins of interest. Enrichment and/or fractionation steps can be introduced at the protein and/or peptide level if sample complexity needs to be reduced or when a specific subset of proteins or peptides are of interest.

Whole proteome analysis of cells or tissue can be undertaken to yield a crude lysate using a buffer containing detergent. Detergents are organic amphipathic (with hydrophobic tail and a hydrophilic head) surfactants. They are used to compromise the cell integrity by separating the lipid molecules of the cells membranes using the hydrophobic part of detergent. The cell is lysed and releases its contents. Centrifugation

removes cell debris and the protein-rich supernatant, which is referred to as the lysate, is collected. Lysis reagents are the most common method for sample preparation from cells or tissue in proteomics as it avoids shear forces which occur in techniques such as sonication and mechanical disruption that tend to damage cellular proteins.

1.1.1.1 Subcellular fractionation and enrichment

Despite major improvements in the sensitivity and speed of modern mass spectrometers and their high performance liquid chromatography systems, loading capacity and ion suppression can still limit the overall coverage of complex proteomic samples. Ion suppression occurs in the ionization process when molecules originating from the sample matrix or co-eluting compounds co-elute with the analyte(s) of interest and interfere with the ionization process in the mass spectrometer (Van Eeckhaut et al. 2009). The main cause of ion suppression in LC-MS systems is a change in the spray droplet solution properties caused by the presence of non-volatile or less volatile solutes such as salts, drugs and endogenous compounds (King et al., 2000). Ion suppression can affect the amount of charged ions in the gas phase that ultimately reach the detector, causing signal suppression which reduces the signal-to-noise ratio. To overcome the limitations in the dynamic range of proteomic technologies and to increase the signal-to-noise ratio, protein lysates can be separated into multiple fractions which are analysed separately by LC-MS/MS to improve proteome coverage. Unfractionated whole-cell lysate analyses are dominated by the most abundant proteins which can mask the identification of low abundant proteins which are often of most interest. Traditional methods for fractionation at the protein level include the separation of proteins based on pH, size and hydrophobicity. The advantage of these techniques is that they attain pure fractions with a defined attribute, e.g. proteins >50kDa in size. Organelle fractionation is often used to extract proteins from specific subcellular locations. The most common and efficient method for subcellular fraction is high-speed ultracentrifugation normally coupled with density gradients. Rotational speeds in an ultracentrifuge can reach up to 1,000,000 x g creating centrifugal forces which cause sedimentation of macromolecules. Since different compartments of a cell have different sizes and densities each compartment will sediment into a pellet under different centrifugal forces. Fractionation

yields sub-proteomes which have minimal cross contamination and overlap thus resulting in an improved proteomic depth and reduced sample complexity. Disadvantages of fractionation include the need for larger quantities of starting material, long processing times and expensive instrumentation (such as ultracentrifugation) required to carry out these techniques.

Subcellular enrichment is another common method of proteome separation that can be employed if a particular subset of proteins is of interest. This method is superior to fractionation in some regards due to the smaller starting material requirements, shorter processing times and cheaper costs. Commercial kits are widely available and are applied to enrich proteins from various subcellular compartments. Examples are the isolation of integral membrane proteins and nuclear proteins. The largely proprietary formulation of the extraction reagents provided in these kit approaches means that users are often uninformed of the mechanism of extraction; however, their efficiency and compatibility outweighs this issue. Integral membrane proteins and membrane-associated proteins can be enriched using optimized detergents to separate membrane (hydrophobic) proteins from cytosolic (hydrophilic) proteins. Similar reagent-based kit methods can be used to enrich for nuclear, cytoskeletal, lysosomal and mitochondrial proteins. Reducing sample complexity through either subcellular enrichment or fractionation processes can hugely benefit sample analysis; however, sample number, MS run times and costs increase with these approaches.

1.1.1.2 Proteolytic digestion

The most common technique to identify proteins is to enzymatically digest intact proteins into smaller peptide fragments and then separate the peptides by reverse phase chromatography prior to tandem mass spectrometry. Peptides are easier than proteins to fractionate by liquid chromatography and ionize and fragment more efficiently than proteins, and the resulting spectra are easier to interpret for protein identification. When protein has been extracted from the biological material or subcellular compartment of interest it may still remain in a three-dimensional configuration. To allow the proteolytic enzyme to efficiently cleave the intact proteins into peptides, proteins must be

structurally denatured with strong chaotropic agents such as urea or thiourea which are normally included in the lysis buffer. The reduction of disulfide bridges in protein is also necessary to unfold the intact protein; dithiothreitol (DTT) is the most commonly used reducing agent in proteomics. The free sulfhydryl groups on the cysteine residues are then alkylated with a reagent such as iodoacetamide which irreversibly prevents the free sulfhydryl groups from reforming disulfide bonds.

The denatured, reduced and alkylated proteins are then digested to fragment proteins into peptides. Numerous proteolytic enzymes/proteases are commercially available which differ in their cleavage sites. Trypsin is the gold standard for protein digestion and is universally used in the majority of proteomic studies. Trypsin is a serine protease which cleaves proteins at the carboxyl side of arginine and lysine. Trypsin is purified from bovine or porcine pancreas or produced as a heterologous recombinant preparation. Trypsin is considered the protease of choice prior to mass spectrometry since it has a high proteolytic activity, and is very aggressive and stable under a wide variety of conditions. Disadvantages of trypsin include reports that trypsin exhibits lower cleavage efficiency towards lysine than arginine residues which can result in 10-30% undigested, missed cleavage sites (Giansanti et al., 2016). To overcome this, supplementation with Lys-C prior to or alongside trypsin digestion can enhance trypsin performance and digestion efficiency (Glatter et al., 2012). The content of lysine and arginine amino acids underpins the proteomic coverage attainable when using trypsin. If sequence information of a protein or peptide of interest is known and little to none of these amino acids is present then alternative proteases with differing cleavage sites would be more advantageous. Examples of alternative proteases include AspN, GluC, ArgC, chymotrypsin and pepsin.

1.1.1.3 Labelled vs label-free techniques

Quantitative proteomics analysis can be performed on labelled or label-free samples. Labelling approaches can be introduced at the cell growth stage using metabolic labelling or after the protein has been extracted using chemical labelling. Examples of labelling techniques include stable isotope labelling by amino acids in cell culture (SILAC), tandem mass tags (TMT), isotope coded affinity tags (ICAT) and isobaric tags for relative and absolute quantification (iTRAQ). These techniques are all based on standard procedures for quantification by mass spectrometry in which an analyte is quantified by comparison with an introduced isotopomer of the analyte that acts as an internal standard (Thompson et al., 2003). Since the quantity of the standard is known, this means that the quantity of the analyte may be determined from the ratios of their peak intensities. Although labelling-strategies can reduce MS run times, the added cost and limited number of sample labels means study design is quite fixed. Typically you only directly compare those samples that were physically mixed and measured in one run.

In contrast, label-free quantitation approaches are very flexible; even after a study has started new samples can be added to the end of the MS runs for inclusion. Quantitation is most commonly based on the peptide peak signal intensity in the first stage of mass spectrometry (MS1) or less often by spectral counting in the second stage of mass spectrometry (MS2) during tandem mass spectrometry analysis (Wong and Cagney, 2010). In MS/MS proteomic analysis MS1 precursor ions refers to the peptide level and MS2 fragment ions refer to amino acid sequence information. In label-free analysis, samples are analysed under identical run conditions to allow a direct comparison of samples. This makes LC-MS reproducibility critical for accurate quantitation and peak alignment. Label-free approaches are more sensitive to technical variance and require more MS run time than label-based strategies however they are fast, easy to perform, inexpensive, and an unlimited number of samples can be compared.

1.1.2 Mass spectrometry

The single most important analytical technique in proteomics is mass spectrometry. The first fully functional mass spectrometer was reported in 1919 by Francis Aston (D.Sc, 1919). Almost 100 years since its invention, advanced developments have boosted its application and proteomics into the forefront of scientific research. The sensitivity provided using MS allows proteins and peptides at very low concentrations to be identified. The mass accuracy of high-resolution mass spectrometers, which can be less than 5 parts per million (ppm), increases confidence in the identification provided by the bioinformatics search of the raw data. Proteomics based mass spectrometry is used for the detection, fragmentation, quantitation and mass analysis of intact proteins, peptides and their corresponding amino acid sequences. High resolution mass analysis is used to determine amino acid sequence information of a peptide; sequence information can then be compared against a protein database of choice to determine the protein derived from the original biological sample. Increasing peptide identifications leads to increased protein coverage which results in a higher confidence surrounding the protein identification present and less likelihood of false-positive identifications.

The mass spectrometer is defined by one particular trait, i.e. the ability to separate analytes based on their mass-to-charge ratio (*m/z*). To achieve this target molecules are ionized, and subjected to highly controlled electric and/or magnetic fields which deflect the paths of individual ions (Rubakhin and Sweedler, 2010). The paths of two identically moving particles of the same charge will change differently under the influence of identical electric and/or magnetic fields if the particles have different masses. The movement of the heavier particle will be affected less than the lighter one.

Mass spectrometers are composed of three functionally distinct main components: an ion source, a mass analyser that measures the mass-to-charge ratio of ionised analytes, and a detector that registers the number of ions at each *m/z* value (figure 1.1). More than one ion source, mass analyser, and detector are often integrated into a mass spectrometer, thereby providing greater flexibility in the instruments' operating modes and enabling different types of analytes to be characterized.

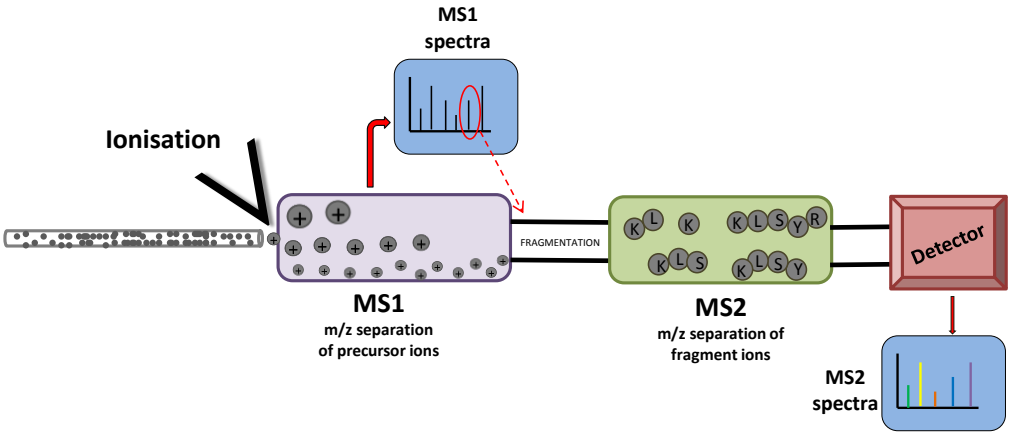


Figure 1.1 Schematic of a basic tandem mass spectrometer and its components. A sample must be ionized to create gas phase, charged analytes. The analytes enter the mass analyser under vacuum where they are separated based on the m/z ratio. Precursor peptide ions (MS1 ions) are then selected and fragmented into product fragment ions (MS2 ions) and analysed in a second mass analyser. The fragment ions are then detected to generate a mass spectrum of signal versus m/z.

1.1.2.1 Ion source

Mass spectrometric measurements are carried out in the gas phase on ionized analytes. To analyse a peptide sample by mass spectrometry it must be vaporised from the liquid phase into gas phase, ionised and transferred into the mass analyser. The two most commonly used ionisation techniques for protein and peptide mass spectrometry analysis are matrix-assisted laser desorption ionisation (MALDI) (Karas and Hillenkamp 1988) and electrospray ionisation (ESI) (Fenn et al., 1989). These "soft" ionisation techniques are predominantly used for proteomic analyses as they impart little energy onto the target analyte thus preventing fragmentation of these molecules when ionised. In principle, MALDI mixes the sample uniformly with a large quantity of matrix which consists of crystallised molecules in water and an organic solvent. An example of a matrix commonly used for peptide analysis is α-Cyano-4-hydroxycinnamic acid (CHCA). The matrix is often acidic to act as a proton source to encourage ionisation of the analyte and have a strong optical absorption to efficiently and rapidly absorb laser light energy (Zenobi and Knochenmuss, 1998). When the matrix dries and the laser is applied, it heats rapidly and vaporises together with the sample, into gaseous phase ions. ESI ionises analytes out of solution using an electrospray which applies high

voltage electricity to a peptide solution creating charged droplets which form an aerosol. ESI is readily coupled to liquid-based separation techniques such as chromatography. MALDI-MS is normally used to analyse relatively simple peptide mixtures, whereas integrated liquid-chromatography ESI-MS systems (LC-MS) are preferred for the analysis of complex samples (Aebersold and Mann, 2003). ESI differs from MALDI in that it produces multiply charged peptide ions; the *m/z* values of the resulting ions become lower and fall in the mass ranges of all common mass analyzers (Banerjee and Mazumdar, 2012). This results in extending the mass range to accommodate the kDa orders of magnitude observed in peptides. ESI carried out at the interface between the nano liquid chromatography (HPLC) and the mass spectrometer is termed nanoelectrospray ionisation (nano-ESI). Nano-ESI only requires a few µL of a sample which is forced through very small needle orifices with µm diameter and is sprayed from the needle tip by applying a voltage into the mass analyser. Karas et al. have shown that, in comparison with conventional electrospray, nano-ESI reduces interference effects from salts and other species and provides better sensitivity toward a variety of analytes, including peptides attributed to the reduced droplet size (Juraschek et al., 1999; Karas et al., 2000). Peptides eluting from the HPLC are continuously ionised with high efficiency before entering the mass analyser. In this thesis nano-electrospray ionization followed by mass spectrometry using two types of mass analysers has been performed.

1.1.2.2 Mass analysers

The future goal of mass spectrometry based proteomics is the complete characterisation of all proteins. Currently, bottom-up analysis of proteomes can result in thousands of peptide fragments with a large dynamic range of abundances. Detection and measurement of these peptides has advanced significantly by using high-resolution, sensitive and fast MS instruments. Accurate mass analysis and maximum sequence coverage in fast acquisition times is becoming an attainable goal to achieve in-depth proteome coverage. The mass analyser is central to the technological power of mass spectrometry. There are four basic types of mass analyser currently used in proteomics

research. These are the quadrupole, ion trap (IT), time-of-flight (TOF), and Fourier transform (FT-MS) analysers. These mass analysers differ in their mode of operation and performance standards, each with their own strengths and weaknesses. In quadrupole mass analysers, four parallel rods are subjected to alternating voltages allowing ions with specific *m/z* to pass through to the detector. Ion trap analysers were patented in the 1950s as an alternative to quadrupoles (Paul and Steinwedel 1953). The principle of mass filtering is still based on alternating voltages, but more recently using a three electrode setup in a new configuration. One of the electrodes, the "ring", is used to capture or "trap" ions in an isolated environment for a time frame prior to mass analysis (Parker et al., 2010). In both quadrupole mass analysers and ion traps, oscillating fields are used to create stable trajectories for ions at certain *m/z* values. Time-of-flight analysers, as the name suggests calculates the *m/z* from the length of time it takes for the ion to reach the detector under a field-free vacuum. The mass of the analyte is proportional to the flight time (measured in nanoseconds) with heavier ions travelling more slowly hence taking more time to reach the detector. The Fourier transform MS analyser represented a breakthrough in terms of higher resolving power and mass accuracy. In Fourier transform MS, charged ions rotate in a circular path in a cyclotron when under a fixed magnetic field. The oscillation of a certain mass is represented as a frequency detected using an image current detection system. The time domain signal (containing all the characteristic frequencies of the measured ions at intensities corresponding to the amount of the molecular species in the sample) is converted using the Fourier transform mathematical formula to generate mass spectra with high mass accuracy. The Orbitrap was invented by Makarov in 1999 (Makarov, 2000) and is the first analyser introduced to the market that is based on a new mass filtering method. In the Orbitrap, ions are trapped and spiral/orbit around a central axial spindle-like electrode putting the ions in different orbitals depending on their characteristic *m/z* value. These oscillations induce an image current that is detected at a specific frequency which can then be transformed into mass spectra using Fourier transformation. The Orbitrap was first reported as a tool for proteomics research in 2005 by Hu et al. (Hu et al., 2005).

Mass spectrometers built with a single mass analyser (single-stage) simply measure the mass of the ion generated by the ion source. Multi-stage mass spectrometers are hybrids; these can consist of the same mass analyser in tandem e.g. a triple quadrupole (Yost and Enke, 1978) or a combination of different mass analysers e.g. the LTQ Orbitrap (Makarov et al., 2006). The power of hybrid instruments allows ions to be strongly manipulated to increase sensitivity, improve mass resolution and fragment selected ions for sequence information.

Currently the most popular mass spectrometers are dominated by high mass resolution Orbitrap-based instruments. The Thermo Scientific LTQ-Orbitrap XL mass spectrometer combines ion trap and Orbitrap mass analysers (Makarov et al., 2006). The Thermo Scientific Q Exactive mass spectrometer was later introduced which used a quadrupole for precursor ion filtering with an Orbitrap mass analyser for MS1 and MS2 analysis (Michalski et al., 2011). The more recent Thermo Scientific model, the Orbitrap Fusion Tribrid mass spectrometer combines the best of quadrupole, ion trap and Orbitrap mass analysis in a revolutionary parallelised Tribrid architecture (Senko et al., 2013). The domination of Thermo Scientific's Orbitrap-based instruments has driven the development of novel mass analysis approaches in other mass analysers, most importantly data-independent acquisition (DIA). Traditional data-dependent acquisition (DDA) takes a pre-determined selection of peptide ions (typically abundant) forward for fragmentation and MS2 analysis. Modern mass spectrometers have increased the number of peptide ions that can be brought forward for fragmentation at any one time through increasing scan speeds; however, limitations still exist. In contrast to DDA, DIA fragments every single peptide ion and analyses it in MS2. This can be performed by sequentially isolating and fragmenting ranges of m/z values, referred to as sequential window acquisition of all theoretical fragment ion spectra (SWATH) (Gillet et al., 2012). Analysis of SWATH-MS data most often relies on a targeted data analysis strategy in which target peptides are detected and quantified from the SWATH-MS fragmentation data by extracting and correlating previously generated query parameters for each target (Collins et al., 2017). SWATH is a highly reproducible method of DIA however, the complexity of the resultant spectra and specific software tools required for data analysis require massive computational power and bioinformatics skills.

During my PhD project I had access to both an LTQ Orbitrap XL and Orbitrap Fusion Tribrid mass spectrometer. Chapters 3 and 5 employ the LTQ Orbitrap to profile pancreatic cancer samples and microRNA-378 depleted CHO DP12 cells respectively. Chapter 7 employs the Orbitrap Fusion Tribrid mass spectrometer to analyse microRNA-7 depleted CHO DP12 cells.

1.1.2.3 Fragmentation strategies

Tandem mass spectrometry (MS/MS or MS^n) is a crucial technique for protein and peptide sequencing as well as post-translational modification (PTM) analysis. This involves the separation of ions based on *m/z* and selection of precursor ions in the first mass analyser (MS). Selected precursor ions are then fragmented to produce fragment ions or "daughter ions". The fragment ions of the precursor ion are then separated, mass analysed and detected in the second mass analyser (MS/MS). Fragmentation is thus central to the characterisation of peptide sequences by tandem mass spectrometry. In MS/MS proteomic analysis MS1 precursor ions refers to the peptide level and MS2 fragment ions refer to amino acid sequence information. In this thesis CID fragmentation was used as the fragmentation strategy for all studies.

1.1.2.3.1 Collision-induced dissociation (CID)

CID involves the collision of an ion with a neutral gas and subsequent dissociation of the ion. It is a low-energy fragmentation process which typically results in cleavage of the amide C–N bond in the peptide backbone resulting in a series referred to as b- and y-fragment ions by the Roepstorff nomenclature (Roepstorff and Fohlman, 1984). The fragment ions produced which contain the N-terminus are called b ions; those containing the C-terminus are called y-ions.

1.1.2.3.2 Higher energy C-trap dissociation (HCD)

HCD fragmentation is available on the LTQ-Orbitrap mass spectrometer. It is a variation of CID where ions are fragmented in a collision cell rather than in the linear ion trap. HCD fragmentation with Orbitrap detection has no low-mass cutoff, high resolution ion detection, and increased ion fragments resulting in higher quality MS/MS

spectra (Jedrychowski et al., 2011). HCD as the name suggests also employs higher energy dissociations than those used in ion trap CID. The main disadvantage of HCD is longer spectral acquisition times due to detection in the Orbitrap which requires more ions than the LTQ.

1.1.2.3.3 Electron transfer dissociation (ETD)

ETD induces fragmentation of peptides using ion/ion chemistry by transferring an electron to higher charge state cationic peptides (Syka et al., 2004). ETD fragments the peptide backbone by cleaving the amide bond producing complementary c- and z- type ions. ETD preserves post-translational modifications of peptides such as phosphorylation, sulfonation, glycosylation, nitrosylation, methylation and acetylation that are labile by collision induced dissociations such as CID and HCD (Mikesh et al., 2006). For this reason ETD is preferentially used when analysing PTMs. It has been shown to be a highly complementary approach to CID for peptide analysis as it out performs CID in analyses of peptides with charge states greater than 2 (Good et al., 2007).

1.1.3 Bioinformatics

No area of proteomics has evolved faster than bioinformatics. Proteomics is dependent on bioinformatics to computationally analyse the raw mass spectral data and convert it into peptide and protein identifications. Bioinformatics is not only critical for qualitatively identifying proteins but also in quantitating differences in protein or peptide abundance levels. Bioinformatics is an increasing aspect of proteomic workflows that is becoming more and more important for data interpretation as the wealth of data increases with advancing mass spectrometers. Matching large amounts of mass data against predicted masses from protein sequence databases, statistical analysis of samples and protein inference where multiple, but incomplete peptides from each protein are detected is too challenging by hand. *In silico* pipelines are tailored to the proteomic experiment and can also incorporate annotation and display the distribution of the resultant data in different charts. Bioinformatics software tools vary in their capabilities and range from command-line tools to costly standalone high-throughput

programs. Many of the vendors of mass spectrometers develop and market computational programs specific to their MS, this enables seamless processing of raw data into comprehensive result outputs.

1.1.3.1 Protein Identification

MS spectra provide information on the mass of a precursor peptide and the mass of the amino acid sequence. This raw mass data must be searched against a protein sequence database using stringent search algorithms to decipher the peptide and thus protein identification that matches the experimental data with a degree of certainty. A typical workflow for protein identification from mass spectrometry data is illustrated in figure 1.2, taken from Cottrell (2011). There are many variations of this workflow; however, the output is always spectral data which must be deciphered *in silico* to identify the peptide and/or protein present in the original sample. Bioinformatics software requires search parameters to be applied when processing the data which describe the experimental design before progressing to database searching and scoring peptide matches. Such parameters include the allowed mass tolerance for precursors and fragments, proteolytic enzyme used, maximum missed cleavage sites allowed and any possible modifications present in the sample.

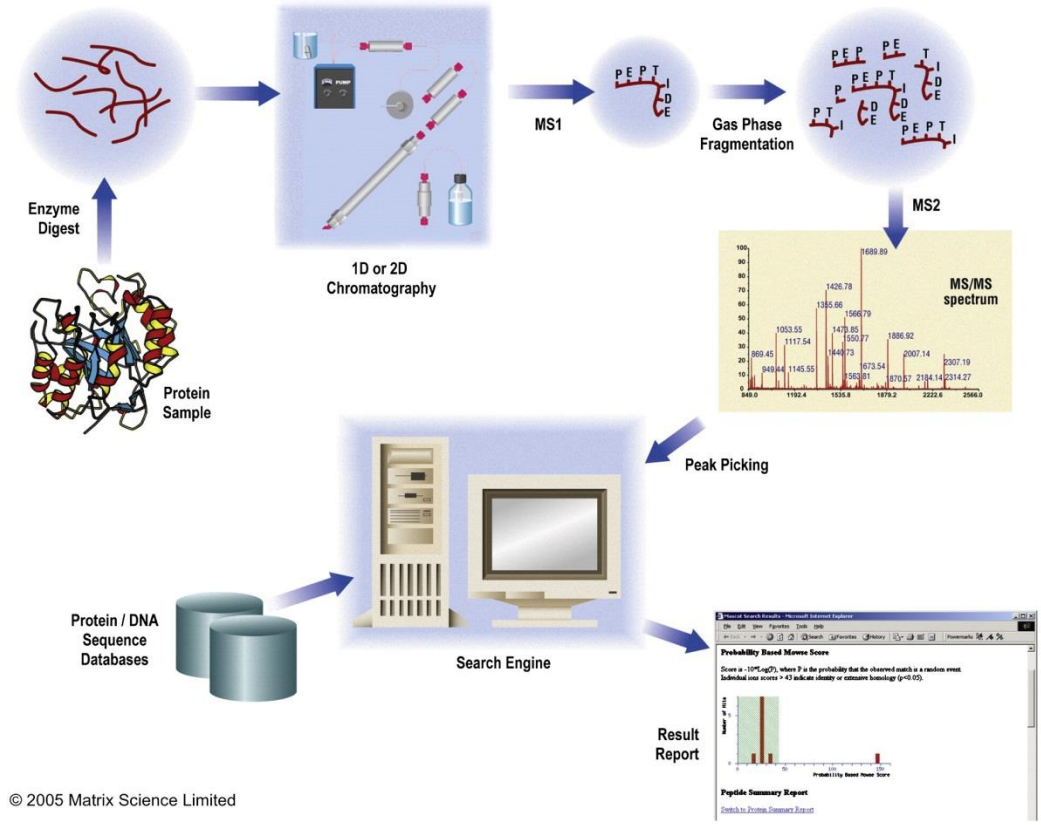


Figure 1.2 A typical workflow for protein identification using MS/MS raw data (Cottrell, 2011).

1.1.3.2 Sequence databases

There are two major repositories for sequence databases, one hosted by the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/) and the other hosted by the European Bioinformatics Institute (http://www.ebi.ac.uk/). Thanks to the Human Genome Project and other sequencing ventures, high quality sequence databases can now be generated at an exceptional rate. The repositories contain databases from a variety of sources such as for example, Swiss-Prot, UniProt, TrEMBL, RefSeq and GenBank. UniProt is the most widely used sequence database (Xu, 2004). The database contains over 60 million sequences, of which over half a million sequences have been curated by experts who critically review experimental and predicted data for each protein (The UniProt Consortium, 2011). Human *(Homo sapiens)* protein identifications are easily deciphered from MS spectral data as the protein sequence database is expertly curated with a minimal level of redundancy.

Protein identification and analysis of Chinese hamster ovary (CHO) cell lines is a major component of this thesis. CHO cells were first isolated 60 years ago in 1958 from a biopsy of an ovary of an adult Chinese hamster (Puck et al., 1958). There are many commercially available sub-clones of the CHO cell line in addition to derivatives of these which are constructed to express different features such as bio-therapeutic products (Gamper et al., 2005). Past proteomic analyses of CHO cell lines were based on homologous peptides whereby the raw MS data was searched against a human, mouse and/or rat sequence database with conserved protein identifications inferring presence in the original CHO sample (Meleady et al. 2012; Kim et al. 2011; Wei et al. 2011). Thankfully, increasing affordability of sequencing techniques resulted in the draft genome sequence of the CHO-K1 cell line (Xu et al., 2011), sequencing information for the CHO-SEAP cell line (Hammond et al., 2011) and cDNA sequencing of CHO K1 and CHO DUKXB11 cells (Becker et al., 2011) all being released in 2011. Two protein sequence databases based on the Becker et al. and Xu et al. sequencing information were generated and made publicly available. A study by Meleady et al. incorporating these databases demonstrated a 40-50% increase in the number of proteins identified in CHO K1 MS data when compared to homology-based searches (Meleady et al., 2012b). More recently the Chinese hamster species *(Cricetulus griseus)* draft genome was sequenced (Lewis et al., 2013), ten years after the first human genome sequencing. The same 2013 study also re-sequenced and analyzed the genomes of six CHO cell lines from the CHO-K1, DG44 and CHO-S lineages. The original CHO K1 genome sequencing (Xu et al., 2011) and the Chinese hamster genome (Lewis et al., 2013) were annotated, further curated and accompanied by a transcriptome reference released by NCBI RefSeq in 2012 and 2014 respectively. The overdue protein sequence databases consequently harbour some redundant, un-reviewed sequences and lack the more thorough curation available in the human protein sequence databases. In a recent comparison of the NCBI RefSeq 2012 and 2014 annotated genomes the latter annotation was preferred because across all 60 CHO cell culture samples sequenced, higher mapped rates were observed to the 2014 annotated genome (Le et al., 2015). Therefore, the annotated NCBI *Cricetulus griseus* protein database is recommended to provide the most accurate protein identification when matching mass spectra and peptide information in proteomic

studies of CHO cells (Heffner et al., 2017). Sequencing studies of the CHO genome using modernised technologies are ongoing with updated assemblies generating a more contiguous and complete genome reference which is contiuously updated on the NCBI database (Rupp et al., 2018).

1.1.3.3 Identification algorithms

When using protein databases for peptide sequence identification, a search algorithm engine is used for matching the MS/MS data to a sequence database. The two most common search engines are SEQUEST (Eng et al., 1994) and MASCOT (Perkins et al., 1999). These search engines require parameters to be set to allow accurate protein identification such as the name of the protein database and taxonomy of the sample, precursor and fragment mass tolerance, proteolytic enzyme used and a missed cleavage sites allowance. The search engine can then *in silico* access the protein database, report all possible peptides based on the digestion enzyme, calculate the theoretical mass of each peptide and correlate the observed mass of a MS/MS spectra to the theoretical mass. The algorithm computes a similarity score to a given pairing of an observed MS/MS spectrum and to that of a theoretical spectrum of a candidate peptide. The candidate observed sequences must pass a minimum algorithm dependent scoring threshold in order to be considered a positive identification. Search algorithm engines are often available on web-servers thus making them easily integrated with bioinformatics processing pipelines. Once individual peptide sequences are confidently identified, the set of peptide sequences is used to infer which proteins may have been present in the original sample.

1.1.4 Quantitative proteomic profiling

Quantitative proteomics is a powerful approach to determine protein expression across samples through abundance evaluation. Quantitative proteomic profiling can measure changes in for example: healthy vs diseased individuals or changes in proteins over time in response to a drug. Traditional methods for studying proteins expression have been biased toward a relatively small subset of proteins for which high-quality, mainly

antibody-based assays have been commercially available (Edwards et al., 2011). Mass spectrometry based methods for quantitative proteomic analysis have now emerged which allow for identification and quantitation of hundreds and thousands of proteins expressed in a given biological sample. These advancements have significantly contributed to unravelling cellular signaling networks, understanding disease states and improving diagnosis.

Mass spectrometry based quantitative analysis can be based on absolute or relative peptide abundances. Absolute quantitation or targeted proteomics is used to determine the exact concentration of a protein in a sample. This method of quantitation uses a synthetic peptide at a known concentration for MS analysis that has been previously identified in the biological sample of interest. This peptide can be serially diluted and used to create a standard curve of intensity values or labelled to shift its mass higher than the endogenous peptide and spiked into the sample. The biological sample containing the peptide of interest can then be MS analysed and the concentration can be quantitated by plotting its intensity value on the standard curve or comparing the labelled synthetic peptide intensity to the endogenous peptide abundance. Absolute quantitation is often used for analysing a blood biomarker in a patient sample to determine the ng/mL of the biomarker which may aid diagnosis. Traditional targeted proteomics uses a triple quadrupole mass spectrometer for sample analysis. Relative quantitation is more commonly used for whole proteome or sub-proteome analyses as it can simultaneously provide information on changes in protein expression of hundreds of proteins at a given time. Differences in peak intensity of the same analyte between multiple samples can accurately reflect relative differences in its abundance. Peak intensity is the signal intensity value of an ion produced from ESI and identified in the MS which correlates linearly with ion concentration (Voyksner and Lee, 1999; Wang et al., 2003). Moreover, when the peak area of all identified peptides for a protein are combined the peak area correlates linearly to the concentration of the protein (r^2 = 0.991) (Chelius and Bondarenko, 2002). Since this discovery, many tools for extracting the peak ion signal intensities and quantitating peptide and protein abundances have been established and implemented into bioinformatics software (Al Shweiki et al., 2017).

1.1.4.1 Label-free quantitative proteomics

Label-free quantitative proteomics measures the differences in proteins abundance across two or more biological samples. In recent years label-free approaches have become more popular thanks to the simple experimental designs and sample preparation methods (Bantscheff et al., 2007). Each biological sample within an experiment must be individually prepared and analysed by LC-MS/MS. Quantitative label-free proteomics directly compares the peak intensities of a given peptide ion across all samples in the analysis and provides a fold-change result which indicates the protein expression profile across the experimental groups. Although theoretically this method is simple, in reality there are some practical restraints that must be acknowledged. First, even the same sample can give differences in the peak intensities of a peptide ion from run to run. These differences are often caused by technical variation during sample preparation so an efficient, reproducible method should be employed to best suit the biological sample. Normalisation is often implemented to account for this kind of variation. Second, any drifts in retention time and *m/z* will significantly confound the direct comparison of multiple LC-MS/MS sample runs. Samples are aligned to each other for peak comparison; a failure to accurately align runs will result in large variability, inaccuracy in quantitation and false-positive identifications. Thus, highly reproducible LC-MS/MS and careful peak alignment are crucial for the success of this quantitative method. Lastly, the large amount of data collected during LC-MS/MS analysis of complex protein samples requires label-free quantitative data analysis to be computationally performed. This has led to the development of bioinformatics software programs, both freeware and commercially available, for the automatic quantitation of protein changes between LC-MS/MS samples with comprehensive large-scale capabilities.

The two most popular software packages for label free proteomic quantitation are Progenesis QI for Proteomics and Maxquant LFQ. Maxquant was developed in the Max Planck Institute of Biochemistry, Germany and released in 2008 (Cox and Mann, 2008). It is freely available from http://www.maxquant.org/. Since its release it has grown substantially and now boasts computational workflows for both quantitative labelled and label-free proteomic analyses as well as an integrated protein identification search algorithm for an all-in-one bioinformatics program. The MaxLFQ workflow for label-

free analysis quantifies proteins across samples using the maximum peptide ratio information from extracted peak ion signal intensities (Cox et al., 2014).

Progenesis QI for Proteomics which is marketed by the Waters company (www.nonlinear.com) also quantifies each protein based on its peptide ion signal peak intensity. In this software individual LC-MS/MS runs are represented as a map with *m/z* values plotted along the x-axis and the retention time of the peptide ions against the y-axis. The peptide ions for each run can thus be graphically separated and, the signal intensity is visualised as black spots whereby increasing spot size represents a peptide ion with a greater intensity / abundance. The samples are then aligned and normalised automatically. The experimental design must then be selected by the user to allow significantly differentially expressed peptide ions (one-way ANOVA p-value <0.05) between the experimental groups to be picked, quantitated and exported for protein identification in a separate bioinformatics program of choice. Significantly differentially expressed peptide ions now attributed to a protein identification are imported into Progenesis QI for Proteomics, abundances of multiple peptides derived from a single protein are cumulated allowing protein level abundance differences to be calculated as a fold-change across experimental groups.

Label-free quantitative proteomic analyses formed the basis for most of the studies presented in this thesis. These analyses were performed using Progenesis QI for Proteomics to determine differentially expressed proteins.

1.1.4.2 Functional annotation

Proteomic analyses, either qualitative or quantitative, result in a report outlining protein descriptions, accession numbers, sequence information and number of peptides identified. These results do not provide any biological context of the phenotype of the sample. In order to understand the biological role of the identified proteins additional bioinformatics analysis must be performed. Most commonly, gene ontology (GO) analysis is performed which annotates proteins in the context of three categorical attributes; biological process, molecular function and cellular component. This can be helpful in the context of chapter 3 whereby cellular component analysis verified that the membrane protein enrichment kit extracted membrane-associated proteins. GO analysis

can also be integrated into platforms such as DAVID (Huang, Sherman, and Lempicki 2009), STRING (Snel et al., 2000), and PANTHER (Thomas et al., 2003). These tools integrate GO analysis to create and visualise protein lists in a biological context. Pathway analysis is another important feature of functional annotation. Mechanistic reasons for phenotypes can be attributed to the identification of an up or down regulated pathway from a protein list. Pathway analysis tools include Kyoto encyclopedia of genes and genomes (KEGG) and ingenuity pathway analysis (IPA) (Kanehisa and Goto, 2000; Krämer et al., 2014). KEGG analysis in chapter 7 revealed a role for the Akt growth pathway, splicesome and RNA transport pathways as significantly upregulated following microRNA-7 depletion in CHO cells.

1.2 Pancreatic cancer

The pancreas is a glandular organ located in the upper left part of the abdomen surrounded by the small intestine, liver, and spleen. The organ serves as two glands in one: a digestive exocrine gland and a hormone producing endocrine gland. Almost the entire pancreas (95%) consists of exocrine tissue that produces and secretes pancreatic juice into the duodenum section of the small intestine. The juice contains bicarbonate which neutralizes acid entering the duodenum from the stomach; and digestive enzymes such as trypsin, amylase and lipase which help to break down proteins, carbohydrates and lipids, respectively. The remaining pancreas tissue consists of the endocrine cells contained in complex clusters of cells called islets of Langerhans after their discoverer Paul Langerhans in 1869. Different specialised cells within these clusters of cells produce hormones such as insulin (beta cells) and glucagon (alpha cells) that synchronously regulate blood sugar levels; other islet hormones include somatostatin and pancreatic polypeptide.

The most common form of pancreatic cancer accounting for 95% of cases develops from epithelial cells of the exocrine pancreas and is classified as pancreatic adenocarcinoma (PDAC) (Becker et al. 2014). A far less common form called endocrine tumours, account for less than 5% of all pancreatic cancer tumours. The latest statistics from the National Cancer Registry of Ireland published in May 2018 reports 564 new cases and 495 deaths per year for PDAC in Ireland (https://www.ncri.ie/). The age profile at diagnosis shows 30% of patients are aged between 65-74 years old and 42% of patients are 75+ years thus meaning over 70% of patients diagnosed with PDAC in Ireland are 65+ years of age. According to GLOBOCAN 2018 produced by the International Agency for Research on Cancer, pancreatic cancer is currently the seventh leading cause of cancer deaths in both males and females (Bray et al., 2018). By 2025, pancreatic cancer is predicted to become the third leading cause of death from cancer in the European Union after lung and colorectal cancers (Ferlay et al., 2016). PDAC is considered a low-incidence but high-mortality disease. The high mortality rates are founded on the fact that most patients are diagnosed at advanced and/or metastatic

stages of the disease, and even when treatment is administered PDAC shows high levels of resistance causing minimal improvements to patient's outcome.

1.2.1 Diagnosis

PDAC develops from well-defined precursor lesions such as pancreatic intraepithelial neoplasia (PanIN), intraductal papillary mucinous neoplasm (IPMN), mucinous cystic neoplasm, and other lesions. These lesions are classified using a two-tiered system of low grade and high grade with the latter lesions having a higher propensity to progress into PDAC requiring careful clinical attention and treatment, while lower grade lesions lack a significant risk for progression to invasive carcinoma, and as such, do not warrant therapy (Basturk et al., 2015). Identification of high-grade precursor lesions represents a major opportunity to diagnose and eliminate pre-invasive precursors of PDAC to reduce incidence and in turn mortality. Although efforts have been made for the development and implementation of methods to detect precursor lesions without an invasive pancreatic biopsy (Canto et al., 2006; Eser et al., 2011; Poruk et al., 2013), no consensus opinion has been reached. Clinical diagnosis of PDAC is often achieved using an endoscopic ultrasound, CT scan, MRI or other imaging technologies, however these are not only expensive and scarcely available, they also have limited sensitivity thus far in detecting very small precursor lesions (Brand et al., 2007). Therefore, currently by themselves, precursor lesions are not a useful screening strategy for the early detection of PDAC.

Efforts to identify screening regimens for early diagnosis of PDAC have focused mainly on serum biomarkers due to their advantages like minimal sampling variability, cost and the minimally invasive nature of the method. The "best" and only biomarker in widespread clinical use is CA 19-9, a carbohydrate tumour-associated antigen which is often released into the blood of patients with pancreatic cancer. The use of CA19-9 is limited to monitoring response in patients undergoing treatment due to its low sensitivity (79%) and specificity (82%) (Goonetilleke and Siriwardena 2007). Moreover CA19-9 levels are often elevated in patients with other gastrointestinal conditions such as chronic and acute pancreatitis, cirrhosis, cholangitis, and obstructive jaundice (Sheen-

Chen et al. 2007; Wu, Kuntz, and Wadleigh 2013). Based on these reasons, CA19-9 is not acceptable as a screening biomarker for the diagnosis of PDAC in the general population.

Several promising candidate biomarkers have been identified recently that successfully discriminate PDAC from healthy controls. A complete review of these is covered in chapter 3 (Coleman et al. 2016). Although extensive efforts have been made to date, a reliable biomarker with high sensitivity and specificity for PDAC which reaches the necessary accuracy for diagnostic use has not been identified. More recently combinations of clinically promising biomarkers such as MMP7, osteopontin and CEACAM1 together with CA19-9 in a biomarker panel screening method increased power to accurately diagnose PDAC compared to a single biomarker alone (Kuhlmann et al. 2007; Koopmann et al. 2004; Simeone et al. 2007). Such studies give promise that a biomarker panel may be identified which increases sensitivity and specificity. The improvement of PDAC diagnosis and the introduction of reliable early detection screening approaches remains a major goal in current research. Early detection is the single most significant and favourable approach for improving PDAC patient outcome which must be introduced in the near future.

1.2.2 Treatment

1.2.2.1 Current treatment options

Upon diagnosis the only realistic treatment option for PDAC patients is complete resection followed by adjuvant treatment (Adamska, Domenichini, and Falasca 2017). Unfortunately however, PDAC typically develops with very few symptoms subsequently resulting in only 10-20% of patients presenting at a stage amenable for surgical resection and possible cure (Poruk et al. 2013). For those patients with localised disease and a tumour size < 2 cm complete surgical resection can offer a 5-year survival rate of 18% to 25% (Yamamoto 2015), an improvement compared to the median 5-year survival rate of 3% .

The role of radiation therapy in PDAC treatment is strongly debated with conflicting evidence for its benefits over the years. Although the ability of radiation to improve

local control has been demonstrated, it has not always led to improved survival outcomes for patients generally as a result of using sub-optimal radiation techniques (Boyle et al. 2015). Recent and ongoing trials of PDAC radiation treatment incorporate more optimized methods such as stereotactic body radiation therapy that allows high dose delivery to tumours with minimal dosing to surrounding normal tissue. It is hoped that these advanced techniques will lead to improved outcomes while reducing toxicity rates to become a standard of treatment in PDAC.

With the majority of PDAC patients presenting at a non-resectable and/or metastatic stage, chemotherapy is the prevailing treatment option. Gemcitabine is a standard chemotherapy treatment with a 5-10% response rate and which marginally increases median overall survival (OS) to 6 months compared to the untreated median survival of about 3 months (Burris et al. 1997). Between the introduction of gemcitabine in 1997 and 2012, 34 randomised phase III trials were performed to try improve the efficacy of gemcitabine using combination therapies (Ciliberto et al. 2013). The most successful to date, the MPACT trial (n=861), identified albumin-bound nab-paclitaxel (Abraxane®) together with gemcitabine increased OS to 8.5 months compared to 6.7 months in single-agent gemcitabine treated patients (Von Hoff et al. 2013; Goldstein et al. 2015). Unfortunately, the positive response to this therapy was accompanied by 77% of participants experiencing grade 3 adverse effects such as neutropenia, leukopenia, anaemia, fatigue and diarrhoea (Goldstein et al. 2015). Nevertheless, the increase in patients' survival rates was the basis for FDA approval and establishment of Abraxane®-gemcitabine as a first-line therapy option for patients with advanced and metastatic pancreatic cancer. A major step forward for the chemotherapeutic treatment of PDAC was the introduction of the combination of chemotherapy drugs including fluorouracil, leucovorin, irinotecan and oxaliplatin called FOLFIRINOX. FOLFIRINOX improved OS to 11.1 months compared with 6.8 months for single-agent gemcitabine in the PRODIGE trial (n=342) of metastatic PDAC patients (Conroy et al., 2011). Considerable toxicity was reported with side effects such as neutropenia, diarrhoea and alopecia reported for the majority of patients undergoing the treatment. However, a significant reduction in quality of life impairment was observed in FOLFIRINOX-treated patients compared to gemcitabine (Gourgou-Bourgade et al., 2013).

FOLFIRINOX is now considered as a first-line option for patients with advanced and metastatic pancreatic cancer; however its use is constrained to patients under the age of 75. To improve PDAC patients' tolerance to the drug modifications of FOLFIRINOX regimens aimed at reducing toxicity have been introduced.

Emerging studies have suggested that the complex nature between PDAC stroma and tumour cells may be hampering chemotherapeutic effectiveness as 90% of PDAC is stromal cells which act as a barrier to drug delivery (Bhaw-Luximon and Jhurry, 2015). In PDAC, Hedgehog pathway (Hh) activation influences tumour growth by promoting desmoplasia, the growth of stromal components (Bailey et al., 2008). Olive et al. reported that inhibition of Hh pathway in a pancreatic cancer mouse model enhances the formation of blood vessels in poorly vascularized tumours and decreases their stromal component (Olive et al., 2009). By combining traditional gemcitabine chemotherapy with stromal depletion using a Hh cellular signaling pathway inhibitor they observed a transient increase in intratumoral vascular density and intratumoural concentration of gemcitabine. The increased drug delivery through stromal reduction increased the survival rate from 11 to 25 days for the mouse model. Numerous emerging studies are now focusing on targeting the stromal density of PDAC tumours to increase drug delivery and improve OS, however, this research is still in its infancy (Bhaw-Luximon and Jhurry, 2015; Kota et al., 2017; Mei et al., 2016).

In summary, the chemotherapeutic drugs approved by the Food and Drug Administration (FDA) for PDAC treatment are gemcitabine (Gemzar®), albumin bound nab-paclitaxel (Abraxane®) and the FOLFIRINOX regimen which constitutes leucovorin, fluorouracil, irinotecan and oxaliplatin. Even with these available options median OS at best doubles to approximately 11 months. It remains debated, by clinicians and PDAC patients alike, whether the off-set toxicity effects of the chemotherapeutic options are worth the minimal extension in OS.

1.2.2.2 Prospective targeted therapies

Conventional chemotherapeutic treatments have been somewhat unsuccessful in significantly improving patient's chances for survival, and general only offering marginal increases to the value of months. Therefore there is a paramount need for the development of novel, more effective and less toxic strategies which advance on the currently available therapeutic options. In the last 20 years there has been a paradigm shift in the treatment of solid tumours from traditional chemotherapies to targeted therapeutics which takes advantage of the unique hallmarks of cancer (Hanahan and Weinberg, 2000).

Frequently targeted therapeutics using small molecule drugs are directed at oncogenic drivers; from the first FDA approved targeted therapy imatinib, a tyrosine kinase inhibitor for chronic myeloid leukaemia and gastrointestinal stromal tumour in 2004 (O'Brien et al., 2003); gefitinib and erlotinib targeting EGFR kinases in non-small cell lung cancer (Zhang et al., 2018); and vemurafenib for BRAF mutated metastatic melanoma treatment (Chapman et al., 2011). Many small molecule inhibitors focus on targeting the dysregulated signalling pathways, kinases or cell cycle checkpoints associated with cancer. These exploitable characteristics are increasingly being revealed by our expanding knowledge surrounding the abnormal biology and genetics of cancers which has accelerated through high-throughput genome sequencing studies (Hoelder et al., 2012). In 2008 Jones et al. performed the first in-depth genetic sequencing of 24 PDAC tumours to identify mutated and dysregulated pathways characteristic of the disease. They found that pancreatic cancer contained on average 63 genetic alterations which defined a core set of 12 cellular signalling pathways and processes that were altered (Jones et al., 2008). The most comprehensive study to date by Biankin et al. sequenced 99 PDAC tumours and reaffirmed the previously identified mutated genes as well as novel genes involved in chromatin modification, DNA damage repair and pathways such as axon guidance and SLIT/ROBO signalling (Biankin et al., 2012). Among those identified, changes in RAS, Hedgehog, Wnt, TGF-β, and JAK-STAT pathway are recognised as key contributors in PDAC progression. Since these breakthrough informative studies, there have been many clinical trials centred round

these altered elements of PDAC; however, most of these have failed during phase II/III as no significant improvement of OS was observed (Adamska et al., 2017). As aforementioned, the only FDA approved targeted therapeutic for PDAC remains to be the EGFR targeting small molecule erlotinib when used in combination with gemcitabine for treatment of locally advanced, unresectable or metastatic PDAC.

The progress with small molecule drugs as targeted therapies is mirrored by the successful introduction of protein-based therapeutics particularly monoclonal antibodies (mAbs). In 2013, global sales revenue for all monoclonal antibody products was nearly $75 billion, representing approximately half of the total sales of all biopharmaceutical products (Ecker et al., 2015). The unprecedented success of trastuzumab for the treatment of HER2 positive breast cancer (Slamon et al., 2001) inspired and accelerated research of antibody-based drug therapies as well as analyses of extracellular proteins of cancer cells. The inherent capacity for specificity of mAbs dubbed them "magic bullets" 30 years ago (Brodsky, 1988). MAbs have been used for treating diseases since the 1980s with the first fully human mAb approved for the treatment of rheumatoid arthritis in 2002. Since the development of transgenic mouse platforms which enable production of fully human antibodies, past issues regarding immunogenicity have been eradicated. Therapeutic mAbs can mediate their actions by various types of direct or indirect mechanisms (Weiner et al., 2010). Direct mechanism typically refers to binding of the mAb to an antigen on the surface of a cancer cell to elicit their effects by triggering events such as internalisation, degradation, dimerization. Other direct mechanisms include complement-dependent cytotoxicity and antibody-dependent cytotoxicity. Indirect mechanisms refers to the subsequent response of the body's defence mechanisms upon mAb binding such as the recruitment of effector cells with the capacity for antibody-dependent cellular cytotoxicity or phagocytosis causing cancer cell death (Foltz Ian N. et al., 2013). Therapeutic antibodies can effectively kill cancer cells in an independent manner such as trastuzumab therapy, or by combination approaches such as conjugation with a toxic agent which increases their anti-tumour efficacy.

Antibodies conjugated to a cytotoxic payload such as a chemotherapeutic agent are referred to as antibody-drug conjugates (ADCs). In ADC therapy, a toxin is linked to an antibody as it homes in on its target plasma membrane protein expressed on tumours cells. After antibody binding to the cell surface antigen the ADC complex is typically internalised, payload and all, allowing the toxin to be released intracellularly where it can interfere with various cellular mechanisms to induce cell death (Sievers and Senter, 2013). ADCs are a potent therapy for cancer treatment as they are highly selective and allow specific delivery of cytotoxic agents to the intended cancer cell target. Depending on the specificity of the target protein, ADCs can in theory discriminate between disease and healthy tissues thus reducing off-set toxicity and damage to normal cells. The first FDA approved ADC 'Mylotarg' was developed in 2001 but later withdrawn in 2010 following a clinical trial that revealed patients died with no improvement to remission or OS compared to standard therapy (Petersdorf et al. 2013). The fatal toxicity of Mylotarg was caused by the release of the cytotoxic agent into the patients' bloodstreams prior to reaching the target site as a result of premature linker cleavage. The linker technology conjugating the antibody and toxic payload thus revealed itself to be a subtle but significant factor in the development of ADCs for controlling delivery. In 2013 a second generation of FDA approved ADCs with improved linker technology were released, namely Kadcyla and Adcetris. The success of Adecetris for Hodgkin's lymphoma treatment is founded on the ADC's target, CD30 which is strongly expressed on the lymphoma tumour cells and otherwise limited to activated T and B cell expression (Falini et al., 1995, p. 30). Kadcyla was the first ADC to receive FDA approval for a solid tumour; it is used for the treatment of HER2 positive metastatic breast cancer patients (Amiri-Kordestani et al., 2014, p. 2). Kadcyla built on the success of trastuzumab for breast cancer by conjugating this antibody to the cytotoxic microtubule inhibitor DM1 using linker technology to create an impressively effective ADC molecule for the improved treatment of HER2 positive breast cancer (Lewis Phillips et al., 2008, p. 1).

Currently more than 50 ADCs are in clinical trials with 70% of these having entered in the last 3 years (Nasiri et al., 2018, p.). Although the focus of ADCs has remained in the

treatment of haematological cancers, the success of Kadcyla in breast cancer gives confidence to the clinical development of novel ADC therapies for solid tumours.

1.2.3 Application of proteomics for PDAC improvement

The lack of improvement in survival rates for patients of PDAC over the last 40 years (figure 1.3) highlights the importance of performing PDAC research. While the majority of cancers have seen vast improvements in survival outcomes, PDAC remains one of the toughest cancers to overcome. The unfortunate combination of several factors which are now characteristic of PDAC, culminate to deliver a bleak outcome. PDAC is considered a low-incidence but high-mortality disease. The low survival rates are founded on the facts that (i) early diagnosis is unachievable as there is no biomarker(s) approved for diagnostic screenings (ii) most patients when ultimately diagnosed are at an advanced and/or metastatic, inoperable, stage of the disease and (iii) chemotherapeutic treatment options for PDAC show minor improvements to overall survival rates, toxic side effects and often high levels of resistance thus making it more for palliative care than an option for remission.

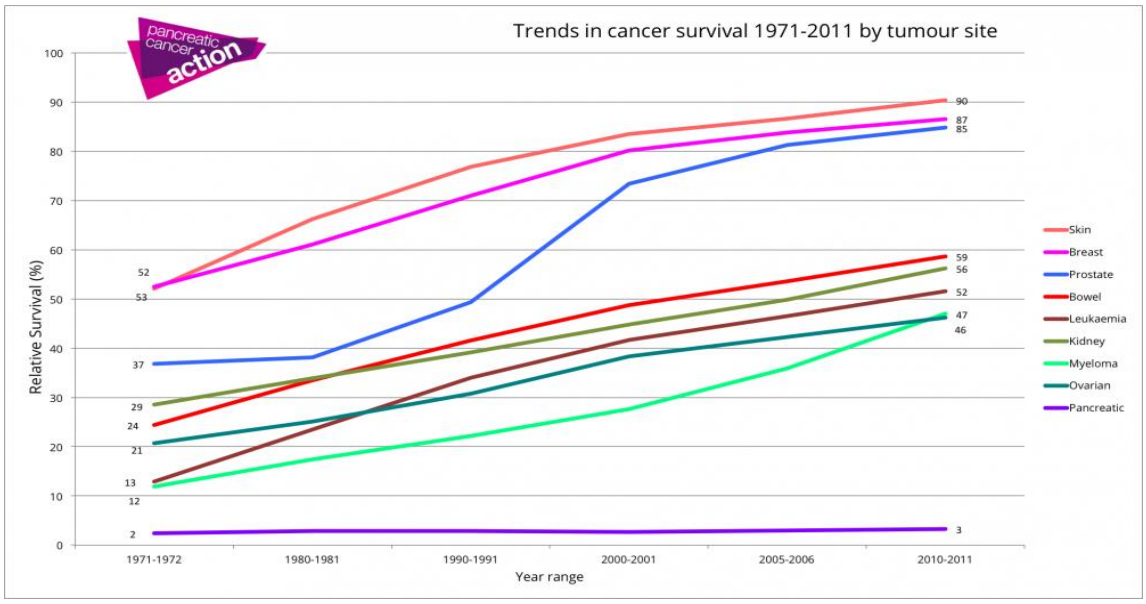


Figure 1.3 Trends in cancer survival rates from 1971-2011 for 9 major cancers. Pancreatic cancer survival has only slightly increased from 2% to 3% over 40 years while cancers such as skin, breast and prostate have increased to over 85% survival rate. Graph taken from the Pancreatic Cancer Action Network charity site www.pancan.org

Evidence from some cancer types such as cervical and colon cancer, suggests that earlier detection of cancer will improve survival (Zauber et al., 2012) (Gustafsson et al., 1997). Despite the apparent advantages of screening the general population for cancers, progress in this field has been limited and only a handful of cancer screening tests have clinical value and approval ("Early detection," 2018). Early detection of PDAC may provide an opportunity to cure or at least greatly improve survival as evidenced by Egawa et al. who showed surgically resected stage I PDAC resulted in a median survival time of 78.2 months and a 5-year survival rate of 58.1% (Egawa et al., 2004). Therefore, development of an early diagnostic test is a major unmet need in PDAC research which may improve the survival outcome of patients. Proteomics is a high-throughput, sensitive and accurately quantitative method which can be applied for the discovery of candidate biomarkers. Mass spectrometry-based proteomics is increasingly contributing to our understanding of the dynamics and roles that proteins play, advancing our understanding of biology on a systems wide level for a range of diseases (Angel et al., 2012). In the discovery phase of biomarker development the objective is to profile model systems or patient samples to identify candidates with differential abundance between PDAC and normal, PDAC and chronic pancreatitis, or PDAC and other diseases of the pancreas (Root et al., 2018). Cell lines, mouse models, and organoids can adequately recapitulate PDAC disease progression and are often more amenable to intervention studies and profiling analyses than primary samples from human patients. Primary patient samples are, however, the best sample type for profiling studies to identify candidate biomarkers. A complete review of proteomic profiling studies for biomarker discovery in PDAC cell lines, primary tissues and secretome is provided in chapter 3 (Coleman et al., 2016). Proteomics is a key tool for profiling PDAC and has already identified many promising proteins which have potential as diagnostic or predictive biomarkers for patients of PDAC. The use of novel, state-of-the-art mass spectrometers is increasing our knowledge surrounding the biology and development of many diseases. The application of modern proteomics to PDAC may succeed in identifying a reliable biomarker for the early detection of this disease which remains a huge challenge.

Another key factor contributing to the poor survival rate for PDAC patients is the lack of any highly effective, significantly life-prolonging treatments. Although the most advantageous treatment in terms of survival outcome is surgical resection of the tumour, only 20% of patients present at a stage amenable to resection. This again highlights the importance of early diagnosis through biomarker identification and approval which may be made possible by proteomic profiling studies. Standard chemotherapy treatments approved for PDAC (discussed in section 2.2.1) only provide an increased survival rate amounting to a handful of months. To achieve a significant improvement to the survival outcome of PDAC patients, new therapeutic options must be identified and implemented. As described in section 2.2.2, novel therapies which strategically target the tumour site have had huge successes in the treatment of a plethora of cancers. Targeted therapeutics rely on the evaluation of tumour profiles to identify targets specific to the disease to deliver the cytotoxic drug to that site with minimal off target effects. The challenge of profiling and identifying tumour targets necessitates the use of novel, in-depth approaches to define cancer cell targets leading to the development of significantly improved therapeutic treatments. The application of MS-based proteomics again provides a viable and promising approach for accurately profiling and quantitating the proteome of PDAC. Comparative analyses can help to identify targets specific or at least highly expressed in PDAC to improve the efficiency of targeted therapeutics. The development of ADCs, among other novel targeted therapeutics, provides great promise for the implementation of a more effective treatment option for PDAC. The successful development of an ADC is dependent on the idea of selecting an appropriate target antigen, with its suitability forming a key determinant of the efficacy of the ADC. Targeted therapies such as ADCs exploit the difference in protein expression between cancer cells and normal cells. It is widely accepted that the target antigen should be selectively expressed on the surface of tumour cells with little or no expression on normal tissues in order to limit off-tumor toxicity (Damelin et al., 2015). The application of proteomics can identify such cell surface antigens in PDAC and thus may help in the development of effective targeted therapeutic such as an ADC for PDAC treatment.

1.2.3.1 Membrane proteomics

The identification of a reliable diagnostic biomarker, and/or a suitable antigen for targeted therapeutic development, could both significantly improve the survival outcome for patients of PDAC. Both of these current absences have the potential to be solved through the application of proteomics. This thesis specifically focused on the application of membrane proteomics for profiling PDAC. Membrane proteins have two desirable characteristics which may be exploited for PDAC improvement; (1) occasionally being shed or cleaved and released into the bloodstream (HAYASHIDA et al., 2010; Tien et al., 2017), and (2) the favourable location of plasma membrane proteins on the surface of cells (Cooper, 2000). These characteristics make membrane proteins uniquely special in that they have the potential to be either or both a non-invasive diagnostic biomarker if released into the bloodstream of patients, and a target antigen for targeted therapeutics if highly expressed on the cell surface of tumours cells.

Membrane proteins are known to be involved in cancer. Membrane receptor proteins such as the human epidermal growth factor receptor (ErbB/HER) family have been extensively studied and shown to play important roles in cell signalling as well as cancer development and progression when dysregulated. The most notable pathways activated by the HER family include the Ras/MAPK pathway, the PI3K/Akt pathway and the JAK/Stat signaling pathway all of which have been described to play crucial roles in tumorigenesis (Wieduwilt and Moasser, 2008). The overexpression of two member proteins of the ErbB/HER family, HER2 and EGFR, is correlated with poor prognosis in breast cancer highlighting the importance of some membrane proteins in cancer development. Many other studies have identified several membrane proteins to be associated with various cancer types such as glioblastoma (Ghosh et al., 2017), colon cancer (Chiang et al., 2014), prostate cancer (Amin et al., 2018), ovarian cancer (Van Simaeys et al., 2014), small-cell lung cancer (Ocak et al., 2014) and cervical cancer (Pappa et al., 2018).

Despite compromising approximately 30% of the encoded human genome (Wallin and Heijne, 1998), membrane proteins are often under-represented in many proteomic

studies. The under-representation of membrane proteins in proteomic profiling studies is mainly due to the hydrophobic and low abundance nature of these proteins (Tan et al., 2008). Successful membrane protein studies are founded on efficient extractions of membrane proteins. The traditional method for extraction of membrane proteins is a multi-step detergent extraction coupled with gradient based ultracentrifugation (Tanford and Reynolds, 1976). This method and variations of it are extremely time-consuming, require large amounts of starting sample material and specialised equipment. More recently, the introduction of commercial membrane protein extraction kits has enabled a faster, cheap and easier method, however, membrane protein fractions do not have the same purity as traditional methods.

1.2.4 Preface to chapters 2 and 3

The following chapters 2 and 3 are two published articles based on the topic of proteomic profiling of PDAC. Chapter 2 is a review article outlining proteomic techniques for clinical proteomic studies and describes proteomic studies in pancreatic cancer cell lines, tissues and the secretome. This review was published after an extensive literature survey of proteomic studies seeking to identify novel biomarkers and drug targets for PDAC. The aim of this part of my thesis is apply MS-based proteomics to profile PDAC to identify potential biomarkers or targets for novel targeted therapeutics, therefore this review article allowed me to understand the current status of the field and identify challenges and pitfalls of previous studies. The majority of studies described in the review did not employ membrane protein profiling to identify biomarkers or drug targets. This was therefore identified as an important angle in PDAC research which was employed for all analyses of PDAC undertaken in this thesis. Chapter 2 was fully researched, designed and written by me.

Chapter 3 is a research article describing the proteomic analysis of primary PDAC tumours, adjacent-normal pancreas tissue and patient-derived tumour xenografts (PDXs). This research article contains the main body of PDAC research performed in this thesis. The analysis of primary tumour samples is a strong advantage of this paper as many PDAC analyses are limited to cell line or model systems. The PDX tumours selectively retain the primary tumour cells and replace the stromal component of the

primary tumour with mouse stroma. This selectivity of the PDX model was exploited in the proteomic analysis. Protein identification was achieved by searching the raw data against a human and mouse dual database due to the mixture of both human and mouse cellular components in the PDX tumour samples. The inherent orthologous nature of mouse and human species resulted in approximately two-thirds of the protein data being sequence identical to both species therefore prohibiting elucidation of the species of origin. However, one-third of the protein identifications were species-specific due to sequence alterations. Forty samples were analysed in this study using a membrane protein enrichment strategy for protein extraction followed by LC-MS/MS using an LTQ-Orbitrap XL mass spectrometer. Label-free quantitative analysis was used to identify differentially expressed proteins between the various sample groups. The identification of overexpressed proteins associated with PDAC presents candidate proteins which may have potential as novel biomarkers and/or drug targets. The lists of human-specific protein identifications that were overexpressed in PDX tumours after engraftment provide a rich source of information as we know these proteins are tumour cell derived and not contaminating stromal proteins. This work was carried out with the collaboration of clinicians and pathologists in St. Vincent's University Hospital, Dublin and St. Luke's hospital Dublin. The research described in this article was principally performed by me. Sample preparation, quantitative data analysis, Western blot validation, manuscript preparation and submission were undertaken by me.

1.3 MicroRNA engineering and proteomic profiling of recombinant CHO cells

A biopharmaceutical is any product manufactured in, extracted from, or synthesised from a biological source. The majority of biopharmaceutical therapeutics are produced using recombinant DNA technology which brings together DNA sequences, usually from different species or a synthetic DNA sequence, to create sequences that would not otherwise be found in the host genome. The resulting protein is referred to as a recombinant protein. The ability to produce mammalian proteins in recombinant systems has had a profound impact in many areas of basic and applied research. The production of recombinant protein therapeutics (i.e. biologics) offers a wide range of possibilities for therapeutic applications in the treatment of human cancer and diseases. Although a variety of expression systems exist, such as micro-organisms, animal cell lines, plants and animals, mammalian cells are the dominant host choice for the commercial production of therapeutic proteins. Microbial systems, especially *E. coli*, have the advantage of low cost in establishing a production strain, quick production cycle and high productivity compared to mammalian expression systems (Zhu, 2012). However, there are various limitations for prokaryotic systems like *E. coli* such as difficulty in producing large complex proteins and the absence of eukaryotic post-translational modifications (PTMs) which mimic human PTMs (Sahdev et al., 2008). The post-translational metabolic machinery required for events like glycosylation, which is often required for optimal biological function of a mAb or fusion protein, is only available in eukaryotic mammalian cells (Beck et al., 2008). The most widely used mammalian host cells are Chinese hamster ovary (CHO) cells, mouse myeloma cells (e.g. NS0 and Sp2/0) and human embryonic kidney cells (HEK293); however, CHO cells are most frequently employed in the biopharmaceutical industry (Walsh, 2018).

Chinese hamster ovary (CHO) cells were first isolated 60 years ago in 1958 from a biopsy of an ovary of an adult Chinese hamster (Puck et al., 1958). Genentech obtained the first approval for CHO cells as recombinant hosts in 1987 for the commercial production of human tissue plasminogen activator (tPA), the first recombinant therapeutic protein produced from mammalian cells (Finkle, 1988). At present, almost

70% of all recombinant protein therapeutics are produced in CHO cells (Jayapal et al., 2007). In 2017 the top 10 selling drugs worldwide were dominated by mAb therapeutics and 5 of the top 10 selling drugs were produced in CHO cell lines (Figure 1.4.) (Urquhart, 2018).

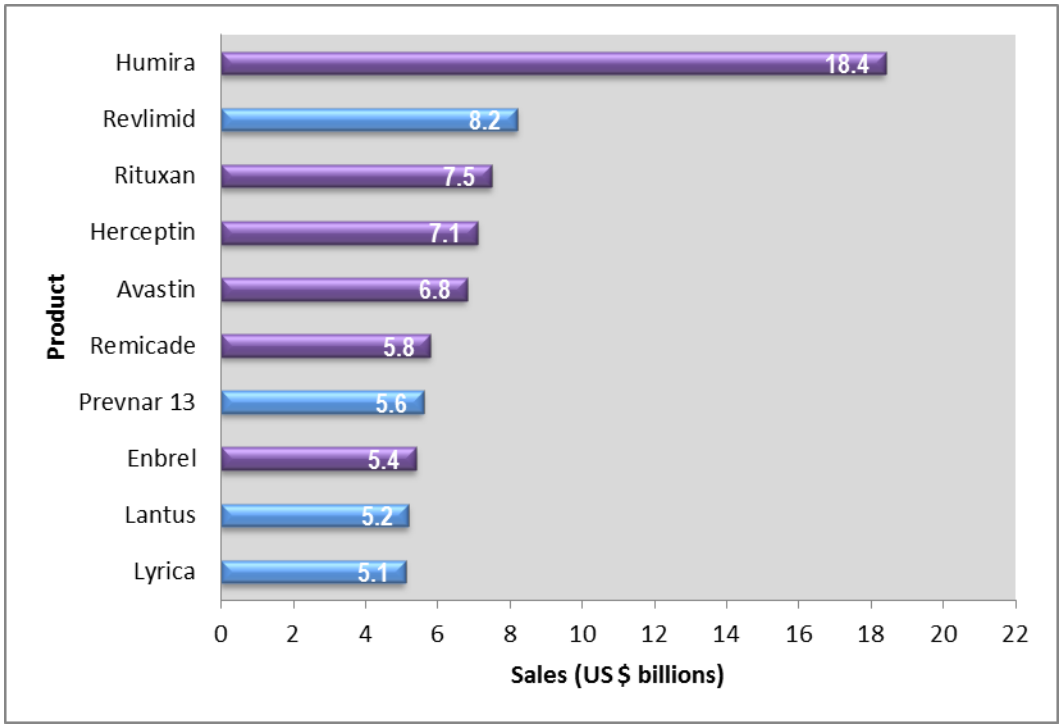


Figure 1.4 Top 10 selling drugs globally in 2017. Monoclonal antibody therapeutics are shown in purple. Humira, Rituxan, Herceptin, Avastin and Enbrel are all produced in CHO cell lines. Graph adapted from Urquhart 2018.

The popularity of CHO cells can be attributed to multiple characteristics. CHO cells can produce recombinant proteins with glycosylation modifications that are both compatible with and bioactive in humans. They are a safe host system with resistance to a vast majority of human pathogenic viruses (Taber et al., 1976). From an industrial perspective their ability to grow in suspension culture is highly desirable as it allows volumetric scalability. Most importantly CHO cells are amenable to genetic modifications allowing for easy introduction of foreign DNA and expression of large quantities of the desired protein. The isolation of CHO cell lines deficient in DHFR and GS in the 1980s (Sanders et al., 1987; Urlaub et al., 1983) led to an effective method for selecting stable clones and amplifying the gene of interest by co-transfection which led

to dramatic increases in specific productivity levels. Over the past 20 years further optimisation strategies for CHO cell production included improvement of media formulations, feeding strategies and bioprocessing regimens. However these efforts remain somewhat limited as productivity yields for CHO cells remains typically 10-100 fold lower than what can be achieved in microbial systems (Schmidt, 2004) (Jayapal et al., 2007). Although optimisation of the external properties surrounding biopharmaceutical production has been successful, the next major steps towards improving production must focus on optimisation of the CHO cell itself (Kuystermans et al., 2007). In order to achieve this, a deeper understanding and profiling of the CHO cell on a multi-layered omics level must be performed (Datta et al., 2013). A deep biological understanding of CHO cell properties will enable cellular optimisation of bioprocessing.

1.3.1 CHO cell engineering

Current approaches to meet the ever-increasing demand for therapeutic drugs are focused on CHO cell engineering. Omics, CHO genome sequencing, and developments in genome editing tools have enabled a more direct route in cell line optimisation through the understanding and modification of specific genes or pathways that directly affect cell culture performance or protein quality. Despite the use of CHO cells for several decades as a host system, the first genome from CHO K1 cells was only sequenced in 2011 and the related Chinese hamster genome was reported in 2013 (Lewis et al., 2013; Xu et al., 2011). In addition to genomics, data surrounding the transcriptome, miRnome, proteome, phosphoproteome and glycoproteome of CHO cells have become available recently through significant and collaborative efforts of the CHO cell research community (Baycin-Hizal et al., 2012; Becker et al., 2011; Clarke et al., 2012b; Hackl et al., 2011; Henry et al., 2017; Yang et al., 2018). The information gathered by these efforts has advanced the fundamental knowledge of CHO cell biology and drivers of advantageous phenotypes. CHO cell characterisation through multi-omic approaches has opened the doors for genetic engineering to potentially establish superior production cell factories.

Genetic engineering of CHO cells traditionally involves either overexpression of beneficial genes or repression of disadvantageous genes by complete genomic knock-out or siRNA-mediated knock-down. These single gene targeted engineering approaches have not only focused on increasing protein expression of the desired product, but also manipulation of other important cell culture performance features such as apoptosis, glycosylation, metabolism, secretion, cell cycle and proliferation (Fischer et al., 2015). Cell death is an important feature for genetic engineering as it affects viable cell densities which forecasts product yield. Fortunately, previous research of other eukaryotes including humans, has identified proteins involved in activating and inhibiting apoptosis allowing straightforward CHO cell engineering to limit apoptosis (Arden and Betenbaugh, 2004). Overexpression of anti-apoptotic proteins that include Bcl-2, Bcl-xL, XIAP and Mcl-1 (Majors et al., 2009; Mastrangelo et al., 2000; Meents et al., 2002; Sauerwald et al., 2002), or down-regulation of pro-apoptotic proteins such as Bak and Bax (Cost et al., 2010; Grav et al., 2015), have resulted in improved apoptosis resistance and/or increased protein production in CHO cells. Similarly pro-proliferative genes have been overexpressed to increase viable cell densities in CHO cells (Doolan et al., 2010; Jaluria et al., 2007). Glycosylation profiles of proteins, as previously mentioned, are important for the functionality and efficacy of mAbs and although CHO cells can post translationally modify proteins the patterns are only human-like and not human-identical (Butler and Spearman, 2014). Genetic engineering of various genes involved in glycosylation and fucosylation have improved the glycan profiles and toxicity of recombinant mAbs produced in CHO cells (Lee et al., 1989; Prati et al., 2002; Yamane-Ohnuki et al., 2004).

The ultimate goal of CHO cell engineering is to increase productivity of CHO cells to meet the increasing demands for therapeutics in a cost-effective fashion and reach yields similar to competitor prokaryotic host systems. Protein production in CHO cells has been greatly enhanced through the introduction of genes involved in protein synthesis machinery and/or secretion. Overexpression of various transcription factors such as ZFP-TF, ATF4 or GADD34 has increased protein production by up to 10-fold in some cases (Kwon et al., 2006; Ohya et al., 2008; Omasa et al., 2008). The mTOR pathway has been extensively studied for its regulation of protein synthesis in eukaryotes through

various components (X. Wang and Proud 2006). Constitutive expression of mTOR in various CHO cell lines resulted in increased viability, robustness, cell size, proliferation, and antibody production (Dadehbeigi and Dickson, 2015; Dreesen and Fussenegger, 2011). Inhibition of mTOR complex 1 at the start of batch culture in CHO K1s enhances product yields and increase the longevity of culture by slowing cell growth (Dadehbeigi and Dickson, 2015). Another strategy employed to increase yields during CHO cell production is engineering the secretory pathway to increase the secretion rate of recombinant therapeutic proteins. The overexpression of various genes involved in secretion such as XBP1, SLY1, CERT, VAMP8 or SRP14 (Florin et al., 2009; Le Fourn et al., 2014; Peng et al., 2011; Peng and Fussenegger, 2009; Tigges and Fussenegger, 2006) have been shown to substantially increase product yield in various CHO cell lines.

1.3.1.1 Engineering CHO cells with microRNAs

Although remarkable achievements in CHO cell engineering over the past few decades has increased cell culture performance it has largely resulted from single gene engineering. Phenotypic changes during CHO cell culture, whether advantageous or disadvantageous, are unlikely to be caused from a single gene alteration rather more likely a plethora of genes or pathways affecting various culture features. CHO cell engineering should therefore rationally shift its focus from one particular gene of interest or attribute to multiple gene targeting approaches to simultaneously exploit various aspects of cell culture performance.

MicroRNAs (miRNAs) have recently become a novel angle for CHO cell engineering (Barron et al. 2011; Müller, Katinger, and Grillari 2008; Fischer et al. 2017). MiRNAs are a class of small non-coding RNAs which are highly conserved over species and have been shown to regulate multiple gene targets (Ambros et al., 2003). Since their discovery in the 1990s, miRNAs have been shown be involved in a myriad of important cellular processes and various cancers (Sun et al., 2010; Jansson and Lund, 2012; Giza et al., 2014; Liz and Esteller, 2016). Studies revealed that miRNAs can regulate gene expression in a post-transcriptional manner by either inhibiting translation of its mRNA target or degradation of it. The regulatory capabilities of miRNAs make them an

exciting option for CHO cell engineering. Advantageously, the non-coding nature of miRNAs means that they do not contribute to host cell translational burden. Secondly, the small effector or "seed" sequence of each miRNA means a single miRNA can target multiple genes (Bartel, 2009); this makes engineering of single miRNAs a more efficient means of affecting global cellular behaviour than single gene strategies (Druz et al., 2013). Subsequent to the release of sequencing information for CHO miRNAs, numerous miRNAs have been identified and studied as potential "engimiRs" in CHO cells (Hackl et al., 2011).

Constitutive exogenous miRNA intervention has been shown to be an effective method for engineering CHO cell behaviours. Finding miRNAs to target for cell engineering can be achieved by profiling differential miRNA abundance and correlating this to desirable cell phenotypes (Clarke et al., 2012) or in a shotgun approach via transient high-content functional screening (Fischer et al., 2014). Constitutive miRNA over-expression or repression has been used to engineer CHO cell specific productivity (Barron et al., 2011; Meleady et al., 2012; Sanchez et al., 2014; Fischer et al., 2015; Kelly et al., 2015; Emmerling et al., 2015; Fischer et al., 2017), inhibit apoptosis (Druz et al., 2011; Kelly et al., 2017) and enhance cell proliferation (Fischer et al., 2014; Strotbek et al., 2013).

1.3.1.2 Application of proteomics

The majority of miRNA studies to date have focused on either determining the mechanism of miRNA mediated gene regulation and/or identifying miRNA targets. This has been accomplished through a variety of techniques such as next generation sequencing, gene analysis microarrays, *in silico* predictions and mass spectrometry based proteomics (Thomson et al., 2011). MiRNA have been found to impact protein output (Baek et al., 2008) thus proteomic profiling approaches have an inherent advantage of assessing the ultimate effect of miRNAs. Quantitative proteomics has emerged as a key technique for experimental identification of miRNA targets by allowing direct determination of proteins whose levels are altered because of miRNA-mediated translational suppression. In addition to identifying individual miRNA targets,

the application of quantitative proteomics can also reveal signaling networks and pathways affected by a miRNA.

The vast majority of studies to date that report miRNA targets are based on computational prediction approaches or altered mRNA levels. The exact mechanism of miRNA target recognition remains largely unknown and as a result bioinformatic algorithms for the prediction of miRNA targets are not completely accurate and often over-predict miRNA targets (Huang et al., 2013). Popular prediction tools such as TargetScan, miRanda, TarBase and miRDB mainly predict miRNA targets based on sequence pairing of the zone between nucleotides 2 to 8 (termed the seed region) of miRNA to the 3' UTR of mRNA (Riffo-Campos et al., 2016). Increasing evidence, however, reports that miRNA binding to mRNA targets can occur at various sites such as the 5' UTR and even coding regions (Grimson et al., 2007; Schnall-Levin et al., 2010; Shin et al., 2010a). Therefore miRNA target prediction by computational algorithms alone is often not sufficient in cataloguing all potential mRNA targets of miRNA. Similarly it has been found the miRNA can regulate gene expression by mRNA storage without mRNA degradation (Pillai et al., 2005) indicating that mRNA quantitation may not be entirely accurate for identifying miRNA targets. It is therefore essential to experimentally identify targets of miRNA through powerful methods such as MS-based proteomics. In the context of CHO cell engineering, the identification of miRNA targets can form a basis for directed engineering of a single gene or pathway to further enhance desirable phenotypes. The multifaceted nature of miRNAs allowing them to target hundrerds of mRNAs can potentially mean an mRNA of importance involved in, for example, proliferation may be repressed by the same miRNA that represses another mRNA involved in anti-apoptosis. The selection of beneficial miRNAs for CHO cell culture performance enhancement will require a complete understanding of the biological effects of individual miRNAs through target identification. Any adverse effects on protein production and/or quality would eliminate any advantage of miRNA-induced enhanced CHO cell phenotypes.

1.3.2 Preface to chapters 4 to 7

Chapters 4 to 7 are a set of published works performed during my PhD. These chapters are all focused on the proteomic profiling of CHO cells following miRNA manipulation. Proteomic analysis of CHO cell lines performed in these chapters not only aim to identify miRNA targets but also, through in-depth comprehensive fractionation analyses an insight into the proteome of CHO cells displaying improved cell culture phenotypes has been provided. The information gathered by these analyses has advanced the knowledge of the CHO cell proteome and identified drivers of advantageous phenotypes.

Chapter 4 is a methodology book chapter published in Springers *Methods in Molecular Biology* book series entitled "Heterologous Protein Production in CHO cells". This methodology was adapted from the original method released by Wisniewski et al. in 2009 (Wiśniewski et al., 2009). I have applied this method, in large, to the two subsequent research articles. This method for protein sample preparation of CHO cells provides an efficient and reproducible method of sample preparation prior to label-free mass spectrometry. The absence of an internal standard spiked early in sample preparation protocols means that label-free methods are sensitive to technical variance; consequently, label-free proteomics requires high instrument performance and standardisation of sample preparation methods. The main objective for a label-free sample preparation method is to obtain stable peak intensities between replicate sample preparations to ensure accurate quantitation. The application of this method for the proteomic analyses of CHO cells performed in this thesis ensured a robust and high-throughput method of sample preparation with minimal technical variance.

Chapter 5 and 6 are two research articles which stemmed from one study on the effect of miR-378 depletion on CHO cells and the knock-on proteomic alterations. MiR-378 was previously identified by a member of our group as downregulated in CHO cell clones which exhibit an increased cellular growth rate (Clarke et al., 2012b). For the study conducted during my PhD, stable depletion of miR-378 in the CHO DP12 cell line resulted in a 59% increase in viable cell growth. The aim of this study was to provide a

comprehensive proteomic dataset to elucidate miR-378 targets. Cell samples were collected for two culture time points; day 4 and day 8. Subcellular protein enrichment was performed to yield 2 protein fractions per sample and resulted in increased proteomic coverage. Quantitative data analysis identified over 100 significantly differentially expressed proteins with approximately 70 proteins upregulated miR-378 depleted CHO DP12 cells in both the cytosolic and membrane-enriched fractions. In silico prediction of miR-378 targets followed by mRNA and protein level validation of target upregulation identified Usp14 as a strong candidate target of miR-378. Direct overexpression of Usp14 in CHO DP12 cells had a significant increase on cell growth supporting a role of Usp14 in the increased peak cell density seen with miR-378-3p depletion. My involvement in chapter 5 was the proteomic analysis of miR-378 depleted CHO DP12 cells by performing cell sample collection, protein sample preparation and all associated proteomic data analyses. The work regarding miRNA analyses and UPR response was performed by my co-author Alan Costello. The subsequent chapter 6 is a data article which was suggested during the submission process of the research article to *Journal of Biotechnology*. The MS data including both qualitative and differentially quantitative analyses for each condition and culture time point were published in this data article to facilitate a thorough explanation of the proteomic data separately and increase visibility. All of the included work in chapter 6 was performed by me during my PhD as well as subsequent manuscript design, preparation and submission.

Chapter 7 is an in-depth proteomic analysis of the miRNA-7 depleted CHO DP12 cell line. MiRNA-7 has been extensively studied in our group for its impact on CHO cell culture growth and productivity (N. Barron et al., 2011; Meleady et al., 2012a; Sanchez et al., 2013, 2014a, p. 7). MiR-7 was initially identified as downregulated in a temperature-shift study whereby CHO cultures grown at 31 degrees showed enhanced growth and recombinant protein productivity. The connection between miR-7 and these desirable cultures phenotypes was further elucidated and conclusively miR-7 expression represses desirable CHO culture phenotypes, thus depletion of endogenous miR-7 enhances CHO performance for both cell growth and productivity. The study undertaken in my PhD was to identify targets of miR-7 through quantitative MS-based proteomics

when miR-7 was stably depleted from the the mAb-producing CHO DP12 cell line. Stable depletion of miR-7 in this study resulted in a 65% increase in cell growth and >3-fold increase in mAb product yield compared to control cells. The aim of the study was to provide a comprehensive proteomic dataset to elucidate miR-7 targets. For a complete analysis of the culture effects of miR-7 two time points during batch culture; exponential and stationary phase of cell growth, were analysed by cell collection on day 3 and day 8. To attain an in-depth proteomic profile, fractionation by a subcellular enrichment approach was performed to increase the protein identifications; this was successfully achieved. In addition to the preparation of subcellular protein-enriched fractions, whole cell total proteomic analysis was also performed. High-throughput, sensitive mass spectrometry analysis was achieved by the use of the Orbitrap Fusion Tribrid mass spectrometer. This mass spectrometer identified over 3000 proteins per sample in a 140-minute gradient. The added strategy of subcellular protein enrichment as a fractionation method prior to MS resulted in over 5000 proteins identified per condition. Quantitative label-free analysis identified significantly differentially expressed proteins for both the early and late, i.e. exponential and stationary phases of CHO cell culture. Proteins that were upregulated following stable depletion of miR-7 were potentially driving the improved phenotype observed in these CHO DP12 cells. Over 100 proteins during early and late stage of culture were significantly upregulated over 1.5-fold times in miR-7 depleted cells. Analysis of these proteins suggested that the Akt pathway which promotes cell survival and growth, and ribosome biogenesis associated proteins were upregulated and contributing to the observed improved phenotype. My involvement in this study was cell sample collection, protein sample preparation, all data analyses, literature research, manuscript design preparation and submission.

1.4 References

Adamska, A., Domenichini, A., Falasca, M., 2017. Pancreatic Ductal Adenocarcinoma: Current and Evolving Therapies. Int. J. Mol. Sci. 18. https://doi.org/10.3390/ijms18071338

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

Agarwal, V., Bell, G.W., Nam, J.-W., Bartel, D.P., 2015. Predicting effective microRNA target sites in mammalian mRNAs. elife 4, e05005.

Al Shweiki, M.R., Mönchgesang, S., Majovsky, P., Thieme, D., Trutschel, D., Hoehenwarter, W., 2017. Assessment of Label-Free Quantification in Discovery Proteomics and Impact of Technological Factors and Natural Variability of Protein Abundance. J. Proteome Res. 16, 1410–1424. https://doi.org/10.1021/acs.jproteome.6b00645

Altamirano, C., Paredes, C., Illanes, A., Cairo, J.J., Godia, F., 2004. Strategies for fed-batch cultivation of t-PA producing CHO cells: substitution of glucose and glutamine and rational design of culture medium. J. Biotechnol. 110, 171–179.

Ambros, V., Bartel, B., Bartel, D.P., Burge, C.B., Carrington, J.C., Chen, X., Dreyfuss, G., Eddy, S.R., Griffiths-Jones, S., Marshall, M., Matzke, M., Ruvkun, G., Tuschl, T., 2003. A uniform system for microRNA annotation. RNA N. Y. N 9, 277–279.

Amin, M.K.B.A., Shimizu, A., Zankov, D.P., Sato, A., Kurita, S., Ito, M., Maeda, T., Yoshida, T., Sakaue, T., Higashiyama, S., Kawauchi, A., Ogita, H., 2018. Epithelial membrane protein 1 promotes tumor metastasis by enhancing cell migration via copine-III and Rac1. Oncogene 37, 5416. https://doi.org/10.1038/s41388-018-0286-0

Amiri-Kordestani, L., Blumenthal, G.M., Xu, Q.C., Zhang, L., Tang, S.W., Ha, L., Weinberg, W.C., Chi, B., Candau-Chacon, R., Hughes, P., Russell, A.M., Miksinski, S.P., Chen, X.H., McGuinn, W.D., Palmby, T., Schrieber, S.J., Liu, Q., Wang, J., Song, P., Mehrotra, N., Skarupa, L., Clouse, K., Al-Hakim, A., Sridhara, R., Ibrahim, A., Justice, R., Pazdur, R., Cortazar, P., 2014. FDA approval: ado-trastuzumab emtansine for the treatment of patients with HER2-positive metastatic breast cancer. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. 20, 4436–4441. https://doi.org/10.1158/1078-0432.CCR-14-0012

Angel, T.E., Aryal, U.K., Hengel, S.M., Baker, E.S., Kelly, R.T., Robinson, E.W., Smith, R.D., 2012. Mass spectrometry based proteomics: existing capabilities and future directions. Chem. Soc. Rev. 41, 3912–3928. https://doi.org/10.1039/c2cs15331a

Arden, N., Betenbaugh, M.J., 2004. Life and death in mammalian cell culture: strategies for apoptosis inhibition. Trends Biotechnol. 22, 174–180. https://doi.org/10.1016/j.tibtech.2004.02.004

Artigues, A., Nadeau, O.W., Rimmer, M.A., Villar, M.T., Du, X., Fenton, A.W., Carlson, G.M., 2016. Protein Structural Analysis via Mass Spectrometry-Based Proteomics. Adv. Exp. Med. Biol. 919, 397–431. https://doi.org/10.1007/978-3-319-41448-5_19

Baek, D., Villén, J., Shin, C., Camargo, F.D., Gygi, S.P., Bartel, D.P., 2008. The impact of microRNAs on protein output. Nature 455, 64–71. https://doi.org/10.1038/nature07242

Bailey, J.M., Swanson, B.J., Hamada, T., Eggers, J.P., Singh, P.K., Caffery, T., Ouellette, M.M., Hollingsworth, M.A., 2008. Sonic Hedgehog Promotes Desmoplasia in Pancreatic Cancer. Clin. Cancer Res. 14, 5995–6004. https://doi.org/10.1158/1078-0432.CCR-08-0291

Banerjee, S., Mazumdar, S., 2012. Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte [WWW Document]. Int. J. Anal. Chem. https://doi.org/10.1155/2012/282574

Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., Kuster, B., 2007. Quantitative mass spectrometry in proteomics: a critical review. Anal. Bioanal. Chem. 389, 1017–1031. https://doi.org/10.1007/s00216-007-1486-6

Barron, N., Kumar, N., Sanchez, N., Doolan, P., Clarke, C., Meleady, P., O'Sullivan, F., Clynes, M., 2011. Engineering CHO cell growth and recombinant protein productivity by overexpression of miR-7. J. Biotechnol. 151, 204–211.

Barron, Niall, Sanchez, N., Kelly, P., Clynes, M., 2011. MicroRNAs: tiny targets for engineering CHO cell phenotypes? Biotechnol. Lett. 33, 11–21. https://doi.org/10.1007/s10529-010-0415-5

Bartel, D.P., 2009. MicroRNAs: target recognition and regulatory functions. Cell 136, 215–233. https://doi.org/10.1016/j.cell.2009.01.002

Basturk, O., Hong, S.-M., Wood, L.D., Adsay, N.V., Albores-Saavedra, J., Biankin, A.V., Brosens, L.A.A., Fukushima, N., Goggins, M., Hruban, R.H., Kato, Y., Klimstra, D.S., Klöppel, G., Krasinskas, A., Longnecker, D.S., Matthaei, H., Offerhaus, G.J.A., Shimizu, M., Takaori, K., Terris, B., Yachida, S., Esposito, I., Furukawa, T., 2015. A REVISED CLASSIFICATION SYSTEM AND RECOMMENDATIONS FROM THE BALTIMORE CONSENSUS MEETING FOR NEOPLASTIC PRECURSOR LESIONS IN THE PANCREAS. Am. J. Surg. Pathol. 39, 1730–1741. https://doi.org/10.1097/PAS.0000000000000533

Baycin-Hizal, D., Tabb, D.L., Chaerkady, R., Chen, L., Lewis, N.E., Nagarajan, H., Sarkaria, V., Kumar, A., Wolozny, D., Colao, J., Jacobson, E., Tian, Y., O'Meally, R.N., Krag, S.S., Cole, R.N., Palsson, B.O., Zhang, H., Betenbaugh, M., 2012. Proteomic Analysis of Chinese Hamster Ovary Cells. J. Proteome Res. 11, 5265–5276. https://doi.org/10.1021/pr300476w

Beck, A., Wagner-Rousset, E., Bussat, M.-C., Lokteff, M., Klinguer-Hamour, C., Haeuw, J.-F., Goetsch, L., Wurch, T., Van Dorsselaer, A., Corvaïa, N., 2008. Trends in glycosylation, glycoanalysis and glycoengineering of therapeutic antibodies and Fc-fusion proteins. Curr. Pharm. Biotechnol. 9, 482–501.

Becker, A.E., Hernandez, Y.G., Frucht, H., Lucas, A.L., 2014. Pancreatic ductal adenocarcinoma: Risk factors, screening, and early detection. World J. Gastroenterol. WJG 20, 11182–11198. https://doi.org/10.3748/wjg.v20.i32.11182

Becker, J., Hackl, M., Rupp, O., Jakobi, T., Schneider, J., Szczepanowski, R., Bekel, T., Borth, N., Goesmann, A., Grillari, J., Kaltschmidt, C., Noll, T., Pühler, A., Tauch, A., Brinkrolf, K., 2011. Unraveling the Chinese hamster ovary cell line

transcriptome by next-generation sequencing. J. Biotechnol. 156, 227–235. https://doi.org/10.1016/j.jbiotec.2011.09.014

Bhaw-Luximon, A., Jhurry, D., 2015. New avenues for improving pancreatic ductal adenocarcinoma (PDAC) treatment: Selective stroma depletion combined with nano drug delivery. Cancer Lett. 369, 266–273. https://doi.org/10.1016/j.canlet.2015.09.007

Biankin, A.V., Waddell, N., Kassahn, K.S., Gingras, M.-C., Muthuswamy, L.B., Johns, A.L., Miller, D.K., Wilson, P.J., Patch, A.-M., Wu, J., Chang, D.K., Cowley, M.J., Gardiner, B.B., Song, S., Harliwong, I., Idrisoglu, S., Nourse, C., Nourbakhsh, E., Manning, S., Wani, S., Gongora, M., Pajic, M., Scarlett, C.J., Gill, A.J., Pinho, A.V., Rooman, I., Anderson, M., Holmes, O., Leonard, C., Taylor, D., Wood, S., Xu, Q., Nones, K., Fink, J.L., Christ, A., Bruxner, T., Cloonan, N., Kolle, G., Newell, F., Pinese, M., Mead, R.S., Humphris, J.L., Kaplan, W., Jones, M.D., Colvin, E.K., Nagrial, A.M., Humphrey, E.S., Chou, A., Chin, V.T., Chantrill, L.A., Mawson, A., Samra, J.S., Kench, J.G., Lovell, J.A., Daly, R.J., Merrett, N.D., Toon, C., Epari, K., Nguyen, N.Q., Barbour, A., Zeps, N., Australian Pancreatic Cancer Genome Initiative, Kakkar, N., Zhao, F., Wu, Y.Q., Wang, M., Muzny, D.M., Fisher, W.E., Brunicardi, F.C., Hodges, S.E., Reid, J.G., Drummond, J., Chang, K., Han, Y., Lewis, L.R., Dinh, H., Buhay, C.J., Beck, T., Timms, L., Sam, M., Begley, K., Brown, A., Pai, D., Panchal, A., Buchner, N., De Borja, R., Denroche, R.E., Yung, C.K., Serra, S., Onetto, N., Mukhopadhyay, D., Tsao, M.-S., Shaw, P.A., Petersen, G.M., Gallinger, S., Hruban, R.H., Maitra, A., Iacobuzio-Donahue, C.A., Schulick, R.D., Wolfgang, C.L., Morgan, R.A., Lawlor, R.T., Capelli, P., Corbo, V., Scardoni, M., Tortora, G., Tempero, M.A., Mann, K.M., Jenkins, N.A., Perez-Mancera, P.A., Adams, D.J., Largaespada, D.A., Wessels, L.F.A., Rust, A.G., Stein, L.D., Tuveson, D.A., Copeland, N.G., Musgrove, E.A., Scarpa, A., Eshleman, J.R., Hudson, T.J., Sutherland, R.L., Wheeler, D.A., Pearson, J.V., McPherson, J.D., Gibbs, R.A., Grimmond, S.M., 2012. Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. Nature 491, 399–405. https://doi.org/10.1038/nature11547

Bollati-Fogolín, M., Forno, G., Nimtz, M., Conradt, H.S., Etcheverrigaray, M., Kratje, R., 2005. Temperature reduction in cultures of hGM-CSF-expressing CHO cells: effect on productivity and product quality. Biotechnol. Prog. 21, 17–21. https://doi.org/10.1021/bp049825t

Bonander, N., Darby, R.A., Grgic, L., Bora, N., Wen, J., Brogna, S., Poyner, D.R., O'Neill, M.A., Bill, R.M., 2009. Altering the ribosomal subunit ratio in yeast maximizes recombinant protein yield. Microb. Cell Factories 8, 10.

Boyle, J., Czito, B., Willett, C., Palta, M., 2015. Adjuvant radiation therapy for pancreatic cancer: a review of the old and the new. J. Gastrointest. Oncol. 6, 436–444. https://doi.org/10.3978/j.issn.2078-6891.2015.014

Brand, R.E., Lerch, M.M., Rubinstein, W.S., Neoptolemos, J.P., Whitcomb, D.C., Hruban, R.H., Brentnall, T.A., Lynch, H.T., Canto, M.I., 2007. Advances in counselling and surveillance of patients at risk for pancreatic cancer. Gut 56, 1460–1469. https://doi.org/10.1136/gut.2006.108456

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., Jemal, A., 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA. Cancer J. Clin. 68, 394–424. https://doi.org/10.3322/caac.21492

Brodsky, F.M., 1988. Monoclonal antibodies as magic bullets. Pharm. Res. 5, 1–9.

Burris, H.A., Moore, M.J., Andersen, J., Green, M.R., Rothenberg, M.L., Modiano, M.R., Cripps, M.C., Portenoy, R.K., Storniolo, A.M., Tarassoff, P., Nelson, R., Dorr, F.A., Stephens, C.D., Von Hoff, D.D., 1997. Improvements in survival and clinical benefit with gemcitabine as first-line therapy for patients with advanced pancreas cancer: a randomized trial. J. Clin. Oncol. 15, 2403–2413. https://doi.org/10.1200/JCO.1997.15.6.2403

Butler, M., Spearman, M., 2014. The choice of mammalian cell host and possibilities for glycosylation engineering. Curr. Opin. Biotechnol. 30, 107–112. https://doi.org/10.1016/j.copbio.2014.06.010

Cai, W., Yang, H., 2016. The structure and regulation of Cullin 2 based E3 ubiquitin ligases and their biological functions. Cell Div. 11, 7. https://doi.org/10.1186/s13008-016-0020-7

Canto, M.I., Goggins, M., Hruban, R.H., Petersen, G.M., Giardiello, F.M., Yeo, C., Fishman, E.K., Brune, K., Axilbund, J., Griffin, C., Ali, S., Richman, J., Jagannath, S., Kantsevoy, S.V., Kalloo, A.N., 2006. Screening for Early Pancreatic Neoplasia in High-Risk Individuals: A Prospective Controlled Study. Clin. Gastroenterol. Hepatol. 4, 766–781. https://doi.org/10.1016/j.cgh.2006.02.005

Chapman, P.B., Hauschild, A., Robert, C., Haanen, J.B., Ascierto, P., Larkin, J., Dummer, R., Garbe, C., Testori, A., Maio, M., Hogg, D., Lorigan, P., Lebbe, C., Jouary, T., Schadendorf, D., Ribas, A., O'Day, S.J., Sosman, J.A., Kirkwood, J.M., Eggermont, A.M.M., Dreno, B., Nolop, K., Li, J., Nelson, B., Hou, J., Lee, R.J., Flaherty, K.T., McArthur, G.A., 2011. Improved Survival with Vemurafenib in Melanoma with BRAF V600E Mutation. N. Engl. J. Med. 364, 2507–2516. https://doi.org/10.1056/NEJMoa1103782

Chelius, D., Bondarenko, P.V., 2002. Quantitative Profiling of Proteins in Complex Mixtures Using Liquid Chromatography and Mass Spectrometry. J. Proteome Res. 1, 317–323. https://doi.org/10.1021/pr025517j

Chiang, S.-F., Tsai, M.-H., Tang, R., Hsieh, L.-L., Chiang, J.-M., Yeh, C.-Y., Hsieh, P.-S., Tsai, W.-S., Liu, Y.-P., Liang, Y., Chen, J.-S., Yu, J.-S., 2014. Membrane Proteins as Potential Colon Cancer Biomarkers: Verification of 4 Candidates from a Secretome Dataset. Surg. Sci. 05, 418. https://doi.org/10.4236/ss.2014.510067

Ciliberto, D., Botta, C., Correale, P., Rossi, M., Caraglia, M., Tassone, P., Tagliaferri, P., 2013. Role of gemcitabine-based combination therapy in the management of advanced pancreatic cancer: a meta-analysis of randomised trials. Eur. J. Cancer 49, 593–603.

Clarke, C., Henry, M., Doolan, P., Kelly, S., Aherne, S., Sanchez, N., Kelly, P., Kinsella, P., Breen, L., Madden, S.F., 2012a. Integrated miRNA, mRNA and protein expression analysis reveals the role of post-transcriptional regulation in controlling CHO cell growth rate. BMC Genomics 13, 656.

Clarke, C., Henry, M., Doolan, P., Kelly, S., Aherne, S., Sanchez, N., Kelly, P., Kinsella, P., Breen, L., Madden, S.F., Zhang, L., Leonard, M., Clynes, M., Meleady, P., Barron, N., 2012b. Integrated miRNA, mRNA and protein expression analysis reveals the role of post-transcriptional regulation in controlling CHO cell growth rate. BMC Genomics 13, 656. https://doi.org/10.1186/1471-2164-13-656

Coleman, O., Henry, M., Clynes, M., Meleady, P., 2017. Filter-Aided Sample Preparation (FASP) for Improved Proteome Analysis of Recombinant Chinese Hamster Ovary Cells. Methods Mol. Biol. Clifton NJ 1603, 187–194. https://doi.org/10.1007/978-1-4939-6972-2_12

Coleman, O., Henry, M., McVey, G., Clynes, M., Moriarty, M., Meleady, P., 2016. Proteomic strategies in the search for novel pancreatic cancer biomarkers and drug targets: recent advances and clinical impact. Expert Rev. Proteomics 13, 383–394.

Collins, B.C., Hunter, C.L., Liu, Y., Schilling, B., Rosenberger, G., Bader, S.L., Chan, D.W., Gibson, B.W., Gingras, A.-C., Held, J.M., Hirayama-Kurogi, M., Hou, G., Krisp, C., Larsen, B., Lin, L., Liu, S., Molloy, M.P., Moritz, R.L., Ohtsuki, S., Schlapbach, R., Selevsek, N., Thomas, S.N., Tzeng, S.-C., Zhang, H., Aebersold, R., 2017. Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry. Nat. Commun. 8, 291. https://doi.org/10.1038/s41467-017-00249-5

Conroy, T., Desseigne, F., Ychou, M., Bouché, O., Guimbaud, R., Bécouarn, Y., Adenis, A., Raoul, J.-L., Gourgou-Bourgade, S., Fouchardière, C. de la, 2011. FOLFIRINOX versus gemcitabine for metastatic pancreatic cancer. N. Engl. J. Med. 364, 1817–1825.

Cooper, G.M., 2000. Structure of the Plasma Membrane. Cell Mol. Approach 2nd Ed.

Cost, G.J., Freyvert, Y., Vafiadis, A., Santiago, Y., Miller, J.C., Rebar, E., Collingwood, T.N., Snowden, A., Gregory, P.D., 2010. BAK and BAX deletion using zinc-finger nucleases yields apoptosis-resistant CHO cells. Biotechnol. Bioeng. 105, 330–340. https://doi.org/10.1002/bit.22541

Costello, A., Coleman, O., Lao, N.T., Henry, M., Meleady, P., Barron, N., Clynes, M., 2018. Depletion of endogenous miRNA-378-3p increases peak cell density of CHO DP12 cells and is correlated with elevated levels of ubiquitin carboxyl-terminal hydrolase 14. J. Biotechnol. 288, 30–40. https://doi.org/10.1016/j.jbiotec.2018.10.008

Cottrell, J.S., 2011. Protein identification using MS/MS data. J. Proteomics 74, 1842–1851.

Cox, J., Hein, M.Y., Luber, C.A., Paron, I., Nagaraj, N., Mann, M., 2014. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol. Cell. Proteomics MCP 13, 2513–2526. https://doi.org/10.1074/mcp.M113.031591

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

Dadehbeigi, N., Dickson, A.J., 2015. Chemical manipulation of the mTORC1 pathway in industrially relevant CHOK1 cells enhances production of therapeutic proteins. Biotechnol. J. 10, 1041–1050. https://doi.org/10.1002/biot.201500075

Damelin, M., Zhong, W., Myers, J., Sapra, P., 2015. Evolving Strategies for Target Selection for Antibody-Drug Conjugates. Pharm. Res. 32, 3494–3507. https://doi.org/10.1007/s11095-015-1624-3

Datta, P., Linhardt, R.J., Sharfstein, S.T., 2013. An 'omics approach towards CHO cell engineering. Biotechnol. Bioeng. 110, 1255–1271. https://doi.org/10.1002/bit.24841

Deeb, S.J., Cox, J., Schmidt-Supprian, M., Mann, M., 2014. N-linked Glycosylation Enrichment for In-depth Cell Surface Proteomics of Diffuse Large B-cell Lymphoma Subtypes. Mol. Cell. Proteomics 13, 240–251. https://doi.org/10.1074/mcp.M113.033977

Doolan, P., Meleady, P., Barron, N., Henry, M., Gallagher, R., Gammell, P., Melville, M., Sinacore, M., McCarthy, K., Leonard, M., Charlebois, T., Clynes, M., 2010. Microarray and proteomics expression profiling identifies several candidates, including the valosin-containing protein (VCP), involved in regulating high cellular growth rate in production CHO cell lines. Biotechnol. Bioeng. 106, 42–56. https://doi.org/10.1002/bit.22670

Dreesen, I.A.J., Fussenegger, M., 2011. Ectopic expression of human mTOR increases viability, robustness, cell size, proliferation, and antibody production of chinese hamster ovary cells. Biotechnol. Bioeng. 108, 853–866. https://doi.org/10.1002/bit.22990

Druz, A., Son, Y., Betenbaugh, M., Shiloach, J., 2013. Stable inhibition of mmu-miR-466h-5p improves apoptosis resistance and protein production in CHO cells. Metab. Eng. 16, 87–94. https://doi.org/10.1016/j.ymben.2012.12.004

D.Sc, F.W.A.M.A., 1919. LXXIV. A positive ray spectrograph. Lond. Edinb. Dublin Philos. Mag. J. Sci. 38, 707–714. https://doi.org/10.1080/14786441208636004

Early detection: a long road ahead, 2018. . Nat. Rev. Cancer 18, 401. https://doi.org/10.1038/s41568-018-0021-8

Ecker, D.M., Jones, S.D., Levine, H.L., 2015. The therapeutic monoclonal antibody market. mAbs 7, 9–14. https://doi.org/10.4161/19420862.2015.989042

Edwards, A.M., Isserlin, R., Bader, G.D., Frye, S.V., Willson, T.M., Yu, F.H., 2011. Too many roads not taken. Nature 470, 163–165. https://doi.org/10.1038/470163a

Egawa, S., Takeda, K., Fukuyama, S., Motoi, F., Sunamura, M., Matsuno, S., 2004. Clinicopathological aspects of small pancreatic cancer. Pancreas 28, 235–240.

Eng, J.K., McCormack, A.L., Yates, J.R., 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J. Am. Soc. Mass Spectrom. 5, 976–989. https://doi.org/10.1016/1044-0305(94)80016-2

Erde, J., Loo, R.R.O., Loo, J.A., 2014. Enhanced FASP (eFASP) to Increase Proteome Coverage and Sample Recovery for Quantitative Proteomic Experiments. J. Proteome Res. 13, 1885–1895. https://doi.org/10.1021/pr4010019

Eser, S., Messer, M., Eser, P., Werder, A. von, Seidler, B., Bajbouj, M., Vogelmann, R., Meining, A., Burstin, J. von, Algül, H., Pagel, P., Schnieke, A.E., Esposito, I., Schmid, R.M., Schneider, G., Saur, D., 2011. In vivo diagnosis of murine

pancreatic intraepithelial neoplasia and early-stage pancreatic cancer by molecular imaging. Proc. Natl. Acad. Sci. 108, 9945–9950. https://doi.org/10.1073/pnas.1100890108

Estes, S., Melville, M., 2014. Mammalian Cell Line Developments in Speed and Efficiency, in: Zhou, W., Kantardjieff, A. (Eds.), Mammalian Cell Cultures for Biologics Manufacturing, Advances in Biochemical Engineering/Biotechnology. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 11–33. https://doi.org/10.1007/10_2013_260

Falini, B., Pileri, S., Pizzolo, G., Durkop, H., Flenghi, L., Stirpe, F., Martelli, M.F., Stein, H., 1995. CD30 (Ki-1) molecule: a new cytokine receptor of the tumor necrosis factor receptor superfamily as a tool for diagnosis and immunotherapy. Blood 85, 1–14.

Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., Whitehouse, C.M., 1989. Electrospray ionization for mass spectrometry of large biomolecules. Science 246, 64–71.

Ferlay, J., Partensky, C., Bray, F., 2016. More deaths from pancreatic cancer than breast cancer in the EU by 2017. Acta Oncol. 55, 1158–1160. https://doi.org/10.1080/0284186X.2016.1197419

Finkle, B., 1988. New medicines from industry. J. Chem. Technol. Biotechnol. 43, 313–327. https://doi.org/10.1002/jctb.280430411

Fischer, S., Buck, T., Wagner, A., Ehrhart, C., Giancaterino, J., Mang, S., Schad, M., Mathias, S., Aschrafi, A., Handrick, R., Otte, K., 2014. A functional high-content miRNA screen identifies miR-30 family to boost recombinant protein production in CHO cells. Biotechnol. J. 9, 1279–1292. https://doi.org/10.1002/biot.201400306

Fischer, S., Handrick, R., Otte, K., 2015. The art of CHO cell engineering: A comprehensive retrospect and future perspectives. Biotechnol. Adv. 33, 1878–1896. https://doi.org/10.1016/j.biotechadv.2015.10.015

Fischer, S., Marquart, K.F., Pieper, L.A., Fieder, J., Gamer, M., Gorr, I., Schulz, P., Bradl, H., 2017. miRNA engineering of CHO cells facilitates production of difficult-to-express proteins and increases success in cell line development. Biotechnol. Bioeng. 114, 1495–1510. https://doi.org/10.1002/bit.26280

Florin, L., Pegel, A., Becker, E., Hausser, A., Olayioye, M.A., Kaufmann, H., 2009. Heterologous expression of the lipid transfer protein CERT increases therapeutic protein productivity of mammalian cells. J. Biotechnol. 141, 84–90. https://doi.org/10.1016/j.jbiotec.2009.02.014

Foltz Ian N., Karow Margaret, Wasserman Scott M., 2013. Evolution and Emergence of Therapeutic Monoclonal Antibodies. Circulation 127, 2222–2230. https://doi.org/10.1161/CIRCULATIONAHA.113.002033

Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., Mering, C.V., 2012. STRING v9. 1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 41, D808–D815.

Friedman, R.C., Farh, K.K.-H., Burge, C.B., Bartel, D.P., 2009. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. 19, 92–105. https://doi.org/10.1101/gr.082701.108

Gammell, P., Barron, N., Kumar, N., Clynes, M., 2007. Initial identification of low temperature and culture stage induction of miRNA expression in suspension CHO-K1 cells. J. Biotechnol. 130, 213–218.

Gamper, N., Stockand, J.D., Shapiro, M.S., 2005. The use of Chinese hamster ovary (CHO) cells in the study of ion channels. J. Pharmacol. Toxicol. Methods, Electrophysiological Methods in Neuropharmacology 51, 177–185. https://doi.org/10.1016/j.vascn.2004.08.008

Gautreau, A., Poullet, P., Louvard, D., Arpin, M., 1999. Ezrin, a plasma membrane-microfilament linker, signals cell survival through the phosphatidylinositol 3-kinase/Akt pathway. Proc. Natl. Acad. Sci. U. S. A. 96, 7300–7305.

Ghosh, D., Funk, C.C., Caballero, J., Shah, N., Rouleau, K., Earls, J.C., Soroceanu, L., Foltz, G., Cobbs, C.S., Price, N.D., Hood, L., 2017. A Cell-Surface Membrane Protein Signature for Glioblastoma. Cell Syst. 4, 516-529.e7. https://doi.org/10.1016/j.cels.2017.03.004

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

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

Giza, D.E., Vasilescu, C., Calin, G.A., 2014. Key principles of miRNA involvement in human diseases. Discoveries 2, e34. https://doi.org/10.15190/d.2014.26

Glatter, T., Ludwig, C., Ahrné, E., Aebersold, R., Heck, A.J.R., Schmidt, A., 2012. Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin proteolysis over trypsin digestion. J. Proteome Res. 11, 5145–5156. https://doi.org/10.1021/pr300273g

Goldstein, D., El-Maraghi, R.H., Hammel, P., Heinemann, V., Kunzmann, V., Sastre, J., Scheithauer, W., Siena, S., Tabernero, J., Teixeira, L., Tortora, G., Van Laethem, J.-L., Young, R., Penenberg, D.N., Lu, B., Romano, A., Von Hoff, D.D., 2015. nab-Paclitaxel plus gemcitabine for metastatic pancreatic cancer: long-term survival from a phase III trial. J. Natl. Cancer Inst. 107. https://doi.org/10.1093/jnci/dju413

Good, D.M., Wirtala, M., McAlister, G.C., Coon, J.J., 2007. Performance Characteristics of Electron Transfer Dissociation Mass Spectrometry. Mol. Cell. Proteomics 6, 1942–1951. https://doi.org/10.1074/mcp.M700073-MCP200

Goonetilleke, K.S., Siriwardena, A.K., 2007. Systematic review of carbohydrate antigen (CA 19-9) as a biochemical marker in the diagnosis of pancreatic cancer. Eur. J. Surg. Oncol. EJSO 33, 266–270.

Gourgou-Bourgade, S., Bascoul-Mollevi, C., Desseigne, F., Ychou, M., Bouché, O., Guimbaud, R., Bécouarn, Y., Adenis, A., Raoul, J.-L., Boige, V., Bérille, J., Conroy, T., 2013. Impact of FOLFIRINOX compared with gemcitabine on quality of life in patients with metastatic pancreatic cancer: results from the

PRODIGE 4/ACCORD 11 randomized trial. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. 31, 23–29. https://doi.org/10.1200/JCO.2012.44.4869

Grav, L.M., Lee, J.S., Gerling, S., Kallehauge, T.B., Hansen, A.H., Kol, S., Lee, G.M., Pedersen, L.E., Kildegaard, H.F., 2015. One-step generation of triple knockout CHO cell lines using CRISPR/Cas9 and fluorescent enrichment. Biotechnol. J. 10, 1446–1456. https://doi.org/10.1002/biot.201500027

Grimson, A., Farh, K.K.-H., Johnston, W.K., Garrett-Engele, P., Lim, L.P., Bartel, D.P., 2007. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol. Cell 27, 91–105.

Gustafsson, L., Pontén, J., Zack, M., Adami, H.O., 1997. International incidence rates of invasive cervical cancer after introduction of cytological screening. Cancer Causes Control CCC 8, 755–763.

Hackl, M., Borth, N., Grillari, J., 2012. miRNAs – pathway engineering of CHO cell factories that avoids translational burdening. Trends Biotechnol. 30, 405–406. https://doi.org/10.1016/j.tibtech.2012.05.002

Hackl, M., Jakobi, T., Blom, J., Doppmeier, D., Brinkrolf, K., Szczepanowski, R., Bernhart, S.H., Siederdissen, C.H. zu, Bort, J.A.H., Wieser, M., 2011. Next-generation sequencing of the Chinese hamster ovary microRNA transcriptome: Identification, annotation and profiling of microRNAs as targets for cellular engineering. J. Biotechnol. 153, 62–75.

Hammond, S., Swanberg, J.C., Kaplarevic, M., Lee, K.H., 2011. Genomic sequencing and analysis of a Chinese hamster ovary cell line using Illumina sequencing technology. BMC Genomics 12, 67. https://doi.org/10.1186/1471-2164-12-67

Hammond, S.M., 2015. An overview of microRNAs. Adv. Drug Deliv. Rev. 87, 3–14.

Hanahan, D., Weinberg, R.A., 2000. The Hallmarks of Cancer. Cell 100, 57–70. https://doi.org/10.1016/S0092-8674(00)81683-9

HAYASHIDA, K., BARTLETT, A.H., CHEN, Y., PARK, P.W., 2010. Molecular and Cellular Mechanisms of Ectodomain Shedding. Anat. Rec. Hoboken NJ 2007 293, 925–937. https://doi.org/10.1002/ar.20757

Heffner, K.M., Hizal, D.B., Yerganian, G.S., Kumar, A., Can, Ö., O'Meally, R., Cole, R., Chaerkady, R., Wu, H., Bowen, M.A., Betenbaugh, M.J., 2017. Lessons from the Hamster: Cricetulus griseus Tissue and CHO Cell Line Proteome Comparison. J. Proteome Res. 16, 3672–3687. https://doi.org/10.1021/acs.jproteome.7b00382

Henry, M., Power, M., Kaushik, P., Coleman, O., Clynes, M., Meleady, P., 2017. Differential Phosphoproteomic Analysis of Recombinant Chinese Hamster Ovary Cells Following Temperature Shift. J. Proteome Res. 16, 2339–2358. https://doi.org/10.1021/acs.jproteome.6b00868

Hoelder, S., Clarke, P.A., Workman, P., 2012. Discovery of small molecule cancer drugs: Successes, challenges and opportunities. Mol. Oncol., Personalized cancer medicine 6, 155–176. https://doi.org/10.1016/j.molonc.2012.02.004

Hu, Q., Noll, R.J., Li, H., Makarov, A., Hardman, M., Graham Cooks, R., 2005. The Orbitrap: a new mass spectrometer. J. Mass Spectrom. JMS 40, 430–443. https://doi.org/10.1002/jms.856

Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4, 44–57. https://doi.org/10.1038/nprot.2008.211

Huang, D.W., Sherman, B.T., Tan, Q., Collins, J.R., Alvord, W.G., Roayaei, J., Stephens, R., Baseler, M.W., Lane, H.C., Lempicki, R.A., 2007. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol. 8, R183. https://doi.org/10.1186/gb-2007-8-9-r183

Huang, E.P., Marquis, C.P., Gray, P.P., 2007. Development of Super-CHO protein-free medium based on a statistical design. J. Chem. Technol. Biotechnol. Int. Res. Process Environ. Clean Technol. 82, 431–441.

Huang, E.P., Marquis, C.P., Gray, P.P., 2004. Process development for a recombinant Chinese hamster ovary (CHO) cell line utilizing a metal induced and amplified metallothionein expression system. Biotechnol. Bioeng. 88, 437–450.

Huang, T.-C., Pinto, S.M., Pandey, A., 2013. Proteomics for understanding miRNA biology. Proteomics 13, 558–567.

Jaluria, P., Betenbaugh, M., Konstantopoulos, K., Shiloach, J., 2007. Enhancement of cell proliferation in various mammalian cell lines by gene insertion of a cyclin-dependent kinase homolog. BMC Biotechnol. 7, 71. https://doi.org/10.1186/1472-6750-7-71

Jansson, M.D., Lund, A.H., 2012. MicroRNA and cancer. Mol. Oncol., Cancer epigenetics 6, 590–610. https://doi.org/10.1016/j.molonc.2012.09.006

Jayapal, K.P., Wlaschin, K.F., Hu, W.S., Yap, M.G.S., 2007. Recombinant protein therapeutics from CHO Cells - 20 years and counting. Chem. Eng. Prog. 103, 40–47.

Jedrychowski, M.P., Huttlin, E.L., Haas, W., Sowa, M.E., Rad, R., Gygi, S.P., 2011. Evaluation of HCD- and CID-type Fragmentation Within Their Respective Detection Platforms For Murine Phosphoproteomics. Mol. Cell. Proteomics MCP 10. https://doi.org/10.1074/mcp.M111.009910

Jones, S., Zhang, X., Parsons, D.W., Lin, J.C.-H., Leary, R.J., Angenendt, P., Mankoo, P., Carter, H., Kamiyama, H., Jimeno, A., Hong, S.-M., Fu, B., Lin, M.-T., Calhoun, E.S., Kamiyama, M., Walter, K., Nikolskaya, T., Nikolsky, Y., Hartigan, J., Smith, D.R., Hidalgo, M., Leach, S.D., Klein, A.P., Jaffee, E.M., Goggins, M., Maitra, A., Iacobuzio-Donahue, C., Eshleman, J.R., Kern, S.E., Hruban, R.H., Karchin, R., Papadopoulos, N., Parmigiani, G., Vogelstein, B., Velculescu, V.E., Kinzler, K.W., 2008. Core Signaling Pathways in Human Pancreatic Cancers Revealed by Global Genomic Analyses. Science 321, 1801–1806. https://doi.org/10.1126/science.1164368

Juraschek, R., Dülcks, T., Karas, M., 1999. Nanoelectrospray--more than just a minimized-flow electrospray ionization source. J. Am. Soc. Mass Spectrom. 10, 300–308. https://doi.org/10.1016/S1044-0305(98)00157-3

Käll, L., Canterbury, J.D., Weston, J., Noble, W.S., MacCoss, M.J., 2007. Semi-supervised learning for peptide identification from shotgun proteomics datasets. Nat. Methods 4, 923–925. https://doi.org/10.1038/nmeth1113

Kallehauge, T.B., Li, S., Pedersen, L.E., Ha, T.K., Ley, D., Andersen, M.R., Kildegaard, H.F., Lee, G.M., Lewis, N.E., 2017. Ribosome profiling-guided depletion of an mRNA increases cell growth rate and protein secretion. Sci. Rep. 7, 40388.

Kanehisa, M., Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28, 27–30.

Karas, M., Bahr, U., Dülcks, T., 2000. Nano-electrospray ionization mass spectrometry: addressing analytical problems beyond routine. Fresenius J. Anal. Chem. 366, 669–676. https://doi.org/10.1007/s002160051561

Kaushik, P., Henry, M., Clynes, M., Meleady, P., 2018. The Expression Pattern of the Phosphoproteome Is Significantly Changed During the Growth Phases of Recombinant CHO Cell Culture. Biotechnol. J. 13, e1700221. https://doi.org/10.1002/biot.201700221

Kiezun, A., Artzi, S., Modai, S., Volk, N., Isakov, O., Shomron, N., 2012. miRviewer: a multispecies microRNA homologous viewer. BMC Res. Notes 5, 92.

Kim, J.Y., Kim, Y.-G., Han, Y.K., Choi, H.S., Kim, Y.H., Lee, G.M., 2011. Proteomic understanding of intracellular responses of recombinant Chinese hamster ovary cells cultivated in serum-free medium supplemented with hydrolysates. Appl. Microbiol. Biotechnol. 89, 1917–1928. https://doi.org/10.1007/s00253-011-3106-9

King, R., Bonfiglio, R., Fernandez-Metzler, C., Miller-Stein, C., Olah, T., 2000. Mechanistic investigation of ionization suppression in electrospray ionization. J. Am. Soc. Mass Spectrom. 11, 942–950. https://doi.org/10.1016/S1044-0305(00)00163-X

Kishishita, S., Katayama, S., Kodaira, K., Takagi, Y., Matsuda, H., Okamoto, H., Takuma, S., Hirashima, C., Aoyagi, H., 2015. Optimization of chemically defined feed media for monoclonal antibody production in Chinese hamster ovary cells. J. Biosci. Bioeng. 120, 78–84.

Köhler, A., Hurt, E., 2007. Exporting RNA from the nucleus to the cytoplasm. Nat. Rev. Mol. Cell Biol. 8, 761–773. https://doi.org/10.1038/nrm2255

Koopmann, J., Fedarko, N.S., Jain, A., Maitra, A., Iacobuzio-Donahue, C., Rahman, A., Hruban, R.H., Yeo, C.J., Goggins, M., 2004. Evaluation of osteopontin as biomarker for pancreatic adenocarcinoma. Cancer Epidemiol. Biomark. Prev. Publ. Am. Assoc. Cancer Res. Cosponsored Am. Soc. Prev. Oncol. 13, 487–491.

Kota, J., Hancock, J., Kwon, J., Korc, M., 2017. Pancreatic cancer: Stroma and its current and emerging targeted therapies. Cancer Lett. 391, 38–49. https://doi.org/10.1016/j.canlet.2016.12.035

Krämer, A., Green, J., Pollard, J., Tugendreich, S., 2014. Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics 30, 523–530. https://doi.org/10.1093/bioinformatics/btt703

Kuhlmann, K.F.D., van Till, J.W.O., Boermeester, M.A., de Reuver, P.R., Tzvetanova, I.D., Offerhaus, G.J.A., ten Kate, F.J.W., Busch, O.R.C., van Gulik, T.M., Gouma, D.J., Crawford, H.C., 2007. Evaluation of Matrix Metalloproteinase 7 in Plasma and Pancreatic Juice as a Biomarker for Pancreatic Cancer. Cancer Epidemiol. Biomark. Prev. Publ. Am. Assoc. Cancer Res. Cosponsored Am. Soc. Prev. Oncol. 16, 886–891. https://doi.org/10.1158/1055-9965.EPI-06-0779

Kuystermans, D., Krampe, B., Swiderek, H., Al-Rubeai, M., 2007. Using cell engineering and omic tools for the improvement of cell culture processes. Cytotechnology 53, 3–22. https://doi.org/10.1007/s10616-007-9055-6

Kwon, R.-J., Kim, S.K., Lee, S.-I., Hwang, S.-J., Lee, G.M., Kim, J.-S., Seol, W., 2006. Artificial transcription factors increase production of recombinant antibodies in Chinese hamster ovary cells. Biotechnol. Lett. 28, 9–15. https://doi.org/10.1007/s10529-005-4680-7

Landberg, G., Tan, E.M., 1994. Characterization of a DNA-binding nuclear autoantigen mainly associated with S phase and G2 cells. Exp. Cell Res. 212, 255–261.

Laskey, R.A., Fairman, M.P., Blow, J.J., 1989. S phase of the cell cycle. Science 246, 609–614.

Le Fourn, V., Girod, P.-A., Buceta, M., Regamey, A., Mermod, N., 2014. CHO cell engineering to prevent polypeptide aggregation and improve therapeutic protein secretion. Metab. Eng. 21, 91–102. https://doi.org/10.1016/j.ymben.2012.12.003

Le, H., Chen, C., Goudar, C.T., 2015. An evaluation of public genomic references for mapping RNA-Seq data from Chinese hamster ovary cells. Biotechnol. Bioeng. 112, 2412–2416. https://doi.org/10.1002/bit.25649

Lee, C.-T., Risom, T., Strauss, W.M., 2007. Evolutionary conservation of microRNA regulatory circuits: an examination of microRNA gene complexity and conserved microRNA-target interactions through metazoan phylogeny. DNA Cell Biol. 26, 209–218.

Lee, E.U., Roth, J., Paulson, J.C., 1989. Alteration of terminal glycosylation sequences on N-linked oligosaccharides of Chinese hamster ovary cells by expression of beta-galactoside alpha 2,6-sialyltransferase. J. Biol. Chem. 264, 13848–13855.

Lewis, B.P., Burge, C.B., Bartel, D.P., 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120, 15–20.

Lewis, N.E., Liu, X., Li, Y., Nagarajan, H., Yerganian, G., O'Brien, E., Bordbar, A., Roth, A.M., Rosenbloom, J., Bian, C., Xie, M., Chen, W., Li, N., Baycin-Hizal, D., Latif, H., Forster, J., Betenbaugh, M.J., Famili, I., Xu, X., Wang, J., Palsson, B.O., 2013. Genomic landscapes of Chinese hamster ovary cell lines as revealed by the *Cricetulus griseus* draft genome. Nat. Biotechnol. 31, 759–765. https://doi.org/10.1038/nbt.2624

Lewis Phillips, G.D., Li, G., Dugger, D.L., Crocker, L.M., Parsons, K.L., Mai, E., Blättler, W.A., Lambert, J.M., Chari, R.V.J., Lutz, R.J., Wong, W.L.T., Jacobson, F.S., Koeppen, H., Schwall, R.H., Kenkare-Mitra, S.R., Spencer, S.D., Sliwkowski, M.X., 2008. Targeting HER2-positive breast cancer with trastuzumab-DM1, an antibody-cytotoxic drug conjugate. Cancer Res. 68, 9280–9290. https://doi.org/10.1158/0008-5472.CAN-08-1776

Li, C., Xiong, Q., Zhang, J., Ge, F., Bi, L.-J., 2012. Quantitative proteomic strategies for the identification of microRNA targets. Expert Rev. Proteomics 9, 549–559. https://doi.org/10.1586/epr.12.49

Lin, N., Mascarenhas, J., Sealover, N.R., George, H.J., Brooks, J., Kayser, K.J., Gau, B., Yasa, I., Azadi, P., Archer-Hartmann, S., 2015. Chinese hamster ovary (CHO) host cell engineering to increase sialylation of recombinant therapeutic proteins

by modulating sialyltransferase expression. Biotechnol. Prog. 31, 334–346. https://doi.org/10.1002/btpr.2038

Liu, H., Yue, D., Chen, Y., Gao, S.-J., Huang, Y., 2010. Improving performance of mammalian microRNA target prediction. BMC Bioinformatics 11, 476. https://doi.org/10.1186/1471-2105-11-476

Liz, J., Esteller, M., 2016. lncRNAs and microRNAs with a role in cancer development. Biochim. Biophys. Acta BBA - Gene Regul. Mech., SI: Clues to long noncoding RNA taxonomy 1859, 169–176. https://doi.org/10.1016/j.bbagrm.2015.06.015

Loganantharaj, R., Randall, T.A., 2017. The Limitations of Existing Approaches in Improving MicroRNA Target Prediction Accuracy. Methods Mol. Biol. Clifton NJ 1617, 133–158. https://doi.org/10.1007/978-1-4939-7046-9_10

Ma, C., Qi, Y., Shao, L., Liu, M., Li, X., Tang, H., 2013. Downregulation of miR-7 upregulates Cullin 5 (CUL5) to facilitate G1/S transition in human hepatocellular carcinoma cells. IUBMB Life 65, 1026–1034.

Majors, B.S., Betenbaugh, M.J., Pederson, N.E., Chiang, G.G., 2009. Mcl-1 overexpression leads to higher viabilities and increased production of humanized monoclonal antibody in Chinese hamster ovary cells. Biotechnol. Prog. 25, 1161–1168. https://doi.org/10.1002/btpr.192

Makarov, A., 2000. Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. Anal. Chem. 72, 1156–1162. https://doi.org/10.1021/ac991131p

Makarov, A., Denisov, E., Kholomeev, A., Balschun, W., Lange, O., Strupat, K., Horning, S., 2006. Performance Evaluation of a Hybrid Linear Ion Trap/Orbitrap Mass Spectrometer. Anal. Chem. 78, 2113–2120. https://doi.org/10.1021/ac0518811

Mastrangelo, A.J., Hardwick, J.M., Bex, F., Betenbaugh, M.J., 2000. Part I. Bcl-2 and Bcl-x(L) limit apoptosis upon infection with alphavirus vectors. Biotechnol. Bioeng. 67, 544–554.

Meents, H., Enenkel, B., Eppenberger, H.M., Werner, R.G., Fussenegger, M., 2002. Impact of coexpression and coamplification of sICAM and antiapoptosis determinants bcl-2/bcl-x(L) on productivity, cell survival, and mitochondria number in CHO-DG44 grown in suspension and serum-free media. Biotechnol. Bioeng. 80, 706–716. https://doi.org/10.1002/bit.10449

Mei, L., Du, W., Ma, W.W., 2016. Targeting stromal microenvironment in pancreatic ductal adenocarcinoma: controversies and promises. J. Gastrointest. Oncol. 7, 487–494. https://doi.org/10.21037/jgo.2016.03.03

Meleady, P., Gallagher, M., Clarke, C., Henry, M., Sanchez, N., Barron, N., Clynes, M., 2012a. Impact of miR-7 over-expression on the proteome of Chinese hamster ovary cells. J. Biotechnol. 160, 251–262.

Meleady, P., Hoffrogge, R., Henry, M., Rupp, O., Bort, J.H., Clarke, C., Brinkrolf, K., Kelly, S., Müller, B., Doolan, P., Hackl, M., Beckmann, T.F., Noll, T., Grillari, J., Barron, N., Pühler, A., Clynes, M., Borth, N., 2012b. Utilization and evaluation of CHO-specific sequence databases for mass spectrometry based proteomics. Biotechnol. Bioeng. 109, 1386–1394. https://doi.org/10.1002/bit.24476

Mering, C. von, Huynen, M., Jaeggi, D., Schmidt, S., Bork, P., Snel, B., 2003. STRING: a database of predicted functional associations between proteins. Nucleic Acids Res. 31, 258–261.

Michalski, A., Damoc, E., Hauschild, J.-P., Lange, O., Wieghaus, A., Makarov, A., Nagaraj, N., Cox, J., Mann, M., Horning, S., 2011. Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. Mol. Cell. Proteomics MCP 10, M111.011015. https://doi.org/10.1074/mcp.M111.011015

Mikesh, L.M., Ueberheide, B., Chi, A., Coon, J.J., Syka, J.E.P., Shabanowitz, J., Hunt, D.F., 2006. The utility of ETD mass spectrometry in proteomic analysis. Biochim. Biophys. Acta 1764, 1811–1822. https://doi.org/10.1016/j.bbapap.2006.10.003

Moritz, B., Woltering, L., Becker, P.B., Göpfert, U., 2016. High levels of histone H3 acetylation at the CMV promoter are predictive of stable expression in Chinese hamster ovary cells. Biotechnol. Prog. 32, 776–786. https://doi.org/10.1002/btpr.2271

Müller, D., Katinger, H., Grillari, J., 2008. MicroRNAs as targets for engineering of CHO cell factories. Trends Biotechnol. 26, 359–365.

Nasiri, H., Valedkarimi, Z., Aghebati-Maleki, L., Majidi, J., 2018. Antibody-drug conjugates: Promising and efficient tools for targeted cancer therapy. J. Cell. Physiol. 233, 6441–6457. https://doi.org/10.1002/jcp.26435

O'Brien, S.G., Guilhot, F., Larson, R.A., Gathmann, I., Baccarani, M., Cervantes, F., Cornelissen, J.J., Fischer, T., Hochhaus, A., Hughes, T., Lechner, K., Nielsen, J.L., Rousselot, P., Reiffers, J., Saglio, G., Shepherd, J., Simonsson, B., Gratwohl, A., Goldman, J.M., Kantarjian, H., Taylor, K., Verhoef, G., Bolton, A.E., Capdeville, R., Druker, B.J., 2003. Imatinib Compared with Interferon and Low-Dose Cytarabine for Newly Diagnosed Chronic-Phase Chronic Myeloid Leukemia. N. Engl. J. Med. 348, 994–1004. https://doi.org/10.1056/NEJMoa022457

Ocak, S., Friedman, D.B., Chen, H., Ausborn, J.A., Hassanein, M., Detry, B., Weynand, B., Aboubakar, F., Pilette, C., Sibille, Y., Massion, P.P., 2014. Discovery of New Membrane-Associated Proteins Overexpressed in Small-Cell Lung Cancer. J. Thorac. Oncol. Off. Publ. Int. Assoc. Study Lung Cancer 9, 324–336. https://doi.org/10.1097/JTO.0000000000000090

Ohya, T., Hayashi, T., Kiyama, E., Nishii, H., Miki, H., Kobayashi, K., Honda, K., Omasa, T., Ohtake, H., 2008. Improved production of recombinant human antithrombin III in Chinese hamster ovary cells by ATF4 overexpression. Biotechnol. Bioeng. 100, 317–324. https://doi.org/10.1002/bit.21758

Olive, K.P., Jacobetz, M.A., Davidson, C.J., Gopinathan, A., McIntyre, D., Honess, D., Madhu, B., Goldgraben, M.A., Caldwell, M.E., Allard, D., Frese, K.K., DeNicola, G., Feig, C., Combs, C., Winter, S.P., Ireland-Zecchini, H., Reichelt, S., Howat, W.J., Chang, A., Dhara, M., Wang, L., Rückert, F., Grützmann, R., Pilarsky, C., Izeradjene, K., Hingorani, S.R., Huang, P., Davies, S.E., Plunkett, W., Egorin, M., Hruban, R.H., Whitebread, N., McGovern, K., Adams, J., Iacobuzio-Donahue, C., Griffiths, J., Tuveson, D.A., 2009. Inhibition of Hedgehog Signaling Enhances Delivery of Chemotherapy in a Mouse Model of

Pancreatic Cancer. Science 324, 1457–1461. https://doi.org/10.1126/science.1171362
Omasa, T., Takami, T., Ohya, T., Kiyama, E., Hayashi, T., Nishii, H., Miki, H., Kobayashi, K., Honda, K., Ohtake, H., 2008. Overexpression of GADD34 enhances production of recombinant human antithrombin III in Chinese hamster ovary cells. J. Biosci. Bioeng. 106, 568–573. https://doi.org/10.1263/jbb.106.568
Pappa, K.I., Christou, P., Xholi, A., Mermelekas, G., Kontostathi, G., Lygirou, V., Makridakis, M., Zoidakis, J., Anagnou, N.P., 2018. Membrane proteomics of cervical cancer cell lines reveal insights on the process of cervical carcinogenesis. Int. J. Oncol. 53, 2111–2122. https://doi.org/10.3892/ijo.2018.4518
Paredes, V., Park, J.S., Jeong, Y., Yoon, J., Baek, K., 2013. Unstable expression of recombinant antibody during long-term culture of CHO cells is accompanied by histone H3 hypoacetylation. Biotechnol. Lett. 35, 987–993. https://doi.org/10.1007/s10529-013-1168-8
Parker, C.E., Warren, M.R., Mocanu, V., 2010. Mass Spectrometry for Proteomics, in: Alzate, O. (Ed.), Neuroproteomics, Frontiers in Neuroscience. CRC Press/Taylor & Francis, Boca Raton (FL).
Peng, R.-W., Abellan, E., Fussenegger, M., 2011. Differential effect of exocytic SNAREs on the production of recombinant proteins in mammalian cells. Biotechnol. Bioeng. 108, 611–620. https://doi.org/10.1002/bit.22986
Peng, R.-W., Fussenegger, M., 2009. Molecular engineering of exocytic vesicle traffic enhances the productivity of Chinese hamster ovary cells. Biotechnol. Bioeng. 102, 1170–1181. https://doi.org/10.1002/bit.22141
Perkins, D.N., Pappin, D.J., Creasy, D.M., Cottrell, J.S., 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20, 3551–3567. https://doi.org/10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2
Pieper, L.A., Strotbek, M., Wenger, T., Gamer, M., Olayioye, M.A., Hausser, A., 2017. Secretory pathway optimization of CHO producer cells by co-engineering of the mitosRNA-1978 target genes CerS2 and Tbc1D20. Metab. Eng. 40, 69–79.
Pillai, R.S., Bhattacharyya, S.N., Artus, C.G., Zoller, T., Cougot, N., Basyuk, E., Bertrand, E., Filipowicz, W., 2005. Inhibition of Translational Initiation by Let-7 MicroRNA in Human Cells. Science 309, 1573–1576. https://doi.org/10.1126/science.1115079
Pinzón, N., Li, B., Martinez, L., Sergeeva, A., Presumey, J., Apparailly, F., Seitz, H., 2017. microRNA target prediction programs predict many false positives. Genome Res. 27, 234–245. https://doi.org/10.1101/gr.205146.116
Poruk, K.E., Firpo, M.A., Adler, D.G., Mulvihill, S.J., 2013. Screening for Pancreatic Cancer: Why, How, and Who? Ann. Surg. 257, 17–26. https://doi.org/10.1097/SLA.0b013e31825ffbfb
Prati, E.G.P., Matasci, M., Suter, T.B., Dinter, A., Sburlati, A.R., Bailey, J.E., 2002. Engineering of coordinated up- and down-regulation of two glycosyltransferases of the O-glycosylation pathway in Chinese hamster ovary (CHO) cells. Biotechnol. Bioeng. 79, 580–585. https://doi.org/10.1002/bit.10442

Puck, T.T., Cieciura, S.J., Robinson, A., 1958. Genetics of Somatic Mammalian Cells: Iii. Long-Term Cultivation of Euploid Cells from Human and Animal Subjects. J. Exp. Med. 108, 945–956. https://doi.org/10.1084/jem.108.6.945

Riffo-Campos, Á.L., Riquelme, I., Brebi-Mieville, P., 2016. Tools for Sequence-Based miRNA Target Prediction: What to Choose? Int. J. Mol. Sci. 17. https://doi.org/10.3390/ijms17121987

Roepstorff, P., Fohlman, J., 1984. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. Biomed. Mass Spectrom. 11, 601. https://doi.org/10.1002/bms.1200111109

Root, A., Allen, P., Tempst, P., Yu, K., 2018. Protein Biomarkers for Early Detection of Pancreatic Ductal Adenocarcinoma: Progress and Challenges. Cancers 10. https://doi.org/10.3390/cancers10030067

Rubakhin, S.S., Sweedler, J.V., 2010. A mass spectrometry primer for mass spectrometry imaging. Methods Mol. Biol. Clifton NJ 656, 21–49. https://doi.org/10.1007/978-1-60761-746-4_2

Rupp, O., MacDonald, M.L., Li, S., Dhiman, H., Polson, S., Griep, S., Heffner, K., Hernandez, I., Brinkrolf, K., Jadhav, V., Samoudi, M., Hao, H., Kingham, B., Goesmann, A., Betenbaugh, M.J., Lewis, N.E., Borth, N., Lee, K.H., 2018. A reference genome of the Chinese hamster based on a hybrid assembly strategy. Biotechnol. Bioeng. 115, 2087–2100. https://doi.org/10.1002/bit.26722

Saha, S., Shan, Y., Mesner, L.D., Hamlin, J.L., 2004. The promoter of the Chinese hamster ovary dihydrofolate reductase gene regulates the activity of the local origin and helps define its boundaries. Genes Dev. 18, 397–410. https://doi.org/10.1101/gad.1171404

Sahdev, S., Khattar, S.K., Saini, K.S., 2008. Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. Mol. Cell. Biochem. 307, 249–264. https://doi.org/10.1007/s11010-007-9603-6

Sanchez, N., Gallagher, M., Lao, N., Gallagher, C., Clarke, C., Doolan, P., Aherne, S., Blanco, A., Meleady, P., Clynes, M., Barron, N., 2013. MiR-7 Triggers Cell Cycle Arrest at the G1/S Transition by Targeting Multiple Genes Including Skp2 and Psme3. PLOS ONE 8, e65671. https://doi.org/10.1371/journal.pone.0065671

Sanchez, N., Kelly, P., Gallagher, C., Lao, N.T., Clarke, C., Clynes, M., Barron, N., 2014a. CHO cell culture longevity and recombinant protein yield are enhanced by depletion of miR-7 activity via sponge decoy vectors. Biotechnol. J. 9, 396–404. https://doi.org/10.1002/biot.201300325

Sanchez, N., Kelly, P., Gallagher, C., Lao, N.T., Clarke, C., Clynes, M., Barron, N., 2014b. CHO cell culture longevity and recombinant protein yield are enhanced by depletion of miR-7 activity via sponge decoy vectors. Biotechnol. J. 9, 396–404.

Sanders, P.G., Hussein, A., Coggins, L., Wilson, R., 1987. Gene amplification: the Chinese hamster glutamine synthetase gene. Dev. Biol. Stand. 66, 55–63.

Santoro, R., Lienemann, P., Fussenegger, M., 2009. Epigenetic engineering of ribosomal RNA genes enhances protein production. PLoS One 4, e6653.

Sasaki, T., Ramanathan, S., Okuno, Y., Kumagai, C., Shaikh, S.S., Gilbert, D.M., 2006. The Chinese Hamster Dihydrofolate Reductase Replication Origin Decision

Point Follows Activation of Transcription and Suppresses Initiation of Replication within Transcription Units. Mol. Cell. Biol. 26, 1051–1062. https://doi.org/10.1128/MCB.26.3.1051-1062.2006

Sauerwald, T.M., Betenbaugh, M.J., Oyler, G.A., 2002. Inhibiting apoptosis in mammalian cell culture using the caspase inhibitor XIAP and deletion mutants. Biotechnol. Bioeng. 77, 704–716.

Schmidt, F.R., 2004. Recombinant expression systems in the pharmaceutical industry. Appl. Microbiol. Biotechnol. 65, 363–372. https://doi.org/10.1007/s00253-004-1656-9

Schnall-Levin, M., Zhao, Y., Perrimon, N., Berger, B., 2010. Conserved microRNA targeting in Drosophila is as widespread in coding regions as in 3'UTRs. Proc. Natl. Acad. Sci. U. S. A. 107, 15751–15756. https://doi.org/10.1073/pnas.1006172107

Schoellhorn, M., Fischer, S., Wagner, A., Handrick, R., Otte, K., 2017. miR-143 targets MAPK7 in CHO cells and induces a hyperproductive phenotype to enhance production of difficult-to-express proteins. Biotechnol. Prog. 33, 1046–1058.

Senko, M.W., Remes, P.M., Canterbury, J.D., Mathur, R., Song, Q., Eliuk, S.M., Mullen, C., Earley, L., Hardman, M., Blethrow, J.D., 2013. Novel parallelized quadrupole/linear ion trap/Orbitrap tribrid mass spectrometer improving proteome coverage and peptide identification rates. Anal. Chem. 85, 11710–11714.

Sheen-Chen, S.-M., Sun, C.-K., Liu, Y.-W., Eng, H.-L., Ko, S.-F., Kuo, C.-H., 2007. Extremely elevated CA19-9 in acute cholangitis. Dig. Dis. Sci. 52, 3140–3142.

Shin, C., Nam, J.-W., Farh, K.K.-H., Chiang, H.R., Shkumatava, A., Bartel, D.P., 2010a. Expanding the microRNA targeting code: functional sites with centered pairing. Mol. Cell 38, 789–802.

Shin, C., Nam, J.-W., Farh, K.K.-H., Chiang, H.R., Shkumatava, A., Bartel, D.P., 2010b. Expanding the microRNA targeting code: functional sites with centered pairing. Mol. Cell 38, 789–802. https://doi.org/10.1016/j.molcel.2010.06.005

Sievers, E.L., Senter, P.D., 2013. Antibody-Drug Conjugates in Cancer Therapy. Annu. Rev. Med. 64, 15–29. https://doi.org/10.1146/annurev-med-050311-201823

Simeone, D.M., Ji, B., Banerjee, M., Arumugam, T., Li, D., Anderson, M.A., Bamberger, A.M., Greenson, J., Brand, R.E., Ramachandran, V., Logsdon, C.D., 2007. CEACAM1, a novel serum biomarker for pancreatic cancer. Pancreas 34, 436–443. https://doi.org/10.1097/MPA.0b013e3180333ae3

Slamon, D.J., Leyland-Jones, B., Shak, S., Fuchs, H., Paton, V., Bajamonde, A., Fleming, T., Eiermann, W., Wolter, J., Pegram, M., Baselga, J., Norton, L., 2001. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. N. Engl. J. Med. 344, 783–792. https://doi.org/10.1056/NEJM200103153441101

Snel, B., Lehmann, G., Bork, P., Huynen, M.A., 2000. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. Nucleic Acids Res. 28, 3442–3444.

Stolfa, G., Smonskey, M.T., Boniface, R., Hachmann, A.-B., Gulde, P., Joshi, A.D., Pierce, A.P., Jacobia, S.J., Campbell, A., 2018. CHO-Omics Review: The Impact of Current and Emerging Technologies on Chinese Hamster Ovary Based

Bioproduction. Biotechnol. J. 13, 1700227. https://doi.org/10.1002/biot.201700227
Sun, W., Julie Li, Y.-S., Huang, H.-D., Shyy, J.Y.-J., Chien, S., 2010. microRNA: A Master Regulator of Cellular Processes for Bioengineering Systems. Annu. Rev. Biomed. Eng. 12, 1–27. https://doi.org/10.1146/annurev-bioeng-070909-105314
Syka, J.E.P., Coon, J.J., Schroeder, M.J., Shabanowitz, J., Hunt, D.F., 2004. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc. Natl. Acad. Sci. 101, 9528–9533. https://doi.org/10.1073/pnas.0402700101
Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., 2014. STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 43, D447–D452.
Taber, R., Alexander, V., Whitford, W., 1976. Persistent reovirus infection of CHO cells resulting in virus resistance. J. Virol. 17, 513–524.
Tan, S., Tan, H.T., Chung, M.C.M., 2008. Membrane proteins and membrane proteomics. PROTEOMICS 8, 3924–3932. https://doi.org/10.1002/pmic.200800597
Tanford, C., Reynolds, J.A., 1976. Characterization of membrane proteins in detergent solutions. Biochim. Biophys. Acta 457, 133–170.
Tanti, G.K., Pandey, S., Goswami, S.K., 2015. SG2NA enhances cancer cell survival by stabilizing DJ-1 and thus activating Akt. Biochem. Biophys. Res. Commun. 463, 524–531.
The UniProt Consortium, 2011. Ongoing and future developments at the Universal Protein Resource. Nucleic Acids Res. 39, D214–D219. https://doi.org/10.1093/nar/gkq1020
Thomas, M., Lieberman, J., Lal, A., 2010. Desperately seeking microRNA targets. Nat. Struct. Mol. Biol. 17, 1169.
Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A., Narechania, A., 2003. PANTHER: A Library of Protein Families and Subfamilies Indexed by Function. Genome Res. 13, 2129–2141. https://doi.org/10.1101/gr.772403
Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Hamon, C., 2003. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal. Chem. 75, 1895–1904.
Thomson, D.W., Bracken, C.P., Goodall, G.J., 2011. Experimental strategies for microRNA target identification. Nucleic Acids Res. 39, 6845–6853. https://doi.org/10.1093/nar/gkr330
Tien, W.-S., Chen, J.-H., Wu, K.-P., 2017. SheddomeDB: the ectodomain shedding database for membrane-bound shed markers. BMC Bioinformatics 18. https://doi.org/10.1186/s12859-017-1465-7
Tigges, M., Fussenegger, M., 2006. Xbp1-based engineering of secretory capacity enhances the productivity of Chinese hamster ovary cells. Metab. Eng. 8, 264–272. https://doi.org/10.1016/j.ymben.2006.01.006
Trummer, E., Fauland, K., Seidinger, S., Schriebl, K., Lattenmayer, C., Kunert, R., Vorauer-Uhl, K., Weik, R., Borth, N., Katinger, H., 2006. Process parameter

shifting: Part I. Effect of DOT, pH, and temperature on the performance of Epo-Fc expressing CHO cells cultivated in controlled batch bioreactors. Biotechnol. Bioeng. 94, 1033–1044.

Urlaub, G., Käs, E., Carothers, A.M., Chasin, L.A., 1983. Deletion of the diploid dihydrofolate reductase locus from cultured mammalian cells. Cell 33, 405–412.

Urquhart, L., 2018. Market watch: Top drugs and companies by sales in 2017. Nat. Rev. Drug Discov. 17, 232. https://doi.org/10.1038/nrd.2018.42

van den Hoogenhof Maarten M.G., Pinto Yigal M., Creemers Esther E., 2016. RNA Splicing. Circ. Res. 118, 454–468. https://doi.org/10.1161/CIRCRESAHA.115.307872

Van Simaeys, D., Turek, D., Champanhac, C., Vaizer, J., Sefah, K., Zhen, J., Sutphen, R., Tan, W., 2014. Identification of Cell Membrane Protein Stress-Induced Phosphoprotein 1 as a Potential Ovarian Cancer Biomarker Using Aptamers Selected by Cell Systematic Evolution of Ligands by Exponential Enrichment. Anal. Chem. 86, 4521–4527. https://doi.org/10.1021/ac500466x

Vizcaíno, J.A., Csordas, A., del-Toro, N., Dianes, J.A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., Xu, Q.-W., Wang, R., Hermjakob, H., 2016. 2016 update of the PRIDE database and its related tools. Nucleic Acids Res. 44, D447-456. https://doi.org/10.1093/nar/gkv1145

Von Hoff, D.D., Ervin, T., Arena, F.P., Chiorean, E.G., Infante, J., Moore, M., Seay, T., Tjulandin, S.A., Ma, W.W., Saleh, M.N., Harris, M., Reni, M., Dowden, S., Laheru, D., Bahary, N., Ramanathan, R.K., Tabernero, J., Hidalgo, M., Goldstein, D., Van Cutsem, E., Wei, X., Iglesias, J., Renschler, M.F., 2013. Increased Survival in Pancreatic Cancer with nab-Paclitaxel plus Gemcitabine [WWW Document]. http://dx.doi.org/10.1056/NEJMoa1304369. https://doi.org/10.1056/NEJMoa1304369

Voyksner, R.D., Lee, H., 1999. Investigating the use of an octupole ion guide for ion storage and high-pass mass filtering to improve the quantitative performance of electrospray ion trap mass spectrometry. Rapid Commun. Mass Spectrom. 13, 1427–1437. https://doi.org/10.1002/(SICI)1097-0231(19990730)13:14<1427::AID-RCM662>3.0.CO;2-5

Wallin, E., Heijne, G.V., 1998. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. Protein Sci. 7, 1029–1038. https://doi.org/10.1002/pro.5560070420

Walsh, G., 2018. Biopharmaceutical benchmarks 2018. Nat. Biotechnol. 36, 1136–1145. https://doi.org/10.1038/nbt.4305

Wang, W., Zhou, H., Lin, H., Roy, S., Shaler, T.A., Hill, L.R., Norton, S., Kumar, P., Anderle, M., Becker, C.H., 2003. Quantification of Proteins and Metabolites by Mass Spectrometry without Isotopic Labeling or Spiked Standards. Anal. Chem. 75, 4818–4826. https://doi.org/10.1021/ac026468x

Wang, X., Proud, C.G., 2006. The mTOR pathway in the control of protein synthesis. Physiol. Bethesda Md 21, 362–369. https://doi.org/10.1152/physiol.00024.2006

Wei, Y.-Y.C., Naderi, S., Meshram, M., Budman, H., Scharer, J.M., Ingalls, B.P., McConkey, B.J., 2011. Proteomics analysis of chinese hamster ovary cells undergoing apoptosis during prolonged cultivation. Cytotechnology 63, 663–677. https://doi.org/10.1007/s10616-011-9385-2

Weiner, L.M., Surana, R., Wang, S., 2010. Monoclonal antibodies: versatile platforms for cancer immunotherapy. Nat. Rev. Immunol. 10, 317–327. https://doi.org/10.1038/nri2744

Wieduwilt, M.J., Moasser, M.M., 2008. The epidermal growth factor receptor family: Biology driving targeted therapeutics. Cell. Mol. Life Sci. CMLS 65, 1566–1584. https://doi.org/10.1007/s00018-008-7440-8

Wiśniewski, J.R., Nagaraj, N., Zougman, A., Gnad, F., Mann, M., 2010. Brain Phosphoproteome Obtained by a FASP-Based Method Reveals Plasma Membrane Protein Topology. J. Proteome Res. 9, 3280–3289. https://doi.org/10.1021/pr1002214

Wiśniewski, J.R., Ostasiewicz, P., Mann, M., 2011. High Recovery FASP Applied to the Proteomic Analysis of Microdissected Formalin Fixed Paraffin Embedded Cancer Tissues Retrieves Known Colon Cancer Markers. J. Proteome Res. 10, 3040–3049. https://doi.org/10.1021/pr200019m

Wiśniewski, J.R., Zougman, A., Nagaraj, N., Mann, M., 2009. Universal sample preparation method for proteome analysis. Nat. Methods 6, 359.

Wong, J.W.H., Cagney, G., 2010. An Overview of Label-Free Quantitation Methods in Proteomics by Mass Spectrometry, in: Hubbard, S.J., Jones, A.R. (Eds.), Proteome Bioinformatics, Methods in Molecular Biology™. Humana Press, Totowa, NJ, pp. 273–283. https://doi.org/10.1007/978-1-60761-444-9_18

Wu, Z., Kuntz, A.I., Wadleigh, R.G., 2013. CA 19-9 tumor marker: is it reliable? A case report in a patient with pancreatic cancer. Clin. Adv. Hematol. Oncol. HO 11, 50–52.

Xu, D., 2004. Protein Databases on the Internet. Curr. Protoc. Mol. Biol. CHAPTER, Unit-19.4. https://doi.org/10.1002/0471142727.mb1904s68

Xu, X., Nagarajan, H., Lewis, N.E., Pan, S., Cai, Z., Liu, X., Chen, W., Xie, M., Wang, W., Hammond, S., Andersen, M.R., Neff, N., Passarelli, B., Koh, W., Fan, H.C., Wang, Jianbin, Gui, Y., Lee, K.H., Betenbaugh, M.J., Quake, S.R., Famili, I., Palsson, B.O., Wang, Jun, 2011. The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. Nat. Biotechnol. 29, 735–741. https://doi.org/10.1038/nbt.1932

Yamamoto, T., 2015. Long-term survival after resection of pancreatic cancer: A single-center retrospective analysis. World J. Gastroenterol. 21, 262. https://doi.org/10.3748/wjg.v21.i1.262

Yamane-Ohnuki, N., Kinoshita, S., Inoue-Urakubo, M., Kusunoki, M., Iida, S., Nakano, R., Wakitani, M., Niwa, R., Sakurada, M., Uchida, K., Shitara, K., Satoh, M., 2004. Establishment of FUT8 knockout Chinese hamster ovary cells: an ideal host cell line for producing completely defucosylated antibodies with enhanced antibody-dependent cellular cytotoxicity. Biotechnol. Bioeng. 87, 614–622. https://doi.org/10.1002/bit.20151

Yang, G., Hu, Y., Sun, S., Ouyang, C., Yang, W., Wang, Q., Betenbaugh, M.J., Zhang, H., 2018. Comprehensive Glycoproteomic Analysis of Chinese Hamster Ovary Cells. Anal. Chem. https://doi.org/10.1021/acs.analchem.8b03520

Yost, R.A., Enke, C.G., 1978. Selected ion fragmentation with a tandem quadrupole mass spectrometer. J. Am. Chem. Soc. 100, 2274–2275. https://doi.org/10.1021/ja00475a072

Zauber, A.G., Winawer, S.J., O'Brien, M.J., Lansdorp-Vogelaar, I., van Ballegooijen, M., Hankey, B.F., Shi, W., Bond, J.H., Schapiro, M., Panish, J.F., Stewart, E.T., Waye, J.D., 2012. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. N. Engl. J. Med. 366, 687–696. https://doi.org/10.1056/NEJMoa1100370

Zenobi, R., Knochenmuss, R., 1998. Ion formation in MALDI mass spectrometry. Mass Spectrom. Rev. 17, 337–366. https://doi.org/10.1002/(SICI)1098-2787(1998)17:5<337::AID-MAS2>3.0.CO;2-S

Zhang, W., Wei, Y., Yu, D., Xu, J., Peng, J., 2018. Gefitinib provides similar effectiveness and improved safety than erlotinib for advanced non-small cell lung cancer. Medicine (Baltimore) 97. https://doi.org/10.1097/MD.0000000000010460

Zhu, J., 2012. Mammalian cell protein expression for biopharmaceutical production. Biotechnol. Adv. 30, 1158–1170. https://doi.org/10.1016/j.biotechadv.2011.08.022

2 Chapter Two: Proteomic strategies in the search for novel pancreatic cancer biomarkers and drug targets: recent advances and clinical impact

Published in *Expert Review of Proteomics* March 2016

DOI: 10.1586/14789450.2016.1167601

Authors: Orla Coleman, Michael Henry, Gerard McVey, Martin Clynes, Michael Moriarty and Paula Meleady

I performed the literature survery and review of published studies surrounding proteomic studies of pancreatic cancer. The manuscript was written, prepared and subsequently submitted by me to *Expert Review of Proteomics* after an invited submission was given to Prof. Martin Clynes for the journal. All revisions for this manuscript were also performed by me.

Abstract

Pancreatic ductal adenocarcinoma (PDAC) is one of the deadliest cancers, despite a low incidence rate it is the fourth leading cause of cancer-related death in the world. Improvement of the diagnosis, prognosis and treatment remains the main focus of pancreatic cancer research. Rapid developments in proteomic technologies have improved our understanding of the pancreatic cancer proteome. Here, we summarise the recent proteomic strategies undertaken in the search for: novel biomarkers for early diagnosis, pancreatic cancer-specific proteins which may be used for novel targeted therapies and proteins which may be useful for monitoring disease progression post-therapy. Recent advances and findings discussed here provide great promise of having a significant clinical impact and improving the outcome of patients with this malignancy.

2.1 Introduction

Pancreatic cancer is a lethal common cancer as it is usually diagnosed at an advanced stage and can be resistant to therapy. Pancreatic ductal adenocarcinoma (PDAC) represents the majority of pancreatic cancers, accounting for 80-90% of pancreatic cancer cases. It is the fourth leading cause of cancer-related mortality in the world and develops in a relatively symptom-free manner leading to rapid disease progression and usually a dismal prognosis at the time of diagnosis. It has the lowest survival rate among human cancers with an overall 5-year survival rate of less than 5% and a median survival of 3 to 6 months for patients with metastatic pancreatic cancer (Vincent,A., Herman, J., Schulick, R., Hruban, R.H.and Goggins, M. 2013, Chue 2008). The most recent estimates for pancreatic cancer from the American Cancer Society for 2015 states about 48,960 people in the United States will be diagnosed with this cancer and about 40,560 Americans will die of pancreatic cancer in 2015 (www.cancer.org). For the majority of cases this cancer is advanced, unresectable and metastatic, for the minority of patients who are diagnosed at an early stage, conventional cancer treatments have limited benefit and little impact on disease progression.

The risk factors and genetic mutations associated with pancreatic cancer have been well studied and characterised. As with most tumours, pancreatic cancer occurs mainly in older individuals with only 10% of patients below the age of 50 (Lowenfels and Maisonneuve 2006). It is estimated that 5 to 10% of pancreatic cancers are due to an inherited predisposition; however hereditary does not pose as big a threat as environmental factors, with cigarette smoking estimated to account for 20-25% of pancreatic cancer incidence (Bosetti et al. 2012). Other factors such as pancreatitis, Peutz-Jeghers syndrome, familial atypical multiple mole melanoma and hereditary nonpolyposis colon cancer also increase the risk of disease progression (Ryan, Hong and Bardeesy 2014).

At present, serum carbohydrate 19-9 (CA19-9) is the only pancreatic cancer biomarker approved for use by the FDA; however, it’s lack of specificity and sensitivity limits its application. CA19-9 can be elevated in many types of gastrointestinal cancers such as colorectal cancer and esophageal cancer, as well as in patients with pancreatitis,

cirrhosis and diseases of the bile ducts (Takahashi et al. 2003, Maestranzi et al. 1998). Due to these limitations CA19-9 should mainly be used to monitor pancreatic cancer treatment or recurrence post surgery but not as an indicator of disease due to the high chance of obtaining a false positive result. Due to a lack of dependable biomarkers early detection of this cancer is currently difficult. This situation emphasises the need for more specific biomarkers against pancreatic cancer to be identified.

Molecular analyses of pancreatic adenocarcinoma have revealed signatures of genetic lesions associated with this aggressive neoplasm including mutations of K-*ras*, CDKN2A (p16), TP53 and BRCA2 (Ryan, Hong and Bardeesy 2014, Li, D. et al. 2004). These mutations are classical tumour-suppressor defects which are seen in various other cancer types. Pancreatic cancer has the highest frequency (>85%) of K-*ras* mutation among all human cancers, with an increase in frequency with disease progression; it therefore seems to be an unavoidable element of this malignancy (Rozenblum et al. 1997a, Schutte et al. 1997). Activating mutations of the *ras*-family of oncogenes results in a multitude of downstream cellular effects including induction of proliferation, survival and invasion by triggering several effector pathways (Fernández-Medarde and Santos 2011, Shields et al. 2000). Pancreatic intraepithelial neoplasia (PanIN) is the most common precursor lesion of pancreatic cancer (Koorstra et al. 2008). Inactivation of the CDKN2A/p16 tumour-suppressor gene is found in 40% of precursor PanIN lesions and disruption of the p16 pathway is seen in 98% of all pancreatic carcinomas (Rozenblum et al. 1997b, Schutte et al. 1997). The p16 pathway is a well described component of cell cycle regulation and alterations to this pathway leads to progression of the cell cycle through the G1/S checkpoint resulting in uncontrolled tumour growth (Cowgill and Muscarella 2003). Rozenblum and colleagues (Rozenblum et al. 1997b) in a comprehensive analysis of 42 pancreatic ductal cancers found that all tumours possessed mutations in the K-*ras* oncogene. Mutational frequencies of the tumour suppressor genes p16, TP53 and BRCA2 were 82%, 76% and 10% respectively.

At the genetic level, pancreatic cancer is well characterised; however these findings have failed to advance our understanding of disease development. The biology of pancreatic cancer is a complex multidimensional process related to mutation and

inactivation of oncogenes and tumour suppressor genes, as well as aberrant expression of growth factors. These abnormalities permit pancreatic cancer cells to survive and grow extensively and almost undetectably. The development of invasive and metastatic phenotypes poses another threat to the fate of patients as such tumours have proved resistant to conventional therapies. While genomics has significantly improved our molecular understanding of this disease, identifying targets suitable for treatment is difficult. Pharmacologic targeting of genetic mutations is not a viable option as the genes exploited by pancreatic cancer are key regulators in cell cycle progression for normal cells. There is undoubtedly an unmet need for novel therapies to be developed with current treatments failing to improve survival rates of patients.

In the post-genomic era, it has become evident that genetic changes alone are not sufficient to understand most disease processes including pancreatic cancer. Proteome-based approaches are important complements to genomic data and provide critical information about the driver molecules behind this disease. Proteomics delineates the functional units of a cell, proteins and their intricate interaction network and signalling pathways for the underlying disease (Boja and Rodriguez 2011). Proteomic studies have been successfully used to identify several protein alterations in tumour cells leading to biomarker discoveries (de Wit et al. 2013, Pontén et al. 2011, Hu et al. 2008, Adam et al. 2001). By applying proteomic technologies, this is an alternative way to identify biomarkers for potentially early diagnosis, targeted therapy and monitoring of disease.

2.2 Current status of PDAC treatment

Pancreatic cancer is an inherently resilient disease which results in a lack of effective treatments for this carcinoma. Current treatments for this disease are dependant on disease stage. Surgical resection of a pancreatic tumour is the best chance of curing patients of PDAC. It is estimated that 95% of patients with this disease present to a doctor at the advanced and unresectable stage (Surlin et al. 2014). Pancreatoduodenectomy (PD), or the Whipple procedure, was first described in 1935 for the surgical removal of pancreatic cancer (Whipple, Parsons and Mullins 1935). Advances in surgical techniques since this has lead to the development of variations of

this method with more precise and simplified operations being routinely employed. Despite the advances in surgical approaches, post-resection survival rates have not improved significantly over the past decade (Winter et al. 2012). More recently, studies have provided evidence that suggests surgical resection followed by treatment with the conventional PDAC chemotherapy Gemcitabine, may lead to a significant increase in overall survival as well as disease-free survival (Oettle et al. 2013).

For over a decade, gemcitabine has been the standard chemotherapy treatment for patients with unresectable and metastatic pancreatic cancer, however many studies have demonstrated the failure of gemcitabine alone as an effective treatment (Samulitis et al. 2015, Lee et al. 2013, Schniewind et al. 2004). Despite this, combinational therapy of therapeutic agents holds great hope for pancreatic cancer treatment. In 2013, Ciliberto et al. performed a meta-analysis of randomised clinical trials to investigate the effectiveness of gemcitabine-based combination therapies compared to gemcitabine alone (Ciliberto et al. 2013). Thirty-four clinical trials which consisted of 10,660 patients in total were analysed in terms of overall survival, response rate and toxicity ratios among other things. Compared to gemcitabine monotherapy, combinational chemotherapy significantly improved overall survival and response rate. However, the combined therapy also resulted in an increase in treatment-related toxicity therefore demonstrating a greater hazard to patients with advanced pancreatic cancer than gemcitabine alone. The recent MPACT clinical trial carried out by Von Hoff et al. (2013) showed that in comparison to gemcitabine therapy alone for patients with metastatic pancreatic adenocarcinoma, albumin-bound paclitaxel in conjunction with gemcitabine significantly improved overall survival, progression-free survival, and response rate (Von Hoff et al. 2013). Similar to other studies of the same concept these advantages are challenged by significant drawbacks such as increased myelosuppression, peripheral neuropathy and toxic effects. Emerging chemotherapies such as FOLFIRINOX has recently been recognised as an alternative first-line pancreatic chemotherapy with improved performance in comparison to gemcitabine (Boone et al. 2013, Conroy et al. 2011).

Molecular targeted therapy has dawned a new era of cancer treatments with hopes for greater efficacy than conventional therapies and less harmful to normal cells. By focusing treatment towards specific key molecules involved in carcinogenesis and tumour growth the consequent side effects are not as destructive as those seen in chemotherapy. Successful targeted therapies preferentially target a molecule, most often a protein, that carries a mutation specific to cancer cells or a molecule that is more highly expressed in cancer cells than normal cells. The success of these therapies therefore undeniably lies in the identification and subsequent targeting of tumour-restricted biomarkers. Vandetanib is a tyrosine kinase inhibitor that targets vascular endothelial growth factor receptor (VEGF), epidermal growth factor receptor (EGFR) and rearranged during transfection receptor (RET), which are important for tumour proliferation and invasion in pancreatic cancer (Falchook et al. 2014, Costello, Greenhalf and Neoptolemos 2012). A phase I trial was performed by Kessler et al. (2015) analysing vandetanib in combination with gemcitabine and capecitabine in patients of pancreatic cancer. The trial set out to combine cytotoxic chemotherapy with a targeted agent to improve clinical benefit without increasing toxicity. The combination of gemcitabine, capecitabine and vandetanib revealed promising activity in pancreatic cancer patients with further validation warranted in a larger cohort of patients. Similarly, Erlotinib is a tyrosine kinase inhibitor that targets EGFR thereby inhibiting tumour cell growth and differentiation. Molecular targeted therapies have been extensively studied in clinical trials, the only success erlotinib in combination with gemcitabine shows modest prolonged survival compared with gemcitabine alone (Moore et al. 2007) and received FDA approval in 2005. More recently a phase II study of Ruxolitinib in combination with capecitabine was performed for patients with metastatic pancreatic cancer where previous gemcitabine treatment had failed (Hurwitz et al. 2015). Ruxolitinib is an inhibitor of JAK1/JAK2 of the Janus kinase pathway, inhibition of which ultimately leads to decreased cancer cell proliferation and differentiation, therefore may improve survival of patients.

2.3 Application of clinical proteomics

The drug discovery model has seen a shift in recent years with the emphasis now being placed on finding a suitable "druggable" protein for a particular disease. Target identification has been approached with a variety of methods; this review focuses on the application of proteomics for novel drug target discovery and drug development for pancreatic cancer. Proteomics has evolved extensively over the past decade, providing new methods to tackle increasing challenges in health care. Analysis of the complete human genome sequence resulted in the identification of ~ 20,687 protein-coding genes (ENCODE Project Consortium 2012). Historically the most widely used proteomic tools are two-dimensional gel electrophoresis mass spectrometry (2DGE-MS) and liquid chromatography mass spectrometry (LC-MS). High-resolution 2-DE systems allow resolution of up to 10,000 protein spots of an entire cell lysate (Wittmann-Liebold, Graack and Pohl 2006) ;however, difficulties are met when several proteins co-migrate to the same spot (Cellulaire 2002). Challenges associated with gel-based techniques have been overcome with the advent of newer methods for protein sample separation and quantification, a basic overview of such techniques and workflow is shown in figure 2.1. Liquid chromatography (LC) is currently one of the most widely used technologies for peptide or protein separation prior to mass spectrometry analysis. Chromatography allows for the collection of both quantitative and structural information, and can achieve picogram/mL^{-1} sensitivities (Theodoridis et al. 2012). Other such techniques include high performance liquid chromatography (HPLC) and capillary electrophoresis (CE), a comparison of their performance when interfaced with the same MS for biomarker analysis was analysed by Mullen et al. (Mullen et al. 2012).

Current methods to reduce sample complexity exploit the inherent nature of proteins. Subcellular fractionation (commercially available kits or centrifugation methods) can separate protein samples into nuclear, membrane and cytoplasmic fractions. Post-translational modification (PTM) isolation kits and resins can enrich PTM substrates and allow for PTM site identification. Protein depletion can allow greater analysis of the proteome by removing/depleting highly abundant proteins in samples allowing for the detection of low abundant ones. Protein depletion spin columns

are optimised to decrease the abundant albumin and antibody components of serum samples in preparation for analysis. Such columns can remove up to the top 20 most highly abundant proteins from plasma samples to enhance the detection of lower abundance proteins in proteomic analyses (Tu et al. 2010). Tissue extraction methods allow extraction of protein from a wide variety of tissue samples. Combining subcellular fractionation with LC-MS/MS over 3,900 proteins were identified in four pancreatic cancer cell lines with 134 proteins significantly differentially expressed between primary and metastatic cell populations (McKinney et al. 2012). The technique of subcellular proteomics for the in-depth profiling of differential protein expression allows for large-scale proteomic comparisons giving a more global insight into the proteome of pancreatic cancer. Similarly, numerous techniques are available for the isolation of cellular proteins based on their post-translational modification which allows for extraction of low-abundant proteins and specific sub-classes of proteins. LC-MS/MS phosphoproteomics of 12 PDAC tissue samples identified 2,101 proteins of which, 152 demonstrated significant difference in abundance between tumour and non-tumour tissue (Britton et al. 2014). More importantly 635 phosphopeptides showed significant regulation and were found to contribute to migration, invasion and proliferation. This study highlights how analysis of the phosphoproteome of pancreatic cancer can be used to better understand the molecular events leading to disease development and such proteins may be used as effective new therapeutic targets for this malignancy.

Sample preparation has also undergone advances with regards the simultaneous identification and quantification of two or more biological samples. Stable Isotope Labelling by Amino acids in cell culture (SILAC) is a metabolic isotopic labelling technique that involves culturing cells in growth media containing amino acids isotopes, usually arginine and lysine, that become incorporated into mammalian proteins (Ong et al. 2002). Proteins from the experimental and control samples are combined and relative quantification of protein differences is enabled through mass spectrometry. Other proteomic labelling technologies include isotope-coded affinity tags (ICAT) (Gygi et al. 1999), isobaric tags for relative and absolute quantitation (iTRAQ)(Wiese et al. 2007) and tandem mass tags (TMT) (Thompson, A. et al. 2003) whereby chemical labelling occurs to the proteins or peptides from a sample rather than at the cell culture stage.

Quantitative label-free LC-MS/MS proteomic analysis is a popular current method of protein identification and quantification whereby sample handling and potential contaminants are reduced allowing for high-throughput of samples. Quantitative label-free proteomics is based on the direct comparison of MS signal intensities between different peptides in a mass spectrometer (Zhu, Smith and Huang 2010). Quantitative information is gathered using of of two methods: spectral counting or area under the curve. All of the aforementioned techniques enable reliable preparation of samples in order to quantitate differences in protein abundance in cells or tissues.

Advances in proteomic sample preparation and separation has been met in parallel with extensive developments in mass spectrometric instruments and software for protein identification and protein expression analysis over the last 10 years. Modern mass spectrometers allow high-resolution sequencing, identification of post-translational modifications and have also significantly increased the number of protein identifications per run. The increased demand for mass spectrometric based proteomic quantitation of biological samples has seen the development of MS instruments from several companies such as Thermo Scientific, Waters, Applied Biosystems Sciex, Bruker Daltonics, Agilent and Shimadzu. Kosanam et al. (2013) performed proteomic analysis of four PDAC tissues and their adjacent benign tissues using a LTQ-Orbitrap XL mass spectrometer (Thermo Fisher Scientific) (Kosanam et al. 2013). This method identified a total of 2190 proteins, sixteen promising candidates were further validated of which laminin gamma 2 (LAMC2) showed significant promise as a putative PDAC biomarker. Scheltema et al. (2014) provide an excellent overview of the state-of-the-art Q Exactive HF benchtop mass spectrometer (Scheltema et al. 2014). This study identified 5,000 proteins in a standard 90-minute gradient of the tryptically digested HeLa mammalian cell lysate. The Thermo range of Q Exactive mass spectrometers are powerful proteomic instruments which significantly increase the number of peptide and protein identifications in a sample. LaConti et al. (2015) carried out extensive serum metabolomics profiling on mouse PDAC models at various stages of disease progression (LaConti et al. 2015). Serum samples were analysed using UPLC coupled to TOF-MS from Waters to identify significant metabolites associated with PDAC progression. Quantification of validated metabolite markers was carried out using MRM on a triple

quadrupole spectrometer ABI QTRAP 4000 LC/MS/MS system. This analysis revealed distinct metabolic patterns in the serum samples which could accurately distinguish between healthy control samples and PDAC samples and highlights the convenience of using animal models for biomarker discovery and validation. Lin et al. (2015) provide an extensive review of recently developed high-resolution mass spectrometers and provide great detail on the principle and functional characteristics of different types and/or models of mass analysers that have recently entered the market (Lin et al. 2015). The developments in mass spectrometers over the past decade has been expansive and these newer technologies will change the way clinical studies identify and quantify biomarkers for various diseases. Optimistically thinking, the introduction of these instruments could potentially be the solution to biomarker discovery, therapeutic target identification and disease staging for patients of pancreatic cancer. Commercially available and open source database search algorithms (e.g. SEQUEST, MASCOT, PEAKS DB, OMSSA, Protein Prospector) are essential for confident protein identification while quantitation software packages allow detection of significant changing proteins in samples being analysed (e.g. Progenesis QI, MaxQuant, SIEVE).

Reliable and reproducible quantification of a protein is an essential part of transitioning a biomarker from the research setting into the clinical environment. Quantitation can be achieved through relative or absolute quantitation of peptides from a protein biomarker. Relative quantitation is achieved by comparing the signal intensity of an analyte from one sample to the signal intensity of the same analyte in a different sample. Relative quantitation methods include labelling techniques previously mentioned such as TMT and iTRAQ that can simultaneously gather signal intensity measurements of an analyte from two samples. Quantitative label-free proteomics can also be used to determine relative abundances of a peptide between samples. Absolute quantitation is achieved through selected-reaction monitoring (SRM) or multiple reaction monitoring (MRM) which is performed on triple-quadrupole mass spectrometers. MRM is a valuable method for validating and quantifying potential biomarkers as candidates across a multitude of samples (Ansari et al. 2014). This technique involves selection of an ion, which represents a peptide for the protein of interest, followed by fragmentation and detection of the product ion with quantitation

based on signal intensity of the detection. MRM analysis allows the quantification of hundreds of peptides and as a result the protein biomarker of interest in the sample. This platform for biomarker validation is feasible and comparable with other techniques such as protein microarrays, ELISA and Luminex assays (Pan et al. 2011).

As proteomic technologies continue to improve, modern mass spectrometric instruments are doubling the sequencing speed at the same resolution which translates to increased peptide sequences and protein identifications. Advances in proteomics technology and methodology have helped to better our understanding of cancer biology and cancer progression. The growth in proteomic research has been stimulated by advances in mass spectrometric methods with faster, more accurate and more sensitive instrumentation continually being developed. The availability of powerful new technologies provides optimism towards future progress in treating pancreatic cancer.

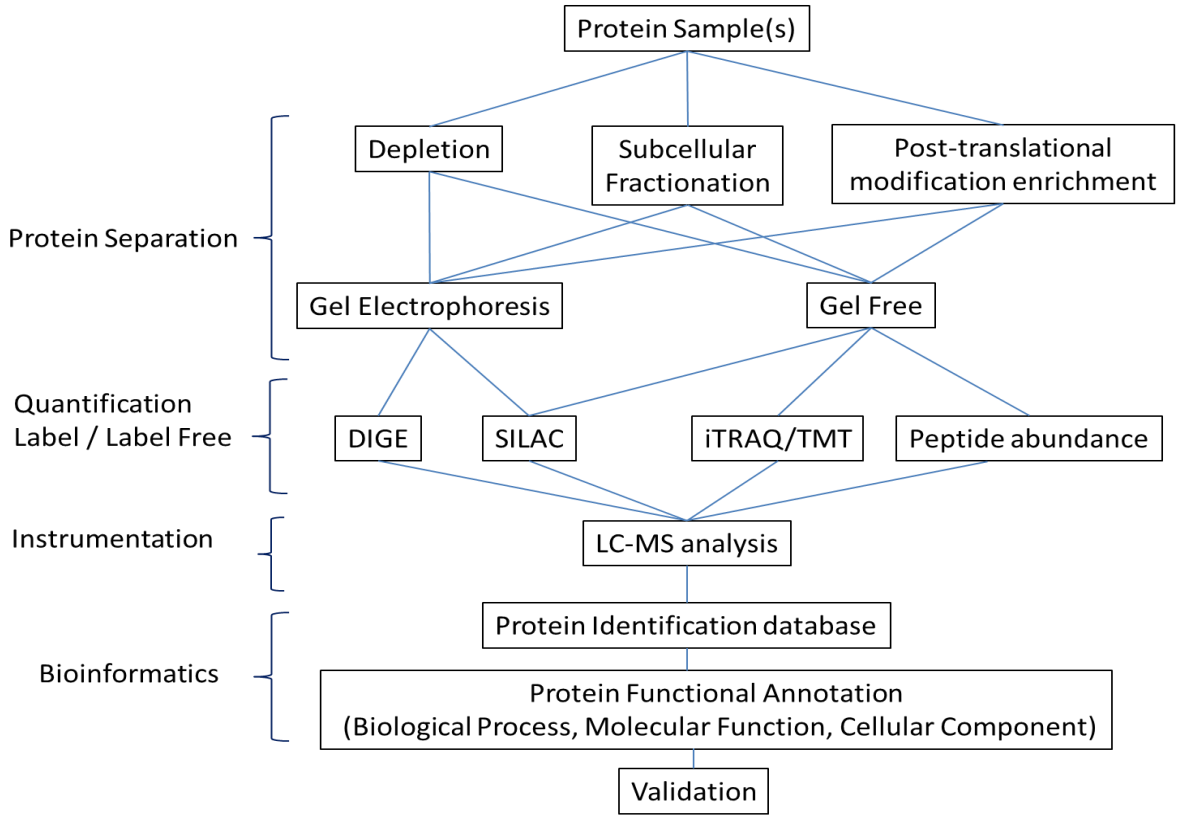


Figure 2.1 Basic overview of main quantitative proteomics techniques.

2.4 Proteomic studies in pancreatic cancer

Biomarker discovery is one strategy to reduce mortality and morbidity of cancer patients. Early detection at the asymptomatic stages of cancer progression is the key to providing a better outcome for patients. Diagnostic, prognostic and predictive markers for a disease are the ideal clinical tools for combating cancer malignancy. Present tumour marker examples include carcinoembryonic antigen (CEA) and prostate-specific antigen (PSA) for prostate cancer, CA-125 for ovarian cancer, CA 15-3 for breast cancer and CA 19-9 for pancreatic cancer. However, the sensitivity and specificity of these biomarkers are not always accurate due to the fact that several of these tumour markers are also produced by non-cancerous tissues. The application of high-end proteomic technologies for biomarker discovery could overcome the problem of using non-specific biomarkers by identifying precise proteome changes in disease versus normal patients. A summary of proteomic pancreatic cancer analyses performed since 2012 is outlined in table 2.1 while other studies are discussed in detail below according to sample type.

A comprehensive review by Harsha et al. (2009) suggests that a compendium of biomarkers for pancreatic cancer already exists and that efforts need to be aimed at transitioning these biomarkers found from scientific research to a clinical setting (Harsha et al. 2009). A total of 2,516 genes are included in the compendium which includes molecules with evidence of overexpression at the mRNA level, protein level or both. Many of the molecules described by the group as potential biomarkers require further interrogation and validation

2.4.1 Proteomic analysis of pancreatic cancer cell lines

Cell lines have long been used as *in-vitro* model systems as they offer many advantages such as being easily obtainable, immortal given the appropriate conditions and allow for experimental manipulation which can not be achieved with tissue specimens. The most relevant limitation of cell lines is that they might not be representative of tissue samples in the clinical setting. Currently, there are approximately 10 common pancreatic cancer

cell lines used for proteomic studies. These cell lines vary phenotypically and genotypically (Deer et al. 2010), therefore choosing a cell line model for pancreatic cancer depends on the analysis and question being asked.

Currently, gemcitabine is the first-line treatment for chemotherapy of pancreatic cancer despite showing resistance in many cases. A proteomic profiling study of pancreatic cancer cell lines employed LC-MS/MS technology to investigate the underlying mechanisms for gemcitabine resistance (Kim, Yikwon et al. 2014). Two PDAC cell lines, BxPc3 and PANC-1 were used (K-ras wildtype and K-ras mutant respectively) and both were compared to the normal cell line HPDE. A total of 837 proteins were identified as differentially expressed between the cell lines, 7 of which were validated using western blotting. To determine protein signatures of the differentially expressed proteins, K-means clustering was performed which yielded 6 distinct clusters each enriched in a specific pathway. Overall, the majority of differentially expressed proteins are involved in tight junctions, adheren junctions, and glutathione and cysteine/methionine metabolism pathways, which are related to drug resistance in pancreatic cancer. Using well-characterised pancreatic cell lines, comprehensive proteomic profiling has been generated that serves as a reference for gemcitabine resistance in pancreatic cancer.

Pancreatic cancer is usually asymptomatic in nature, leading to locally advanced or metastatic disease at the time of diagnosis. In this regard, Liu et al. (2010) investigated the functional proteins that promote metastasis in pancreatic cancer by isolating membrane proteins of pancreatic cell lines and identifying them using LC-MS/MS. By comparing the primary cell line BxPc3 to the metastatic cell line AsPc1, identification of potential target proteins related to metastasis of pancreatic cancer could be achieved. The proteins showing differential levels include cadherins, catenin, integrins, galectins, annexins, collagens and many others, which are known to have roles in tumour cell adhesion or motility. Cadherins are a class of type-1 transmembrane proteins that depend on calcium ions and combined complexes with cathenin to mediate cell adhesion. Cadherin proteins were solely found in primary tumour cells (BxPc3) but not in metastatic tumour cells (AsPc1), suggesting an alteration in cellular adhesion

capabilities in the metastatic AsPc1 cells. Similarly, integrins are glycoprotein members that form heterodimer receptors on the cellular membrane that link the extracellular matrix to the cytoskeleton. This family of proteins also play a major role in adhesion and attachment of the cell and were only identified in BxPc3 cells. This proteomic approach for identification and analysis of membrane proteins in primary and metastatic PDAC cells has revealed many protein biomarkers with the potential to distinguish pancreatic cancer patients for a better treatment decision based on metastasis probability.

Sample type	Proteomic technique	Reference	Principal findings
Tissues	LC-MS/MS	(Ohmine et al. 2015)	Protein expression levels of deoxycytidine kinase (dCK) in pancreatic cancer tissue is a good predictor of progression-free survival
Cell lines	LC-MS/MS	(Singh et al. 2015)	Sanguinarine treated pancreatic cancer cell lines showed downregulation of multiple pathways relevant to pancreatic cancer cell survival
Urine	GeLC-MS/MS	(Radon et al. 2015)	A three-biomarker urine panel consisting of LYVE1, REG1A and TFF1 showing high accuracy and diagnostic performance for PDAC
Cell lines	SILAC and LC-MS/MS	(Kim, M. S. et al. 2014)	Heterogeneity at the proteome level of metastatic pancreatic cancer

Serum	Glycoproteomics, TMT and label-free LC-MS/MS	(Nie et al. 2014)	A glycoprotein biomarker panel with high diagnostic potential for pancreatic cancer
Tissues	iTRAQ and LC-MS/MS	(Wang, Wan-Sheng et al. 2013)	The identification of several key proteins involved in the development of pancreatic cancer associated diabetes mellitus
Tissues	LC-MS/MS and SRM	(Takadate et al. 2013)	Four protein candidates for prognostic markers of PDAC
Plasma	SRM/MRM	(Yoneyama et al. 2013)	Absolute quantitation of proline-hydroxylated α-fibrinogen in plasma revealed its potential for pancreatic cancer diagnosis
Blood	2D-DIGE and MALDI-TOF-MS	(Chen, J. et al. 2013)	Complement 3 protein may be a specific serum biomarker of pancreatic carcinoma
Cell lines	LC-MS/MS	(McKinney et al. 2012)	134 proteins significantly differentially expressed between primary pancreatic cancer cell lines and cell lines derived from metastatic sites of pancreatic cancer
Cell lines	MALDI-TOF-MS	(Wang, W. B. et al. 2012)	Prohibitin 2 and prohibitin may be candidate immunogenic membrane antigens of pancreatic cancer

Table 2.1:A selectionof proteomic studies for pancreatic cancer performed since 2012 using proteomic techniques.

2.4.2 Proteomic analysis of pancreatic cancer tissue

Surprisingly, the majority of proteomic analyses of pancreatic cancer have been carried out on tissue specimens despite the low frequency of this disease. One of the main concerns with pancreatic cancer tissue proteomics is the extensive stroma microenvironment that surrounds the cancer cells. In pancreatic ductal adenocarcinoma the stroma is thought to make up to 90% of the tumour mass and consists of many factors including pancreatic stellate cells, fibroblasts, inflammatory cells, endothelial cells and extra cellular matrix proteins (Phillips, Grippo and Munshi 2012). In 2005, a study by Chen et al. was the first application of isotope-coded affinity tag technology for proteomic analysis of pancreatic cancer tissue (Chen, Ru, Eugene, Donohoe, Pan, Eng, Cooke, Crispin, Lane, Goodlett and Bronner 2005a). The study identified 151 differentially expressed proteins in cancer with at least a 2-fold change. Moreover, several upregulated proteins identified in this study play a role in the cancer cell and stroma relationship. Annexin A2 which was found to be 2.7-fold upregulated in cancer tissue, binds to and activates matrix metalloproteinases on the membrane of cancer cells leading to degradation of the extracellular matrix advancing tumour cell invasion. Many of the differentially expressed proteins identified in this study are thought to orchestrate tumour growth, migration, invasion and metastasis highlighting the importance of the tumour microenvironment in promoting pancreatic cancer progression.

Subsequently over the past decade there has been an explosion in pancreatic cancer proteomic analyses using tissue samples. Pan et al (2013) provides an excellent review article discussing pancreatic cancer tissue proteomic studies in the past few years in light of discovery and technology developments (Pan et al. 2013a). Several important proteins identified in tissue studies of pancreatic cancer include galectin-1 and galectin-3 (Chen, Jian-Hua et al. 2009, Berberat et al. 2001), gelsolin (Thompson, C. C. et al. 2007) and plectin-1 (Bausch et al. 2011).

Galectin-1 and galectin-3 were found to be significantly increased at the mRNA and protein levels in a study comprising of 33 primary pancreatic cancers versus 28 normal pancreases. This study indicates that galectin-1 plays a role in the desmoplastic reaction around cancer cells whereas galectin-3 appears to be involved in cancer cell

proliferation. It is suggested that high levels of galectin-3 could be used as an indicator of cancer cell development and growth in pancreatic cancer. Similarly, when comparing pancreatic cancer tissue to normal tissue using MALDI-TOF-MS, galectin-3 was found to be 3-fold higher in cancer patients (Chen, Jian-Hua et al. 2009). A more recent study by Martinez-Bosch et al. (2014) of galectin-1 in a mouse model of PDAC analysed the functional role of this protein in the pancreatic tumour microenvironment (Martinez-Bosch et al. 2014). Genetic ablation of galectin-1 dampened tumour progression by inhibiting proliferation, angiogenesis, desmoplastic reaction and by stimulating a tumour-associated immune response, yielding a 20% increase in relative lifespan. Molecular analyses showed galectin-1 to have high stromal expression and is mediated in the activation of the Hedgehog pathway in the stroma of pancreatic carcinoma in the mouse model. These findings establish a function for galectin-1 in the tumour-stroma crosstalk in PDAC and provide a clinical reasoning for targeting this protein as a microenvironment-based therapeutic strategy.

Gelsolin, which was initially found to be overexpressed in pancreatic tissue specimens, was tested in a cohort of 60 clinically characterised plasma samples for its use as a diagnostic marker for pancreatic cancer (Pan et al. 2013a). This study applied single-reaction monitoring (SRM) based targeted proteomics to investigate gelsolin as a putative blood biomarker. The SRM technique of mass spectrometry allowed for absolute quantitation of the gelsolin protein, with its concentration range found to be between 17,000−910,000 ng/mL with the upper concentration levels in the cancer plasma samples. Evaluation of gelsolin as a biomarker candidate showed this protein to have 95% specificity and 75% sensitivity. However, due to the low occurrence of pancreatic cancer an ideal biomarker for general population screening would need a specificity of >99%. Nevertheless, these initial findings merit further validation in larger patient and control cohorts for their potential use in clinical applications.

Plectin-1 has recently been identified as a biomarker for early PDAC detection and its suitability as an imaging target was assessed (Bausch et al. 2011). Ideal biomarkers for PDAC would have the ability to distinguish conditions such as chronic pancreatitis from malignancy but also be able to detect the cancer at the preinvasive

stages. Novel biomarkers that satisfy these criteria have been difficult to identify for pancreatic cancer despite the abundance of potential biomarkers discovered to date. A study by Kelly et al. (2008) used mouse models of PDAC to screen for peptides that specifically bind to cell surface antigens on PDAC cells (Kelly et al. 2008). This work identified plectin-1 as a novel biomarker of PDAC which could distinguish PDAC cells from normal pancreatic ductal cells in vitro. Plectin-1 targeted peptides (PTP) conjugated to magnetofluorescent nanoparticles were generated and utilised as a specific imaging probe for PDAC. Carrying on from this work Bausch et al. (2011) designed an experimental cohort consisting of specimens of human PDAC, chronic pancreatitis and normal pancreas to validate plectin-1 as an imaging target. A total of 4 normal pancreata, 15 chronic pancreatitis, 14 PanIN I, 26 PanIN II, 15 PanIN III, 41 PDAC, 8 liver metastases, 11 lymph node metastasis, 10 with matching primary tumours, and 9 peritoneal metastases were obtained. Plectin-1 expression was found to be positive in all PDACs and negative in benign tissues, with its expression increasing during carcinogenesis. Plectin-1 expression was retained in all metastatic specimens and in-vivo imaging using PTP was able to specifically highlight the primary and metastatic tumours. This data suggests plectin-1 may be an ideal biomarker for PDAC and the utilisation of PTP imaging of the malignancy may serve as a non-invasive, feasible method for clinical screening of PDAC patients.

2.4.3 Proteomic analysis of pancreatic cancer secretome

Although tissue specimens and cell lines derived from patients of PDAC are valuable samples for proteomic analyses they are often difficult to obtain as reaching the pancreas of a patient is quite an invasive surgical method. Many other sources of material exist that is more readily accessible including serum, plasma, urine, pancreatic juice and bile among others. Proteins which are secreted from the pancreas, often termed the cancer secretome, is the portion of the proteome often expected to harbour the most promising biomarkers (Gronborg et al. 2006a). Establishing biomarkers for pancreatic cancer which can be detected straightforwardly in a non-invasive manner in a clinical location would be the ideal situation for diagnosis of patients.

A recent investigation of serum taken from pancreatic cancer patients up to 4 years prior to diagnosis identified Thrombospondin-1 as a potential biomarker for PDAC prior to clinical diagnosis (Jenkinson et al. 2015). The use of iTRAQ enabled comparisons of pooled serum from patients with PDAC, chronic pancreatitis, type 2 diabetes mellitus and healthy subjects. Proteomic profiling of the serum samples identified thrombospondin-1 as reduced pre-clinically and MRM analysis confirmed a significant reduction in its levels up to 24 months before PDAC diagnosis. This study highlights the possibility of detecting PDAC at a very early stage of disease progression which consequently could allow for improved prognosis. Similarly Hocker et al. investigated the possibility of detecting early-stage pancreatic cancer using serum profiling (Hocker et al. 2015). Using ESI-MS profiling this group were able to distinguish patients with early PDAC and chronic pancreatitis and also validated blind patient samples to the appropriate patient cohort.

A large study conducted in 2011 by Brand et al. sought to identify serum biomarkers for the detection of pancreatic cancer (Brand et al. 2011). The analysis was carried out on a large cohort of individuals comprised of 333 patients with histologically diagnosed PDAC, 144 patients with benign pancreatic conditions, 227 healthy controls without a history of pancreatic diseases, and patients diagnosed with colon (n = 33), lung (n = 62), and breast (n = 108) cancer. Eighty-three circulating proteins were analysed in sera of the study population resulting in several robust profiles of protein alterations presenting in PDAC patients compared to the healthy and benign groups. As previously mentioned CA 19-9 is currently the only FDA approved biomarker for PDAC, this protein was found to have higher expression in the PDAC group of this study compared to the others. Based on this it was decided to make various 2, 3 and 4-biomarker panels all of which include CA 19-9. The highest performing panel was found to be CA 19–9, ICAM-1 and osteoprotegerin which demonstrated a sensitivity and specificity of 78% and 94% respectively. This panel is selective for PDAC and demonstrated a high level of cancer selectivity when applied to colon, lung, and breast cancer. These findings promote the use of multi-marker panels with high classification power for PDAC and such panels may offer a high level of clinical utility

in combination with conventional imaging modalities to provide a diagnostic role for this malignancy.

A similar study of pancreatic cancer serum was carried in 2013 by Tonack et al. who compared the proteomic profiles of PDAC patients to benign pancreatic disease patients and healthy subjects (Tonack et al. 2013). This study is the first iTRAQ-based analysis of blood-borne proteins from pancreatic cancer patients and controls, including patients with chronic pancreatitis and benign biliary obstruction. Alongside this, LC-MS/MS analyses with and without iTRAQ labelling, for the analysis of pancreatic juice proteins was carried out as pancreatic juice is considered a rich source of potential biomarkers. An overlap of eighty-eight proteins were identified in both experiments and from this, inter-α-trypsin inhibitor heavy chain H3 (ITIH3), α1-β glycoprotein (A1BG) and complement C5 alpha chain (C5) were selected for validation. Similar to Brand et al., this study also recognised that a combined marker panel of C5, CA19-9 and A1BG significantly improved the specificity for PDAC patients compared with CA19-9 alone. Importantly, this analysis compared the performance of markers to distinguish PDAC in the presence or absence of jaundice over chronic pancreatitis. Markers were better at distinguishing PDAC from chronic pancreatitis when PDAC patients had jaundice. However, when jaundice was present in benign disease subjects, the specificity of markers for PDAC was adversely affected. These findings highlight the importance of calibrating markers according to the requirements of individual clinical scenarios. The use of iTRAQ in this study allowed for the simultaneous comparison of protein profiles in several samples and identified important targets as potential biomarkers for pancreatic cancer. The thoughtful selection of blood and pancreatic juice samples revealed candidate serum biomarkers however for some of these, performance was affected by the presence of jaundice. This conclusion emphasises the need to develop and understand the performance of biomarkers in patients with biliary obstruction given that the large proportion of PDAC patients present with jaundice.

The use of microarray technology, especially antibody arrays, is a valuable method for the detection of potential biomarkers and reproducible measurement of their levels over multiple patient samples. Antibody-based microarrays are a rapidly

emerging technique which are likely to have an increasing role in cancer proteomics for biomarker discovery and validation. Emerging platforms such as the Luminex immunoassays are an important technology which will aid both discovery and validation of serum biomarkers for pancreatic cancer (Brand et al. 2011). In 2008, Borrebaeck et al. carried out a large-scale antibody microarray composed of 129 recombinant antibody fragments directed against 60 serum proteins on 44 serum samples of patients with PDAC and healthy individuals. This analysis revealed a pancreatic cancer associated biomarker signature which could accurately classify samples as normal or cancer. In addition, another signature was identified which illustrated the predictive ability of microarrays to determine survival rates of patients with PDAC. In 2009, Haab et al. also employed the use of antibody microarrays to characterise both the protein and glycan levels of specific serum proteins across 23 pancreatic cancer patients and 23 control subjects (Yue et al. 2009). This analysis revealed both protein and glycan measurements at the same capture antibodies. MUC16 showed the strongest protein elevation but no glycan alterations, MUC5AC showing the most glycan alterations and MUC1 showing slight protein elevations with selected highly prevalent glycan alterations. The altered glycan levels were independent of protein elevations for MUC5AC and MUC1. Similarly, Remmers et al. found MUC1 as well as MUC4 to be differentially glycosylated as pancreatic cancer progressed from early intraepithelial neoplasias to metastatic disease when investigating 40 pancreatic tissue samples (Remmers et al. 2013). Thus, glycan measurements could potentially be used for highly effective pancreatic cancer biomarkers independent of protein expression. Gerdtsson et al. carried out a large-scale multicenter trial using microarray analysis of 338 patients (Gerdtsson et al. 2015). A multiplex biomarker signature of 10 serum proteins was identified with sensitivities and specificities in the 91-100% range. The results of this investigation and those aforementioned demonstrates the robustness of serum biomarker signatures which have the potential to be developed into routine protein microarray assays for patient screening.

As is often the case in many cancer studies but more so for pancreatic cancer research, the availability and accessibility to primary samples is often challenging and for many groups unfeasible. In 2011, Makawita et al. integrated proteomic profiling of

cell line conditioned media and pancreatic juice for the identification of pancreatic cancer biomarkers (Makawita et al. 2011). LC-MS/MS was employed to characterise the proteomes of conditioned media from six pancreatic cancer cell lines, the normal human pancreatic ductal epithelial cell line HPDE, and two pools of six pancreatic juice samples from ductal adenocarcinoma patients. Although this group only had access to six pancreatic juice samples the advantage of conditioned media from six pancreatic cancer cell lines strengthened the findings and identifications of this study. Approximately 76% (493 of 648) of proteins identified in the pancreatic juice samples were also identified in the cell line analysis, which underlines the similarity in the proteomes among these biological fluids. Of considerable note from this study is the ability of one of the identified proteins, AGR2, to supplement CA19-9 and raise its sensitivity, from approximately 90% to 100%, at 100% specificity. This once again emphasises that the growing consensus in this field is toward the development of panels of biomarkers, as the combined assessment of multiple molecules can result in increased sensitivity and specificity, in comparison to the assessment of proteins individually. Additionally, this study highlights the benefits of exploiting readily available immortal pancreatic cancer cell lines to reinforce findings from primary samples by increasing cohort size.

One of the most easily accessible bodily fluids is urine which presents a promising myriad of secreted proteins from the pancreas of PDAC patients. Non-invasive biomarkers for the detection, monitoring and therapeutic targeting of pancreatic adenocarcinoma could enable an inexpensive and non-intimidating method of screening patients. Radon et al. (2015) performed an in-depth proteomics analysis of urine samples from healthy controls, chronic pancreatitis and patients with PDAC (Radon et al. 2015). Proteomes of 18 urine samples were analysed using GeLC-MS/MS and selected biomarkers were subsequently validated using ELISA assays. The mass spectrometry data analysis resulted in the identification of approximately 1,500 proteins which underwent semi-quantitative assessment of relative protein abundance between urine sample groups. Three proteins determined to be commonly deregulated were lymphatic vessel endothelial hyaluronan receptor (LYVE1), regenerating islet-derived 1 alpha (REG1A) and trefoil factor 1 (TFF1). This three-biomarker urine panel performed well

at discriminating patients with early-stage PDAC from healthy subjects with high accuracy. Further analyses also suggest that when combined with CA19-9, the accuracy of this multi-marker panel may be increased to over 90%. Although the diagnostic performance of this panel now needs to be further validated, it has awoken great excitement in the pancreatic cancer field. Being completely non-invasive and inexpensive, this urinary multi-marker panel could potentially lead to an improved outcome in patients if used used as an initial screening test for high-risk groups of pancreatic adenocarcinomas.

Recent developments in the search for pancreatic cancer biomarkers are the detection and isolation of cancer derived exosomes from the serum of patients. A study in early 2015 by Costa-Silva et al. investigated the role of exosomes derived from malignant pancreatic lesions in preparing the liver for metastatic cancer cells to arrive in order to improve survival and promote engraftment (Costa-Silva et al. 2015). They found that pancreas derived exosomes are firstly selectively taken up by Kupffer cells in the liver which causes activation of the fibronectin pathway which promotes the arrest of macrophages and neutrophils in the liver thus dampening an immune response. This study shows knockdown of exosomal migration inhibitory factor (MIF) prevents all sequential steps in liver pre-metastatic niche formation. These observations by Costa-Silva et al. not only identify MIF as a potential marker for PDAC metastasis but also describe a previously unknown system by which PDAC-derived exosomes induce the development of liver pre-metastatic niches which drives metastatic disease progression. Melo et al. used mass spectrometry analyses to identify markers for pancreatic cancer derived exosomes and found glypican-1 to be specifically enriched on the cell surface of circulating exosomes (Melo et al. 2015). Quantification of glypican-1 positive circulating exosomes revealed a correlation between abundance and pancreatic tumour burden and thus patient survival. This study suggests glypican-1 positive circulating exosomes as a potential non-invasive diagnostic and screening tool for the detection of pancreatic cancer. The exploitation of tumour exosomes may be a reliable method for detection and monitoring of pancreatic cancer with an emphasis on early detection and metastasis potential. Exosomes are an attractive candidate for early detection and disease monitoring in the circulation of patients with PDAC.

2.5 Clinical Impact

Proteomic analyses offer great promise in the early detection of pancreatic cancer and also providing potential prognostic and predictive indicators of disease. The investigation of protein expression in certain samples at a specific time may help to uncover new tumour biomarkers for the early diagnosis of PDAC and develop personalised therapy. Current studies have focused on the search for differentially expressed proteins using conventional proteomic approaches, including iCAT and iTRAQ labelling, two-dimensional electrophoresis, multi-fractionated liquid chromatography and their combination with sophisticated mass spectrometry techniques. Proteomics technologies also have the capabilities to measure protein abundances, to characterize post-translational modifications (PTMs) and to study protein structure and interactions. The search for pancreatic cancer specific biomarkers with proteomics has major potential to improve early detection, prognosis, diagnosis, treatment selection and disease monitoring. Overexpressed proteins associated with PDAC also have a potential use for tumour detection by ultrasound molecular imaging. The investigation of thymocyte differentiation antigen 1 (Thy1 or CD90) revealed its potential as a target for PDAC neovasculature imaging when tested in transgenic mice that mimic human PDAC (Foygel et al. 2013). This study shows substantially increased ultrasonic imaging signal in tumours developed in a murine model compared to control conditions.

Although proteomic studies of pancreatic cancer are promising and in abundance, only a limited number of protein candidates have ultimately been translated into clinical trials or a clinical setting. The main problem is that most proposed candidates are unlikely to meet the tough criteria necessary for clinical validation. Difficulties arise for clinical proteomics in relation to the nature and quality of clinical samples, experimental design and the complexity of the human proteome (Maes et al. 2015). These challenges often result in the discovery of what seem like promising candidates however when tested in larger cohorts or with a greater stringency, candidates often fail to show the same specificity or sensitivity. Many identified candidate biomarkers which have failed to be approved or clinically advanced, have

been overlooked for their use in monitoring disease progression post therapy, predicting response to treatments, stratifying patients for specific therapies or as novel drug targets. The majority of proteomic studies in pancreatic cancer have focused on biomarker discovery however proteins identified from such studies could also have great potential as a therapy target for an antibody-drug conjugate (ADC). Tumour-targeted delivery of cytotoxic drugs presents several advantages compared to conventional chemotherapies with greater selectivity and less toxicity and harmful effects caused to patients. The development of ADCs has been very active in recent years, as of 2013, 30 ADCs against over 24 targets were in trial for blood cancers and solid tumours (Mullard 2013). A major challenge in the development of ADCs is the limited amount of drugs that can be joined to a single antibody, as overloading the antibody may reduce the binding affinity of the antibody, affect pharmacokinetics or lead to premature release of the cytotoxic payload before reaching the tumour (Sauerborn and van Dongen 2014). The more recent introduction of drug-loaded nanoparticles which contain antibodies integrated to the surface offer the same tumour specificity as ADCs with fewer risks. The underlying principal of action for both therapies remains the same, relying on the antibody specificity to target tumour-restricted proteins for efficient drug delivery. Thus, the application of proteomics for novel drug therapies in pancreatic cancer may lead to encouraging discoveries for better treatments against this malignancy.

2.6 Expert commentary

Serum CA 19-9 remains the only tumour marker approved by the FDA for use in pancreatic cancer, it has significant value as a prognostic marker and as a measurer of disease progression which can ultimately determine treatment decisions. New research studies and clinical strategies focus on the establishment of multi-biomarker panels for the early detection of pancreatic adenocarcinoma which show increased specificity and selectivity than CA 19-9 alone. Ongoing and future studies focused on identifying biomarkers in patients should employ this approach to improve diagnosis which remains a priority for the field. Developments in proteomics have had a profound influence on target identification and continue to rapidly advance its technologies. This may

ultimately lead to complete proteomic profiling of individual patients in the future in an attempt to stratify specific patient groups towards the best treatment for them.

2.7 Five-year view

Given the absence of approved biomarkers and effective therapies for pancreatic ductal adenocarcinoma, emerging multi-biomarker panels and targeted therapies will likely be important for the detection and treatment of this disease. Proteomic profiling of patients may become a standard part of the diagnostic stages of this malignancy. As an alternative to treating all patients of PDAC with a standard therapy, small selected cohorts of patients will be given targeted personalised therapy. Proteomic signatures associated with aggressive and metastatic PDAC will emerge and allow for staging schemes to be introduced. Future insights and findings will significantly improve on the outcome currently given to patients of pancreatic cancer.

2.8 Key Issues

- Pancreatic ductal adenocarcinoma is the fourth leading cause of cancer-related death in the world. Patients are generally given a 5-year survival rate of less than 5% and a median survival rate of 3 to 6 months for patients with metastatic pancreatic cancer.
- Current treatments for this disease modestly improve the prognosis of patients however, despite advances in therapeutics and clinical trials there has been no significant improvement in overall survival rates of patients.
- Serum CA19-9, the only FDA approved biomarker for PDAC in clinical use, often gives false-positive results in patients with other benign pancreatic diseases. It also suffers from low-sensitivity which hampers its use as a diagnostic marker for this malignancy.
- Despite identification of copious potential biomarkers in research, there has been a lack of translation of such candidates for clinical validation. Many promising candidates fail to meet the strict criteria necessary for advancing a target for validation in large cohorts within clinical trials.

- Many identified candidate biomarkers which have failed to be approved or clinically advanced, have been overlooked for their use in monitoring disease progression post therapy, predicting response to treatments, stratifying patients for specific therapies or as novel drug targets for emerging therapies such as ADCs.

2.9 References

Adam, B.L., Vlahou, A., Semmes, O.J., Jr, G.L.W., 2001. Proteomic approaches to biomarker discovery in prostate and bladder cancers. Proteomics 1, 1264–1270. https://doi.org/10.1002/1615-9861(200110)1:10<1264::AID-PROT1264>3.0.CO;2-R

Ansari, D., Aronsson, L., Sasor, A., Welinder, C., Rezeli, M., Marko-Varga, G., Andersson, R., 2014. The role of quantitative mass spectrometry in the discovery of pancreatic cancer biomarkers for translational science. J. Transl. Med. 12, 87-5876-12–87. https://doi.org/10.1186/1479-5876-12-87

Bausch, D., Thomas, S., Mino-Kenudson, M., Fernandez-del, C.C., Bauer, T.W., Williams, M., Warshaw, A.L., Thayer, S.P., Kelly, K.A., 2011. Plectin-1 as a novel biomarker for pancreatic cancer. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. 17, 302–309. https://doi.org/10.1158/1078-0432.CCR-10-0999

Berberat, P.O., Friess, H., Wang, L., Zhu, Z., Bley, T., Frigeri, L., Zimmermann, A., Buchler, M.W., 2001. Comparative analysis of galectins in primary tumors and tumor metastasis in human pancreatic cancer. J. Histochem. Cytochem. Off. J. Histochem. Soc. 49, 539–549.

Boja, E.S., Rodriguez, H., 2011. The path to clinical proteomics research: integration of proteomics, genomics, clinical laboratory and regulatory science. Korean J. Lab. Med. 31, 61–71.

Boone, B.A., Steve, J., Krasinskas, A.M., Zureikat, A.H., Lembersky, B.C., Gibson, M.K., Stoller, R.G., Zeh, H.J., Bahary, N., 2013. Outcomes with FOLFIRINOX for borderline resectable and locally unresectable pancreatic cancer. J. Surg. Oncol. 108, 236–241.

Bosetti, C., Lucenteforte, E., Silverman, D.T., Petersen, G., Bracci, P.M., Ji, B.T., Negri, E., Li, D., Risch, H.A., Olson, S.H., Gallinger, S., Miller, A.B., Bueno-de-Mesquita, H.B., Talamini, R., Polesel, J., Ghadirian, P., Baghurst, P.A., Zatonski, W., Fontham, E., Bamlet, W.R., Holly, E.A., Bertuccio, P., Gao, Y.T., Hassan, M., Yu, H., Kurtz, R.C., Cotterchio, M., Su, J., Maisonneuve, P., Duell, E.J., Boffetta, P., Vecchia, C.L., 2012. Cigarette smoking and pancreatic cancer: an analysis from the International Pancreatic Cancer Case-Control Consortium (Panc4). Ann. Oncol. Off. J. Eur. Soc. Med. Oncol. ESMO 23, 1880–1888. https://doi.org/10.1093/annonc/mdr541 [doi]

Brand, R.E., Nolen, B.M., Zeh, H.J., Allen, P.J., Eloubeidi, M.A., Goldberg, M., Elton, E., Arnoletti, J.P., Christein, J.D., Vickers, S.M., Langmead, C.J., Landsittel, D.P., Whitcomb, D.C., Grizzle, W.E., Lokshin, A.E., 2011. Serum biomarker panels for the detection of pancreatic cancer. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. 17, 805–816. https://doi.org/10.1158/1078-0432.CCR-10-0248 [doi]

Britton, D., Zen, Y., Quaglia, A., Selzer, S., Mitra, V., Lößner, C., Jung, S., Böhm, G., Schmid, P., Prefot, P., 2014. Quantification of pancreatic cancer proteome and phosphorylome: indicates molecular events likely contributing to cancer and activity of drug targets. PloS One 9.

Cellulaire, B., 2002. Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains. Proteomics 2, 3–10.

Chen, J., Wu, W., Tang, H.K., Zheng, C.S., Xia, Y.L., Zhou, H.C., Yang, R.B., Chen, L.J., Hu, L.W., 2013. Analysis of pancreatic cancer peripheral blood by comparative proteomics. Zhonghua Wai Ke Za Zhi 51, 62–65.

Chen, J.-H., Ni, R.-Z., Xiao, M.-B., Guo, J.-G., Zhou, J.-W., 2009. Comparative proteomic analysis of differentially expressed proteins in human pancreatic cancer tissue.

Chen, R., Eugene, C.Y., Donohoe, S., Pan, S., Eng, J., Cooke, K., Crispin, D.A., Lane, Z., Goodlett, D.R., Bronner, M.P., 2005. Pancreatic cancer proteome: the proteins that underlie invasion, metastasis, and immunologic escape. Gastroenterology 129, 1187–1197.

Chue, B.M., 2008. Pancreatic cancer.

Ciliberto, D., Botta, C., Correale, P., Rossi, M., Caraglia, M., Tassone, P., Tagliaferri, P., 2013. Role of gemcitabine-based combination therapy in the management of advanced pancreatic cancer: a meta-analysis of randomised trials. Eur. J. Cancer 49, 593–603.

Conroy, T., Desseigne, F., Ychou, M., Bouché, O., Guimbaud, R., Bécouarn, Y., Adenis, A., Raoul, J.-L., Gourgou-Bourgade, S., Fouchardière, C. de la, 2011. FOLFIRINOX versus gemcitabine for metastatic pancreatic cancer. N. Engl. J. Med. 364, 1817–1825.

Costa-Silva, B., Aiello, N.M., Ocean, A.J., Singh, S., Zhang, H., Thakur, B.K., Becker, A., Hoshino, A., Mark, M.T., Molina, H., 2015. Pancreatic cancer exosomes initiate pre-metastatic niche formation in the liver. Nat. Cell Biol. 17, 816–826.

Costello, E., Greenhalf, W., Neoptolemos, J.P., 2012. New biomarkers and targets in pancreatic cancer and their application to treatment. Nat. Rev. Gastroenterol. Hepatol. 9, 435–444.

Cowgill, S.M., Muscarella, P., 2003. The genetics of pancreatic cancer. Am. J. Surg. 186, 279–286.

Deer, E.L., Gonzalez-Hernandez, J., Coursen, J.D., Shea, J.E., Ngatia, J., Scaife, C.L., Firpo, M.A., Mulvihill, S.J., 2010. Phenotype and genotype of pancreatic cancer cell lines. Pancreas 39, 425–435. https://doi.org/10.1097/MPA.0b013e3181c15963 [doi]

Falchook, G.S., Ordonez, N.G., Bastida, C.C., Stephens, P.J., Miller, V.A., Gaido, L., Jackson, T., Karp, D.D., 2014. Effect of the RET Inhibitor Vandetanib in a Patient With RET Fusion-Positive Metastatic Non-Small-Cell Lung Cancer. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. https://doi.org/JCO.2013.50.5016 [pii]

Fernández-Medarde, A., Santos, E., 2011. Ras in cancer and developmental diseases. Genes Cancer 2, 344–358.

Foygel, K., Wang, H., Machtaler, S., Lutz, A.M., Chen, R., Pysz, M., Lowe, A.W., Tian, L., Carrigan, T., Brentnall, T.A., 2013. Detection of pancreatic ductal adenocarcinoma

in mice by ultrasound imaging of thymocyte differentiation antigen 1. Gastroenterology 145, 885-894. e3.

Gerdtsson, A.S., Malats, N., Säll, A., Real, F.X., Porta, M., Skoog, P., Persson, H., Wingren, C., Borrebaeck, C.A., 2015. A Multicenter Trial Defining a Serum Protein Signature Associated with Pancreatic Ductal Adenocarcinoma. Int. J. Proteomics 2015.

Gronborg, M., Kristiansen, T.Z., Iwahori, A., Chang, R., Reddy, R., Sato, N., Molina, H., Jensen, O.N., Hruban, R.H., Goggins, M.G., Maitra, A., Pandey, A., 2006. Biomarker discovery from pancreatic cancer secretome using a differential proteomic approach. Mol. Cell. Proteomics MCP 5, 157–171. https://doi.org/M500178-MCP200 [pii]

Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., Aebersold, R., 1999. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat. Biotechnol. 17, 994–999.

Harsha, H.C., Kandasamy, K., Ranganathan, P., Rani, S., Ramabadran, S., Gollapudi, S., Balakrishnan, L., Dwivedi, S.B., Telikicherla, D., Selvan, L.D.N., 2009. A compendium of potential biomarkers of pancreatic cancer. PLoS Med. 6, 343.

Hocker, J.R., Postier, R.G., Li, M., Lerner, M.R., Lightfoot, S.A., Peyton, M.D., Deb, S.J., Baker, C.M., Williams, T.L., Hanas, R.J., 2015. Discriminating patients with early-stage pancreatic cancer or chronic pancreatitis using serum electrospray mass profiling. Cancer Lett. 359, 314–324.

Hu, S., Arellano, M., Boontheung, P., Wang, J., Zhou, H., Jiang, J., Elashoff, D., Wei, R., Loo, J.A., Wong, D.T., 2008. Salivary Proteomics for Oral Cancer Biomarker Discovery. Clin. Cancer Res. 14, 6246–6252. https://doi.org/10.1158/1078-0432.CCR-07-5037

Hurwitz, H.I., Uppal, N., Wagner, S.A., Bendell, J.C., Beck, J.T., Wade 3rd, S.M., Nemunaitis, J.J., Stella, P.J., Pipas, J.M., Wainberg, Z.A., Manges, R., Garrett, W.M., Hunter, D.S., Clark, J., Leopold, L., Sandor, V., Levy, R.S., 2015. Randomized, Double-Blind, Phase II Study of Ruxolitinib or Placebo in Combination With Capecitabine in

Patients With Metastatic Pancreatic Cancer for Whom Therapy With Gemcitabine Has Failed. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. 33, 4039–4047. https://doi.org/10.1200/JCO.2015.61.4578 [doi]

Jenkinson, C., Elliott, V., Evans, A., Oldfield, L., Jenkins, R.E., O'Brien, D.P., Apostolidou, S., Gentry-Maharaj, A., Fourkala, E.O., Jacobs, I., 2015. Decreased serum thrombospondin-1 levels in pancreatic cancer patients up to 24 months prior to clinical diagnosis: association with diabetes mellitus. Clin. Cancer Res. clincanres. 0879.2015.

Kelly, K.A., Bardeesy, N., Anbazhagan, R., Gurumurthy, S., Berger, J., Alencar, H., DePinho, R.A., Mahmood, U., Weissleder, R., 2008. Targeted nanoparticles for imaging incipient pancreatic ductal adenocarcinoma. PLoS Med 5, e85.

Kim, M.S., Zhong, Y., Yachida, S., Rajeshkumar, N.V., Abel, M.L., Marimuthu, A., Mudgal, K., Hruban, R.H., Poling, J.S., Tyner, J.W., Maitra, A., Iacobuzio-Donahue, C.A., Pandey, A., 2014. Heterogeneity of pancreatic cancer metastases in a single patient revealed by quantitative proteomics. Mol. Cell. Proteomics MCP 13, 2803–2811. https://doi.org/10.1074/mcp.M114.038547 [doi]

Kim, Y., Han, D., Min, H., Jin, J., Yi, E.C., Kim, Y., 2014. Comparative proteomic profiling of pancreatic ductal adenocarcinoma cell lines. Mol. Cells 37, 888–898. https://doi.org/10.14348/molcells.2014.0207 [doi]

Koorstra, J.-B.M., Feldmann, G., Habbe, N., Maitra, A., 2008. Morphogenesis of pancreatic cancer: role of pancreatic intraepithelial neoplasia (PanINs). Langenbecks Arch. Surg. 393, 561–570.

Kosanam, H., Prassas, I., Chrystoja, C.C., Soleas, I., Chan, A., Dimitromanolakis, A., Blasutig, I.M., Ruckert, F., Gruetzmann, R., Pilarsky, C., Maekawa, M., Brand, R., Diamandis, E.P., 2013. Laminin, gamma 2 (LAMC2): a promising new putative pancreatic cancer biomarker identified by proteomic analysis of pancreatic adenocarcinoma tissues. Mol. Cell. Proteomics MCP 12, 2820–2832. https://doi.org/10.1074/mcp.M112.023507 [doi]

LaConti, J.J., Laiakis, E.C., Mays, A.D., Peran, I., Kim, S.E., Shay, J.W., Riegel, A.T., Jr, A.J.F., Wellstein, A., 2015. Distinct serum metabolomics profiles associated with malignant progression in the KrasG12D mouse model of pancreatic ductal adenocarcinoma. BMC Genomics 16 Suppl 1, S1-2164-16-S1-S1. Epub 2015 Jan 15. https://doi.org/10.1186/1471-2164-16-S1-S1 [doi]

Lee, J.J., Huang, J., England, C.G., McNally, L.R., Frieboes, H.B., 2013. Predictive modeling of in vivo response to gemcitabine in pancreatic cancer. PLoS Comput. Biol. 9, e1003231.

Li, D., Xie, K., Wolff, R., Abbruzzese, J.L., 2004. Pancreatic cancer. The Lancet 363, 1049–1057.

Lin, L., Lin, H., Zhang, M., Dong, X., Yin, X., Qu, C., Ni, J., 2015. Types, principle, and characteristics of tandem high-resolution mass spectrometry and its applications. RSC Adv. 5, 107623–107636.

Lowenfels, A.B., Maisonneuve, P., 2006. Epidemiology and risk factors for pancreatic cancer. Best Pract. Res. Clin. Gastroenterol. 20, 197–209.

Maes, E., Mertens, I., Valkenborg, D., Pauwels, P., Rolfo, C., Baggerman, G., 2015. Proteomics in cancer research: Are we ready for clinical practice? Crit. Rev. Oncol. Hematol. 96, 437–448.

Maestranzi, S., Przemioslo, R., Mitchell, H., Sherwood, R.A., 1998. The effect of benign and malignant liver disease on the tumour markers CA19-9 and CEA. Ann. Clin. Biochem. Int. J. Biochem. Med. 35, 99–103.

Makawita, S., Smith, C., Batruch, I., Zheng, Y., Ruckert, F., Grutzmann, R., Pilarsky, C., Gallinger, S., Diamandis, E.P., 2011. Integrated proteomic profiling of cell line conditioned media and pancreatic juice for the identification of pancreatic cancer biomarkers. Mol. Cell. Proteomics MCP 10, M111.008599. https://doi.org/10.1074/mcp.M111.008599 [doi]

Martinez-Bosch, N., Fernandez-Barrena, M.G., Moreno, M., Ortiz-Zapater, E., Munne-Collado, J., Iglesias, M., Andre, S., Gabius, H.J., Hwang, R.F., Poirier, F., Navas, C.,

Guerra, C., Fernandez-Zapico, M.E., Navarro, P., 2014. Galectin-1 drives pancreatic carcinogenesis through stroma remodeling and Hedgehog signaling activation. Cancer Res. 74, 3512–3524. https://doi.org/10.1158/0008-5472.CAN-13-3013 [doi]

McKinney, K.Q., Lee, J.G., Sindram, D., Russo, M.W., Han, D.K., Bonkovsky, H.L., Hwang, S.I., 2012. Identification of differentially expressed proteins from primary versus metastatic pancreatic cancer cells using subcellular proteomics. Cancer Genomics Proteomics 9, 257–263. https://doi.org/9/5/257 [pii]

Melo, S.A., Luecke, L.B., Kahlert, C., Fernandez, A.F., Gammon, S.T., Kaye, J., LeBleu, V.S., Mittendorf, E.A., Weitz, J., Rahbari, N., 2015. Glypican-1 identifies cancer exosomes and detects early pancreatic cancer. Nature.

Moore, M.J., Goldstein, D., Hamm, J., Figer, A., Hecht, J.R., Gallinger, S., Au, H.J., Murawa, P., Walde, D., Wolff, R.A., Campos, D., Lim, R., Ding, K., Clark, G., Voskoglou-Nomikos, T., Ptasynski, M., Parulekar, W., Group, N.C.I. of C.C.T., 2007. Erlotinib plus gemcitabine compared with gemcitabine alone in patients with advanced pancreatic cancer: a phase III trial of the National Cancer Institute of Canada Clinical Trials Group. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. 25, 1960–1966. https://doi.org/JCO.2006.07.9525 [pii]

Mullard, A., 2013. Maturing antibody-drug conjugate pipeline hits 30. Nat. Rev. Drug Discov. 12, 329–332.

Mullen, W., Albalat, A., Gonzalez, J., Zerefos, P., Siwy, J., Franke, J., Mischak, H., 2012. Performance of different separation methods interfaced in the same MS-reflection TOF detector: A comparison of performance between CE versus HPLC for biomarker analysis. Electrophoresis 33, 567–574.

Nie, S., Lo, A., Wu, J., Zhu, J., Tan, Z., Simeone, D.M., Anderson, M.A., Shedden, K.A., Ruffin, M.T., Lubman, D.M., 2014. Glycoprotein biomarker panel for pancreatic cancer discovered by quantitative proteomics analysis. J. Proteome Res. 13, 1873–1884.

Oettle, H., Neuhaus, P., Hochhaus, A., Hartmann, J.T., Gellert, K., Ridwelski, K., Niedergethmann, M., Zülke, C., Fahlke, J., Arning, M.B., 2013. Adjuvant chemotherapy

with gemcitabine and long-term outcomes among patients with resected pancreatic cancer: the CONKO-001 randomized trial. Jama 310, 1473–1481.

Ohmine, K., Kawaguchi, K., Ohtsuki, S., Motoi, F., Ohtsuka, H., Kamiie, J., Abe, T., Unno, M., Terasaki, T., 2015. Quantitative Targeted Proteomics of Pancreatic Cancer: Deoxycytidine kinase protein level correlates to progression-free survival of patients receiving gemcitabine treatment. Mol. Pharm. 12, 3282–3291.

Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., Mann, M., 2002. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol. Cell. Proteomics MCP 1, 376–386.

Pan, S., Brentnall, T.A., Kelly, K., Chen, R., 2013. Tissue proteomics in pancreatic cancer study: discovery, emerging technologies, and challenges. Proteomics 13, 710–721.

Pan, S., Chen, R., Crispin, D.A., May, D., Stevens, T., McIntosh, M.W., Bronner, M.P., Ziogas, A., Anton-Culver, H., Brentnall, T.A., 2011. Protein alterations associated with pancreatic cancer and chronic pancreatitis found in human plasma using global quantitative proteomics profiling. J. Proteome Res. 10, 2359–2376.

Phillips, P., Grippo, P.J., Munshi, H.G., 2012. Pancretic cancer and tumor microenvironment. Pancreat. Stellate Cells Fibros.

Pontén, F., Schwenk, J.M., Asplund, A., Edqvist, P.-H., 2011. The Human Protein Atlas as a proteomic resource for biomarker discovery. J. Intern. Med. 270, 428–446.

Radon, T.P., Massat, N.J., Jones, R., Alrawashdeh, W., Dumartin, L., Ennis, D., Duffy, S.W., Kocher, H.M., Pereira, S.P., Guarner (posthumous), L., Murta-Nascimento, C., Real, F.X., Malats, N., Neoptolemos, J., Costello, E., Greenhalf, W., Lemoine, N.R., Crnogorac-Jurcevic, T., 2015. Identification of a Three-Biomarker Panel in Urine for Early Detection of Pancreatic Adenocarcinoma. Clin. Cancer Res. 21, 3512–3521. https://doi.org/10.1158/1078-0432.CCR-14-2467

Remmers, N., Anderson, J.M., Linde, E.M., DiMaio, D.J., Lazenby, A.J., Wandall, H.H., Mandel, U., Clausen, H., Yu, F., Hollingsworth, M.A., 2013. Aberrant expression of mucin core proteins and o-linked glycans associated with progression of pancreatic cancer. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. 19, 1981–1993. https://doi.org/10.1158/1078-0432.CCR-12-2662 [doi]

Rozenblum, E., Schutte, M., Goggins, M., Hahn, S.A., Panzer, S., Zahurak, M., Goodman, S.N., Sohn, T.A., Hruban, R.H., Yeo, C.J., Kern, S.E., 1997. Tumor-suppressive pathways in pancreatic carcinoma. Cancer Res. 57, 1731–1734.

Ryan, D.P., Hong, T.S., Bardeesy, N., 2014. Pancreatic adenocarcinoma. N. Engl. J. Med. 371, 1039–1049.

Samulitis, B.K., Pond, K.W., Pond, E., Cress, A.E., Patel, H., Wisner, L., Patel, C., Dorr, R.T., Landowski, T.H., 2015. Gemcitabine resistant pancreatic cancer cell lines acquire an invasive phenotype with collateral hypersensitivity to histone deacetylase inhibitors. Cancer Biol. Ther. 16, 43–51.

Sauerborn, M., Dongen, W. van, 2014. Practical Considerations for the Pharmacokinetic and Immunogenic Assessment of Antibody–Drug Conjugates. BioDrugs 28, 383–391.

Scheltema, R.A., Hauschild, J.P., Lange, O., Hornburg, D., Denisov, E., Damoc, E., Kuehn, A., Makarov, A., Mann, M., 2014. The Q Exactive HF, a Benchtop mass spectrometer with a pre-filter, high-performance quadrupole and an ultra-high-field Orbitrap analyzer. Mol. Cell. Proteomics MCP 13, 3698–3708. https://doi.org/10.1074/mcp.M114.043489 [doi]

Schniewind, B., Christgen, M., Kurdow, R., Haye, S., Kremer, B., Kalthoff, H., Ungefroren, H., 2004. Resistance of pancreatic cancer to gemcitabine treatment is dependent on mitochondria-mediated apoptosis. Int. J. Cancer 109, 182–188.

Schutte, M., Hruban, R.H., Geradts, J., Maynard, R., Hilgers, W., Rabindran, S.K., Moskaluk, C.A., Hahn, S.A., Schwarte-Waldhoff, I., Schmiegel, W., Baylin, S.B., Kern, S.E., Herman, J.G., 1997. Abrogation of the Rb/p16 tumor-suppressive pathway in virtually all pancreatic carcinomas. Cancer Res. 57, 3126–3130.

Shields, J.M., Pruitt, K., McFall, A., Shaub, A., Der, C.J., 2000. Understanding Ras:'it ain't over'til it's over.' Trends Cell Biol. 10, 147–154.

Singh, C.K., Kaur, S., George, J., Nihal, M., Hahn, M.C.P., Scarlett, C.O., Ahmad, N., 2015. Molecular signatures of sanguinarine in human pancreatic cancer cells: A large scale label-free comparative proteomics approach. Oncotarget 6, 10335–10348. https://doi.org/3231 [pii]

Surlin, V., Bintintan, V., Petrariu, F.D., Dobrin, R., Lefter, R., Ciobica, A., Timofte, D., 2014. Prognostic factors in resectable pancreatic cancer. Rev. Med. Chir. Soc. Med. Nat. Iasi 118, 924–931.

Takadate, T., Onogawa, T., Fukuda, T., Motoi, F., Suzuki, T., Fujii, K., Kihara, M., Mikami, S., Bando, Y., Maeda, S., 2013. Novel prognostic protein markers of resectable pancreatic cancer identified by coupled shotgun and targeted proteomics using formalin-fixed paraffin-embedded tissues. Int. J. Cancer 132, 1368–1382.

Takahashi, Y., Takeuchi, T., Sakamoto, J., Touge, T., Mai, M., Ohkura, H., Kodaira, S., Okajima, K., Nakazato, H., Committee, T.M., 2003. The usefulness of CEA and/or CA19-9 in monitoring for recurrence in gastric cancer patients: a prospective clinical study. Gastric Cancer 6, 142–145.

Theodoridis, G.A., Gika, H.G., Want, E.J., Wilson, I.D., 2012. Liquid chromatography–mass spectrometry based global metabolite profiling: a review. Anal. Chim. Acta 711, 7–16.

Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Hamon, C., 2003. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal. Chem. 75, 1895–1904.

Thompson, C.C., Ashcroft, F.J., Patel, S., Saraga, G., Vimalachandran, D., Prime, W., Campbell, F., Dodson, A., Jenkins, R.E., Lemoine, N.R., Crnogorac-Jurcevic, T., Yin, H.L., Costello, E., 2007. Pancreatic cancer cells overexpress gelsolin family-capping proteins, which contribute to their cell motility. Gut 56, 95–106. https://doi.org/gut.2005.083691 [pii]

Tonack, S., Jenkinson, C., Cox, T., Elliott, V., Jenkins, R.E., Kitteringham, N.R., Greenhalf, W., Shaw, V., Michalski, C.W., Friess, H., 2013. iTRAQ reveals candidate pancreatic cancer serum biomarkers: influence of obstructive jaundice on their performance. Br. J. Cancer 108, 1846–1853.

Tu, C., Rudnick, P.A., Martinez, M.Y., Cheek, K.L., Stein, S.E., Slebos, R.J., Liebler, D.C., 2010. Depletion of abundant plasma proteins and limitations of plasma proteomics. J. Proteome Res. 9, 4982–4991.

Vincent, M., A...Herman, J...Schulick, R...Hruban, R.H. an. Goggins, 2013. Pancreatic Cancer. Lancet 378(9791):607–20.

Von Hoff, D.D., Ervin, T., Arena, F.P., Chiorean, E.G., Infante, J., Moore, M., Seay, T., Tjulandin, S.A., Ma, W.W., Saleh, M.N., Harris, M., Reni, M., Dowden, S., Laheru, D., Bahary, N., Ramanathan, R.K., Tabernero, J., Hidalgo, M., Goldstein, D., Van Cutsem, E., Wei, X., Iglesias, J., Renschler, M.F., 2013. Increased Survival in Pancreatic Cancer with nab-Paclitaxel plus Gemcitabine [WWW Document]. http://dx.doi.org/10.1056/NEJMoa1304369. https://doi.org/10.1056/NEJMoa1304369

Wang, W., Zhou, H., Lin, H., Roy, S., Shaler, T.A., Hill, L.R., Norton, S., Kumar, P., Anderle, M., Becker, C.H., 2003. Quantification of Proteins and Metabolites by Mass Spectrometry without Isotopic Labeling or Spiked Standards. Anal. Chem. 75, 4818–4826. https://doi.org/10.1021/ac026468x

Wang, W.B., Zhao, Y.P., Liao, Q., Zhang, T.P., Xu, L., Wu, Y.D., 2012. Validation of candidate immunogenic membrane antigens of human pancreatic cancer screened by proteomics. Zhonghua Wai Ke Za Zhi 50, 260–263.

Whipple, A.O., Parsons, W.B., Mullins, C.R., 1935. Treatment of Carcinoma of the Ampulla of Vater. Ann. Surg. 102, 763–779.

Wiese, S., Reidegeld, K.A., Meyer, H.E., Warscheid, B., 2007. Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. Proteomics 7, 340–350.

Winter, J.M., Brennan, M.F., Tang, L.H., D'Angelica, M.I., DeMatteo, R.P., Fong, Y., Klimstra, D.S., Jarnagin, W.R., Allen, P.J., 2012. Survival after resection of pancreatic adenocarcinoma: results from a single institution over three decades. Ann. Surg. Oncol. 19, 169–175.

Wit, M. de, Fijneman, R.J., Verheul, H.M., Meijer, G.A., Jimenez, C.R., 2013. Proteomics in colorectal cancer translational research: biomarker discovery for clinical applications. Clin. Biochem. 46, 466–479.

Wittmann-Liebold, B., Graack, H.-R., Pohl, T., 2006. Two-dimensional gel electrophoresis as tool for proteomics studies in combination with protein identification by mass spectrometry. Proteomics 6, 4688–4703.

Yoneyama, T., Ohtsuki, S., Ono, M., Ohmine, K., Uchida, Y., Yamada, T., Tachikawa, M., Terasaki, T., 2013. Quantitative Targeted Absolute Proteomics-Based Large-Scale Quantification of Proline-Hydroxylated α-Fibrinogen in Plasma for Pancreatic Cancer Diagnosis. J. Proteome Res. 12, 753–762.

Yue, T., Goldstein, I.J., Hollingsworth, M.A., Kaul, K., Brand, R.E., Haab, B.B., 2009. The Prevalence and Nature of Glycan Alterations on Specific Proteins in Pancreatic Cancer Patients Revealed Using Antibody-Lectin Sandwich Arrays. Mol. Cell. Proteomics 8, 1697–1707. https://doi.org/10.1074/mcp.M900135-MCP200

Zhu, W., Smith, J.W., Huang, C.M., 2010. Mass spectrometry-based label-free quantitative proteomics. J. Biomed. Biotechnol. 2010, 840518. https://doi.org/10.1155/2010/840518 [doi]

3 Chapter Three: A comparative quantitative LC-MS/MS profiling analysis of human pancreatic adenocarcinoma, adjacent-normal tissue, and patient-derived tumour xenografts

Published in *Proteomes* November 2018. Invited submission for the special issue "Pharmaceutical proteomics"

DOI: 10.3390/proteomes6040045

Authors: Orla Coleman, Michael Henry, Fiona O'Neill, Sandra Roche, Niall Swan, Lorraine Boyle, Jean Murphy, Justine Meiller, Neil T. Conlon, Justin Geoghegan, Kevin Conlon, Vincent Lynch, Ninfa L. Straubinger, Robert M. Straubinger, Gerard McVey, Michael Moriarty, Paula Meleady and Martin Clynes

For this publication, I performed the sample preparation including membrane protein enrichment, protein digestion and preparation for MS analysis for all 40 samples analysed. I was responsible for all bioinformatic data analyses. The method to differentiate human-specific proteins in the PDX samples was developed by me. I validated protein expression using Western blot. I was invited to submit to the special issue "Pharmaceutical proteomics" in the journal *Proteomes*. I was responsible for manuscript writing, preparation, submission and revisions.

Abstract

Pancreatic ductal adenocarcinoma (PDAC) is one of the deadliest cancers worldwide; it develops in a relatively symptom-free manner leading to rapid disease progression and metastasis leading to a 5-year survival rate of less than 5%. A lack of dependable diagnostic markers and rapid development of resistance to conventional therapies are among the problems associated with management of the disease. A better understanding of pancreatic tumour biology and discovery of new potential therapeutic targets are important goals in pancreatic cancer research. This study describes the comparative quantitative LC-MS/MS proteomic analysis of the membrane-enriched proteome of 10 human pancreatic ductal adenocarcinomas, 9 matched adjacent-normal pancreas and patient derived xenografts (PDX) in mice (10 at F1 generation and 10 F2). Quantitative label-free LC-MS/MS data analysis identified 129 proteins upregulated and 109 downregulated in PDAC compared to adjacent-normal tissue. In this study, analysing peptide MS/MS data from the xenografts, great care was taken to distinguish species specific peptides definitively derived from human sequences or, from murine, and which could not be distinguished. The human-only peptides from the PDXs is of particular value, since only human tumour cells survive, and stromal cells are replaced during engraftment in the mouse; this list is therefore enriched in tumour-associated proteins some of which might be potential therapeutic or diagnostic targets. Using human-specific sequences 32 proteins were found to be upregulated and 113 downregulated in PDX F1 tumours compared to primary PDAC. Differential expression of CD55 between PDAC and normal pancreas, and expression across PDX generations, was confirmed by Western blotting. This data indicates the value of using PDX models in PDAC research. This study is the first comparative proteomic analysis of PDAC which employs PDX models to identify patient tumour cell-associated proteins in an effort to find robust targets for therapeutic treatment of PDAC.

3.1 Introduction

Pancreatic ductal adenocarcinoma (PDAC), which constitutes 90% of pancreatic cancers, is one of the most lethal solid malignancies and the fourth leading cause of cancer-related mortality in the world. Despite the improvement in screening and therapies of many solid tumours, pancreatic cancer prognosis remains dismal with 91% of patients not surviving within five years of diagnosis and only 26% of patients surviving within the first year of diagnosis (www.pancreatic.org). The absence of symptoms in its initial stages and insufficient early detection tools lead to rapid disease progression, distant metastases and poor prognoses; it is estimated that only 10% of presented patients have tumours that are potentially curable by resection (Balaban, Mangu and Yee 2016, O'lorcain, Deady and Comber 2006) (www.ncri.ie). For most of the cases this cancer is advanced, unresectable and metastatic, and for the minority of patients who are diagnosed at an early stage, conventional cancer treatments have limited benefit and little impact on disease progression.

Since 1997, gemcitabine therapy has been the standard first-line treatment for patients with unresectable locally advanced or metastatic pancreatic cancer (Burris et al. 1997). Most recently, a phase III trial of adjuvant modified FOLFIRINOX significantly improved DFS, MFS and OS compared to Gemcitabine among patients with surgically removed PDAC and is now being considered a new standard of care for pancreatic cancer after resection (Conroy et al. 2018). Monoclonal antibodies have proved to have a successful role in cancer therapy, specifically ADC therapy is emerging as a chemotherapeutic with great promise for solid cancer treatment (Lambert and Morris 2017). With the introduction of such targeted therapeutics onto the market pancreatic cancer research is focused on the identification of novel drug targets to produce immunoconjugates against this disease. Choosing the appropriate target for ADCs is a critical step that affects the efficacy of these biotherapeutics (Beck et al. 2017).

Besides the lack of effective treatments for pancreatic cancer at present, serum carbohydrate 19-9 (CA19-9) is the only pancreatic cancer biomarker (prognostic biomarker only) approved for use by the US Food and Drug Administration (FDA). CA19-9 can be elevated in many types of gastrointestinal cancers such as colorectal

cancer and oesophageal cancer, as well as in patients with pancreatitis, obstructive jaundice, cholangitis, cirrhosis and diseases of the bile ducts (Takahashi et al. 2003, Maestranzi et al. 1998). Unfortunately, despite sensitivity and specificity of approximately 79% and 82% respectively, CA-19-9 is inadequate for the early detection of PDAC in asymptomatic patients and cannot satisfactory diagnose PDAC; however, if elevated it is useful in following patients with known disease. The lack of diagnostic biomarkers and general screening results in an inability to detect early-stage pancreatic cancer which enables metastasis to occur and significantly decreases survival rates. This situation emphasises the need for more specific biomarkers against pancreatic cancer to be identified so that it can be diagnosed at an early stage to prevent poor outcomes for patients.

Comprehensive reviews by Coleman et al. (Coleman et al. 2016) and Pan et al. (Pan et al. 2013b), among others, outlines proteomic studies in pancreatic cancer to date. Although there have been reports of potential biomarkers and diagnostic targets none have been clinically approved. Many proteomic profiling studies have been performed on cell lines (Haun et al. 2015a, Kim, Y. et al. 2014, Gronborg et al. 2006b) but great care must be taken when studying cell lines as they are cultured repeatedly for years and may not retain the cancer proteome of the original patient. Pancreatic tissue proteomics has made some significant findings largely due to the increase in sensitivity of mass spectrometers and an improvement in its associated preparations. Isotope-coded affinity tag technology (ICAT) has been used to identify proteins in PDAC tissues which are involved in invasion and metastasis (Chen, Ru, Eugene, Donohoe, Pan, Eng, Cooke, Crispin, Lane, Goodlett and Bronner 2005b). A similar study using ICAT found that 50% of proteins overexpressed in pancreatitis were previously identified in pancreatic cancer studies and these proteins potentially contribute to false positive biomarkers of pancreatic cancer (Chen, R. et al. 2007). Multiple studies have identified Galectin-1 as overexpressed in PDAC tissues through various proteomic approaches including 2-dimensional gel electrophoresis, Western blotting, immunohistochemistry and mass spectrometry (Shen et al. 2004, Berberat et al. 2001). More recently it has been established as a stromal protein in PDAC which mediates tumour-stroma crosstalk to regulate PDAC progression and suggests it has therapeutic potential (Orozco et al.

2018). Membrane proteomic studies of PDAC to date are limited to cell line analyses (Wang, Weibin et al. 2016, Haun et al. 2015b, Liu et al. 2010). Cell surface glycoprotein labelling or membrane enrichment kits followed by mass spectrometry were the methods of choice for membrane proteome profiling in these studies.

This study set out to compare the proteomes of a panel of human pancreatic adenocarcinomas, adjacent-normal tissue and patient derived xenografts (PDX) from these tumours. The PDX sample set were analysed at the first and second generation in mice - F1 and F2. The tumours selected were characterised by a consultant Histopathologist. The xenograft samples were analysed to determine their stability as models for PDAC and maintaining the PDAC proteome across generations of PDX mice. We specifically enriched membrane and membrane-associated proteins from the samples as candidate membrane proteins found to be overexpressed in the pancreatic tumours (primary and/or xenograft) have potential as targets for novel targeted therapeutics such as antibody-drug conjugates or monoclonal antibody therapies. Through the engraftment process, and as the tumours grow in mice, the patient stroma is overtaken with tissue of mouse origin (Delitto et al. 2015a). This allows an interrogation of the complex tumour/stroma relationship. On the proteomic level using high-resolution mass spectrometry we have selectively determined the tumour cell derived human-specific proteins contributing to PDAC tumorigenesis. This study is an important step forward for PDAC research and presents tumour- associated proteins which have potential as biomarkers or targets for ADC therapy or other antibody-derived targeted therapies. This study presents the first membrane protein analysis of PDAC primary tissues and PDX PDAC tissues.

3.2 Materials and Methods

3.2.1 Patient Demographics and PDX information

Ten pancreatic cancer primary tumour tissue specimens and 9 normal-adjacent specimens, which were patient-matched, were obtained from St. Vincent's University Hospital (SVUH), Dublin between 2013 and 2015. All patients were undergoing pancreatic adenocarcinoma resection. The average age of the patients was 66 years old, of those n=7 are male and n=4 are female. All samples were collected and processed in compliance with the SVUH and Dublin City University (DCU) ethics committees. Pancreatic adenocarcinomas patient-derived xenograft (PDX) tissues are generated by subcutaneous seeding in SCID mice in-house at DCU. Twenty successful PDX tissues n=10 for both the F1 and F2 generations were used for this study. F1 generation refers to mice subcutaneously engrafted with primary patient tumour material whereas the F2 generation of mice were injected with a fragment of the F1 tumour. All samples were cryopreserved at -80^{o} C on the day of extraction until sample preparation was performed. Representative tumour tissue was formalin fixed paraffin embedded for each PDX tumour. Patient and sample details are provided in Supplementary Table S1. Primary patient samples were confirmed as pancreatic adenocarcinoma by a pathologist (N.S.) in SVUH. PDX samples were also confirmed by pathology examination to maintain the human tumour content and morphology of the original tumour.

3.2.2 Membrane Protein Enrichment and Protein Digestion

Tissue specimens were subjected to membrane protein enrichment using the Mem-PER Plus Membrane protein extraction kit (Thermo Scientific, 89842) which applies a mild detergent-based selective extraction protocol to enrich integral membrane proteins and membrane associated proteins. The extraction was performed essentially according to manufacturer's instructions for hard tissue except that buffer volumes were adjusted for the samples depending on the weight. The membrane-enriched fraction was used for all further analysis of the samples in this study. The membrane-enriched fraction from each of the tissue samples was cleaned up using the ReadyPrep 2D Clean-up Kit (Bio-Rad) according to manufacturer's instructions. The cleaned protein pellets were resuspended in a buffer containing 6 M urea, 2 M thiourea, and 10 mM Tris, pH 8.5 and assayed for

protein concentration using the QuickStart Bradford Protein Assay (Bio-Rad). Fifteen micrograms of protein were suspended with 50 mM ammonium bicarbonate and protein digestion was carried out as previously described (Linge et al. 2014).

3.2.3 Quantitative Label-free LC-MS/MS and Data Analysis

Nano LC−MS/MS analysis was carried out using a Dionex Ultimate 3000 RSLCnano system (Thermo Fisher Scientific) coupled to a hybrid linear ion trap/Orbitrap mass spectrometer (LTQ Orbitrap XL; Thermo Fisher Scientific). LC-MS/MS methods were applied as previously described (Linge et al. 2014). Quantitative label-free data analysis was carried out using Progenesis QI for Proteomics (version 2.0; Nonlinear Dynamics, a Waters company, Newcastle upon Tyne, UK), essentially as recommended by the manufacturer (www.nonlinear.com). Peptide and protein identification was achieved with Proteome Discoverer 2.1 using Sequest HT (Thermo Fisher Scientific), Mascot and Percolator. Data for the membrane-enriched fraction of all tumours and adjacent-normal differential analyses were searched against the NCBI Uniprot Swissprot *Homo sapiens* database containing 20,148 sequences (fasta file downloaded May 2017). Data for the PDX analyses were searched against a dual human: mouse fasta database to accurately determine the origin of the protein identifications i.e. primary human tumour or SCID mice stroma. This database was created in-house by merging the NCBI Uniprot Swissprot *Homo sapiens* database containing 20,148 sequences (fasta file downloaded May 2017) and NCBI Uniprot Swissprot *Mus musculus* database containing 16,863 sequences (fasta file downloaded May 2017).

To better analyse the PDX model and evaluate the tumour/stroma (human/mouse) origin of the orthologous proteins in the xenograft experiments, only peptide spectrum matches (PSMs) matching to (1) an unambiguous protein and (2) 1 protein identification, were allowed. The PSM option omits peptides that cannot be unambiguously matched to one protein i.e. if a PSM matches two proteins: the mouse and human version of the same protein, it is omitted therefore only species-specific PSMs and thus peptides are retained. If evidence for BOTH human and mouse peptides from an orthologous protein were observed, then peptides that cannot distinguish between the two species were ignored. Assembly of unambiguous, species specific confidently identified PSMs

resulted in an approximate loss of two-thirds of the proteomic data. However; all differential peptide data carried forward for analysis with Progenesis QI could now be confidently distinguished from peptides derived from the tumour and those orthologues with the mouse form of the protein.

The following search parameters were used for protein identification: (1) peptide mass tolerance set to 20 ppm, (2) MS/MS mass tolerance set to 0.6 Da, (3) up to two missed cleavages were allowed, (4) carbamidomethylation of cysteine set as a fixed modification and (5) methionine oxidation set as a variable modification. Only highly confident peptide identifications with an FDR ≤ 0.01 (identified using a SEQUEST HT workflow coupled with Percolator validation in Proteome Discoverer 2.1 (Thermo Fisher Scientific)) were imported into Progenesis QI software for further analysis. Protein identifications were reviewed and only those which passed the following criteria were considered differentially expressed between experimental groups with high confidence and statistical significance: (i) an ANOVA p-value of ≤ 0.05 between experimental groups; (ii) proteins with ≥ 2 unique peptides contributing to the identification; (iii) a ≥ 1.5- fold change in relative abundance between the two experimental conditions. For the patient matched analysis of adjacent normal tissues and primary PDAC a repeated measures ANOVA was used for paired sample statistical analysis. To calculate the maximum fold change for a protein, Progenesis calculates the mean abundance for that protein in each experimental condition. These mean values are then placed in a condition-vs-condition matrix to find the maximum fold change between any two conditions' mean protein abundances. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository.

Data visualization was achieved using violin plots and heat maps which were generated using the ggplot package in R. We determined the coefficient of variation (CV) of the patients for each differentially expressed protein and plotted the CV distribution as violin plots. Violin plots are similar to box plots, but instead of showing summary statistics such as median and interquartile ranges, the violin plot shows the full variable distribution of all the protein abundances with the widest points of the violin indicating

the most common CV values. Violin plots were generated using R-studio and the ggplot2 library. The relative expression levels of the differentially expressed proteins across the experimental groups was visualised using heat maps generated with ggplot2.

3.2.4 Immunohistochemistry

All immunohistochemical (IHC) staining was performed using the DAKO Autostainer (DAKO). Antigen retrievals were performed using the Epitope Retrieval 3-in-1 solution (pH6) (DAKO). For epitope retrieval, slides were heated to 97°C for 20 minutes and then cooled to 65°C. DAKO Real Envision Detection System, Peroxidase/DAB+ detection system was used.

On the autostainer (DAKO, Glostrup, Denmark) slides were blocked for 10 minutes with 200µl HRP Block (DAKO, Glostrup, Denmark). Slides were washed with 1X wash buffer and 200µl of the primary antibody were assed to the slides for 30 minutes. Slides were once again washed with 1X wash buffer and then incubated with 200µl Real EnVision (DAKO, Glostrup, Denmark) for 30 minutes. Slides were washed again with wash buffer and then stained with 200µl DAB+ substrate for 5 minutes and this procedure was repeated twice. All slides were then counterstained with haematoxylin (DAKO) for 5 minutes and were rinsed with deionised water and then with wash buffer. A negative control (NC) sample was also tested for each sample using antibody diluent without any antibody present. This was used to evaluate any non-specific staining. Following the counterstaining with haematoxylin, the slides were then dehydrated. This was achieved by immersing the slide in 70%, 90% and 100% EtOH, twice in each EtoH solutions for 3 minutes. The slides were then immersed into xylene, twice, for 5 minutes each. Once the slides were cleared they were mounted using DPX (BDH). In order to visualise and evaluate the amount of mouse stromal infiltration in to the human tumour cells and tissue, a mouse monoclonal to human mitochondria (1/1000 Ab92824, Abcam, Cambridge, UK) antibody was used.

3.2.5 Western blotting

Equal concentrations (10 µg) of the membrane-enriched fraction protein samples were prepared in Laemmli SDS-PAGE loading buffer and denatured for 5 minutes at 95 degrees Celsius. Samples were loaded onto 4-12% NuPAGE Bis-Tris gels and separated

at 180V for 1 hour. PageRuler Plus pre-stained protein ladder (Thermo Scientific) was used as the molecular weight marker. Electrophoretic transfer was achieved using the Power Blotter (Thermo Scientific) and the nitrocellulose membrane was blocked for 1 hour using 5% non-fat dried milk powder. Development of blots was carried out using chemiluminescence in the dark room. Western blots for membrane protein enrichment analysis were probed with the following primary antibodies: anti-Caveolin-1 (Ab2910, Abcam) and anti-Sodium Potassium ATPase (Ab76020, Abcam). Western blot validation of CD55 expression was carried out using anti-CD55 (Ab133684, Abcam). Coomassie Blue stained gels corresponding to the Western blot of CD55 are shown in supplementary figure 3.

3.3 Results

3.3.1 Membrane Proteome Coverage

For discovery of candidate human proteins, membrane proteins were enriched from adjacent-normal pancreatic tissues, primary PDAC tumours, and PDAC tumours derived from PDX models at the first (F1) and second (F2) generations totalling a sample set of 40 specimens. The efficiency of the membrane protein enrichment kit was tested using a pancreatic cancer cell line: AsPc1. Figure **Error! Reference source not found.**3.1 hows a Western blot of membrane specific –enrichment using antibodies against two membrane-specific proteins, caveolin 1 and sodium ATPase. Both western blots show increased expression for both membrane proteins in the membrane fraction of the kit, absence in the cytosolic fraction and low expression in the total protein lysate. These western blot results confirm that the membrane protein enrichment kit used has successfully selected membrane-associated proteins in the membrane-enriched fraction.

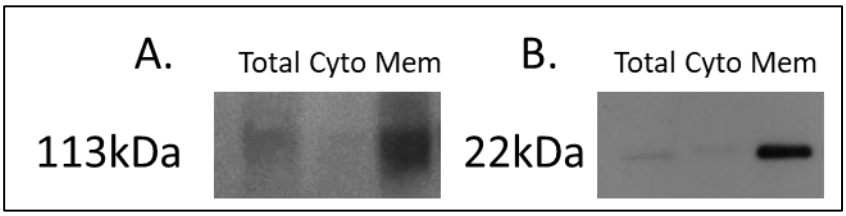


Figure 3.1 Western blot analysis of (A) Sodium Potassium ATPase and (B) Caveolin-1 across AsPc1 cell line fractions. A total protein lysate (Total), the cytosolic (Cyto) fraction and membrane-enriched (Mem) protein fractions from the MemPER kit were analysed. Both membrane protein markers are increased in expression in the membrane fraction compared to the total proteome and cytosolic fraction.

Quantitative label-free proteomic analysis by LC-MS/MS was performed on the membrane-enriched protein fractions. The total number of protein identifications from enriched fractions for adjacent-normal tissues, PDAC tumours, and PDX F1 and F2 generations are 2387, 2485, 2617 and 2588 respectively (Supplementary table S2). Gene Ontology annotation using the ProteinCenter software within Proteome Discoverer indicated that membrane associated proteins comprised 48.8%, 66.5%, 47.8% and 59.5% of the total proteins identified for adjacent-normal, tumour, PDX F1 and PDX F2 samples respectively (Supplementary figure 1).

3.3.2 Differentially Expressed Proteins between Matched Adjacent-Normal and Tumour Tissues

Nine matched adjacent-normal and PDAC tumour tissues were subjected to membrane protein enrichment and equal concentrations of each fraction were digested with trypsin prior to analysis by mass spectrometry for a "bottom-up" proteomic approach. Raw MS data was interrogated using Progenesis QI for Proteomics label-free LC-MS software to identify differentially expressed proteins between the two sample groups. The software selects one of the raw files as the reference run to which all other samples were aligned, allowing relative quantitation of proteins between experimental groups to be calculated.

Differentially expressed proteins between the adjacent-normal and tumour tissues were determined by using a repeated measures ANOVA *p*-value of <0.05, a fold-change cut-off of ± 1.5-fold and a minimum of 2 unique peptides contributing to a protein identification. Principal components analysis (Figure 3.2A) and unsupervised Pearson clustering (Figure 3.2B) of the differentially expressed proteins shows an evident separation of the experimental groups into two categories relating to the adjacent-normal and tumour tissues. Next, we assessed the CV among the patients for each of the sample cohorts and plotted the distribution as violin plots (Figure 3.2C). The average CV of the tumour specimens is lower than that of the adjacent-normal group, perhaps due to the heterogeneity of the latter samples. All PDAC tumours were removed from the pancreas and are all malignant in nature. Adjacent-normal tissues are taken from any region of the pancreas once it is sufficiently distant from the tumour to be deemed non-cancerous, thus leading to a variety of tissue regions acting as an adjacent-normal control and

contributing to the degree of variability shown here. The inherent variability across ten patients also adds to the variability seen within both sample groups.

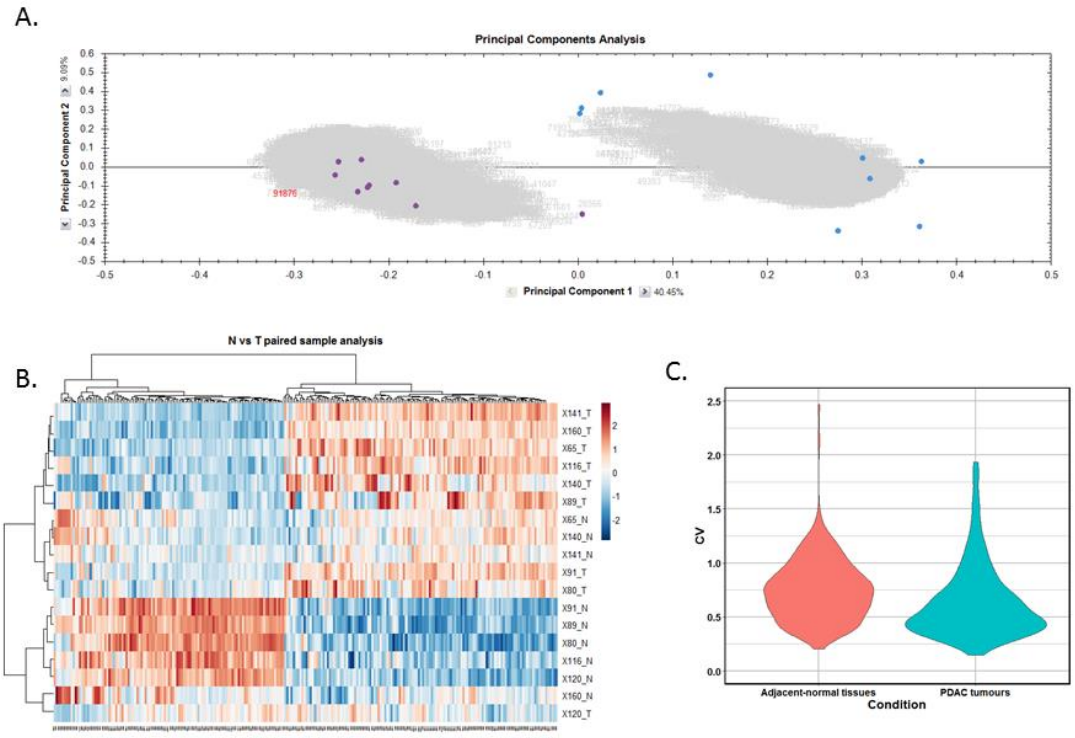


Figure 3.2: Proteomic clustering of pancreatic cancer tumours and adjacent-normal tissues. A. Principal components analysis from Progenesis QI for Proteomics shows separation of the PDAC tumours (purple, n=9) and the adjacent-normal tissues (blue, n=9). B. Heat map using unsupervised Pearson clustering of all differentially expressed proteins visualises their expression levels for each sample. C. The CV was calculated for each DE protein in the sample groups and those values are displayed as violin plots.

From this analysis, a total of 238 proteins were identified as differentially expressed between matched adjacent-normal and tumour tissues using paired sample statistical analysis. Within these proteins, 129 proteins were overexpressed and 109 under-expressed in the PDAC tissue. The accession number, number of unique peptides identified, fold-change and *p*-value of these proteins are outlined in Supplementary Table S3. The top 25 differentially expressed proteins with increased expression in PDAC tumour tissues compared to adjacent-normal tissues are shown in Table 3.1. Of the 129 proteins significantly overexpressed in the PDAC specimens, proteins with the most elevated expression levels, with values of 5-fold or higher, were integrin beta-6, fibronectin, thrombospondin-1 and 2, periostin, immunoglobulin superfamily containing leucine-rich repeat protein, 14-3-3 protein sigma, transforming growth factor beta-1-induced transcript 1 protein, adipocyte enhancer-binding protein 1 and complement decay-accelerating factor (CD55). To investigate further the potential clinical relevance of these identified candidates, we analysed the Badea 2008 gene expression data set (Badea et al. 2008) using Oncomine software (Life Technologies), which compares human PDAC cases to normal pancreas tissues. We analysed 4 proteins of interest; ITGB6, FN1, THBS2 and CD55 from our data. The mRNA levels of these candidate proteins are significantly upregulated in the Badea 2008 PDAC tumour data set (n = 39), validating our results in the context of a larger, independent PDAC cohort. This emphasises the potential of these proteins as markers of PDAC for drug targeting or diagnostics (see Figure 3.3).

Table 3.1 List of top 25 differentially expressed proteins with increased expression in PDAC tumour tissues compared to adjacent-normal tissues as determined by quantitative label-free mass spectrometric analysis using repeated measures ANOVA for paired sample analysis.

Accession	Protein Name	Gene	Peptides	p-value	Fold-change
P18564	Integrin beta-6	ITGB6	2	8.80E-03	8.69
P02751	Fibronectin	FN1	3	4.67E-02	7.95
P07996	Thrombospondin-1	THBS1	5	2.56E-02	7.65
Q15063	Periostin	POSTN	8	2.06E-02	7.34
O14498	Immunoglobulin superfamily containing leucine-rich repeat protein	ISLR	2	4.27E-02	7.10
P31947	14-3-3 protein sigma	SFN	3	1.19E-03	6.06
P35442	Thrombospondin-2	THBS2	5	1.33E-02	6.02
O43294	Transforming growth factor beta-1-induced transcript 1 protein	TGFB1I1	2	5.98E-03	5.75
Q8IUX7	Adipocyte enhancer-binding protein 1	AEBP1	3	1.72E-02	5.59
P08174	Complement decay-accelerating factor	CD55	2	8.43E-04	5.58
Q03518	Antigen peptide transporter 1	TAP1	2	4.62E-03	5.00
Q03519	Antigen peptide transporter 2	TAP2	2	2.42E-02	4.24
O15533	Tapasin	TAPBP	2	1.14E-02	4.07

P23142	Fibulin-1	FBLN1	6	1.44E-02	3.81
Q08380	Galectin-3-binding protein	LGALS3BP	3	7.24E-03	3.80
P50454	Serpin H1	SERPINH1	2	5.73E-04	3.64
P02792	Ferritin light chain	FTL	3	3.77E-02	3.56
P98095	Fibulin-2	FBLN2	2	1.10E-02	3.41
P20592	Interferon-induced GTP-binding protein Mx2	MX2	2	2.43E-04	3.38
Q6UX06	Olfactomedin-4	OLFM4	3	4.26E-02	3.27
Q01518	Adenylyl cyclase-associated protein 1	CAP1	2	3.22E-02	3.22
Q01995	Transgelin	TAGLN	6	2.51E-02	3.20
Q96HE7	ERO1-like protein alpha	ERO1A	11	2.39E-03	3.16
Q92597	Protein NDRG1	NDRG1	2	3.28E-02	3.15
Q6PIU2	Neutral cholesterol ester hydrolase 1	NCEH1	4	1.51E-05	3.13

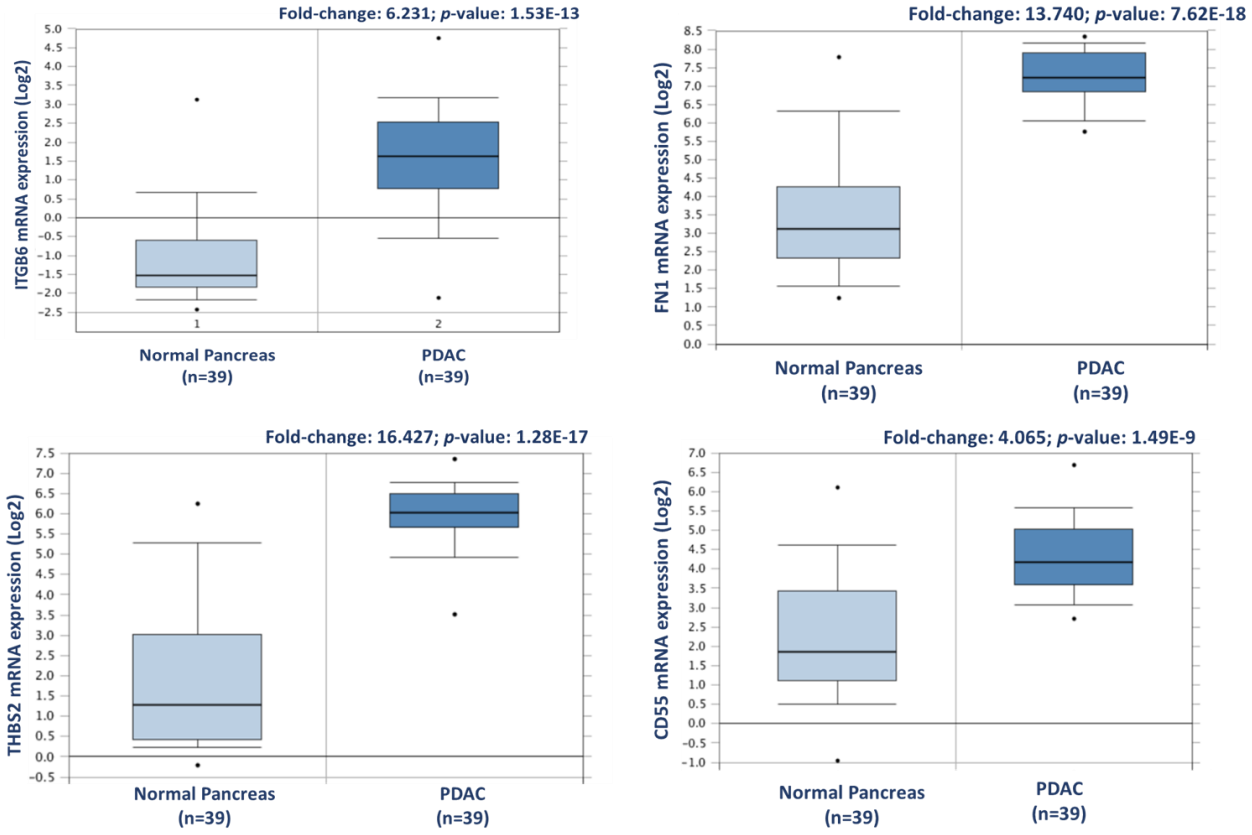


Figure 3.3 Validation of differential expression levels for 4 candidate proteins across a larger, separate PDAC cohort using Oncomine. ITGB6, FN1, THBS2 and CD55 expression levels were all significantly increased in PDAC (n=39) compared to normal pancreas in the Badea et al. 2008 gene expression set.

3.3.3 Differentially Expressed Proteins between Matched Tumour and PDX F1 Tissues

3.3.3.1 Proteome Analysis of PDAC Xenograft Tumours

From the 20 PDAC PDX samples, a total of 2617 and 2588 proteins were identified for the F1 and F2 generation, respectively. These proteins represent a combination of human and mouse identifications because of the mouse proteome integration with the tumour during the engraftment process (Delitto et al. 2015a). To overcome the challenge of identifying tumour derived proteins and thus potential biomarkers and targets for PDAC, all species-indistinguishable proteins from this study were removed and analysed separately. On average, the protein-coding regions of the mouse and human genomes are 85 percent identical; some genes are 99 percent identical while others are

only 60 percent identical (National Human Genome Research Institute, 2010). This orthologous nature of mouse and human genomes makes it difficult to confidently assign protein identifications from PDX models to one species. The combined human: mouse FASTA database allows the Proteome Discoverer identification software and search engine algorithms to impartially search all MS/MS spectra against every annotated human and/or mouse protein sequence and provide it with an algorithm score and accession number. Where an MS/MS spectrum matches both the human and mouse counterparts of a protein with equivalence, the corresponding human and mouse accession numbers are reported in the peptide-spectrum match (PSM) field. By filtering for only PSMs which have a single protein match we can assume that such peptides are species-specific. Such peptide identifications can confidently decipher the origin of the protein, i.e., if it is tumour-derived or mouse-derived. On average, two-thirds of all PSMs from the PDX samples are homologous to the human and mouse counterparts, making them species indistinguishable. The species-specific and species-indistinguishable sets of PSMs were analysed and differential protein lists were generated. This evolutionary change of amino acids within a protein is advantageous, as it indicates that the identified proteins are most likely derived from the patient PDAC tumour cells proliferating within the mouse tumour microenvironment (Delitto et al. 2015a) . Two representative chromatograms of human-specific peptides identified in this study and their corresponding BLASTp search are shown in supplementary figure 2. These results demonstrate the feasibility of identifying large numbers of human-specific proteins from xenograft mouse models of PDAC and applying this method to other solid tumour xenografts.

3.3.3.2 Differential expression of Human-specific Proteins

An analysis of the PDAC patient tumours and PDX F1 tumours was carried out to identify candidate proteins and confirm the maintenance of the proteomic profile for PDAC in PDX models. Within this comparison, 9 of the 10 PDX F1 samples were patient-matched to the original tumour specimens (as shown in supplementary table 1), allowing a stringent and well-controlled comparison of the PDX and patient tumour proteomes. As described above, differentially expressed proteins between the primary

and PDX F1 tumours were determined using an ANOVA p-value of <0.05, a fold-change cut-off of ± 1.5-fold, and a minimum of 2 unique peptides contributing to a protein identification. For this analysis, only PSMs that matched to a single protein identification and that identification is human and not mouse, were used.

Upon tumour engraftment into the mice, mouse-derived cells replace the non-cancer human cells of the original tumour, and thus only the cancer cells of the patient tumour remain (Delitto et al. 2015b, Siolas and Hannon 2013). Figure 3.4 shows representative immunohistochemistry staining of two PDX F1 tissue sections for patients 80 and 99 with a human mitochondria specific antibody. IHC staining in figure 3.4B confirms that the stromal component of the PDAC tissues is unstained and thus is of mouse origin after engraftment.

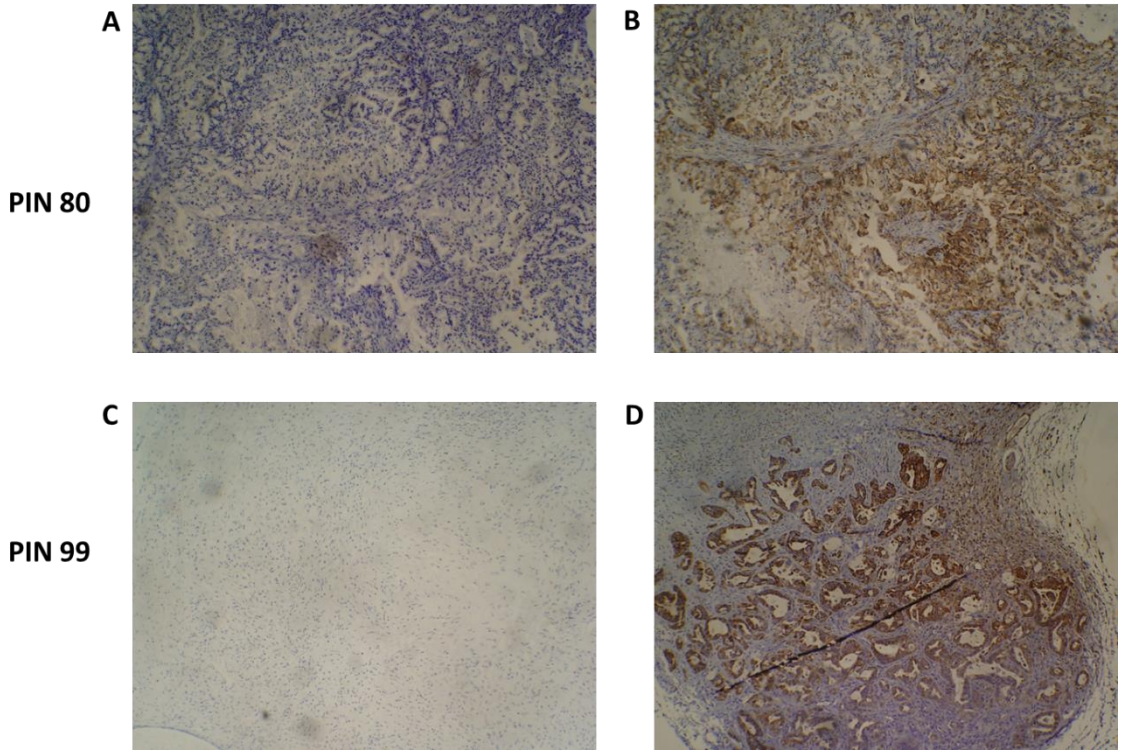


Figure 3.4 Species-specific immunohistochemical analysis of PDX tumour samples. An antibody specific for human mitochondria was used to identify the presence of human cancer cells present in the PDX tumours. Representative images of PIN 80 (A&B) and PIN 99 (C&D) show negative staining (A&C) and human species-specific staining of mitochondria (B&D). Antibody-positive cells are marked by dense brown staining in panels B&D.

We selected human-only proteins from the differential proteome analysis of PDX F1 versus primary tumours, excluding mouse proteins, as our interest is to identify tumour specific proteins. A total of 143 differentially expressed, human-specific proteins were identified, including 32 overexpressed and 111 under expressed in the PDX F1 tumours when compared to the original patient PDAC tumours. The accession number, gene name, number of peptides for identification, and p-value for these proteins are outlined in Supplementary table S3. Principal components analysis (Figure 3.5A) and unsupervised Pearson clustering (Figure 3.5B) of the differentially expressed proteins shows an evident separation of the experimental groups into two clear categories corresponding to the tumour tissues and PDX F1 tissues. We determined the CV within the 10-sample set for each differentially expressed protein and plotted the CV distribution as violin plots (Figure 3.5C). The average CV of the PDX F1 tumours is lower than that of the patient tumours. This could be a result of the tumour implantation into the mice acting as a selection process whereby only aggressive, highly expressed pivotal proteins are retained in the PDX model, thus yielding a unique signature of proteins with little variation across the multiple specimens. Alternatively, a fraction of the differentially-expressed proteins in the patient sample were stroma-derived and were replaced by mouse proteins in the PDX F1 tumour. The original PDAC patient tumours show greater variation in expression of the differentially-expressed proteins, perhaps reflecting inter-patient heterogeneity and the different PDAC stages of the resected cancer. The PDX models are a much more controlled study group with regards to, for example, tumour volume and duration of tumour growth.

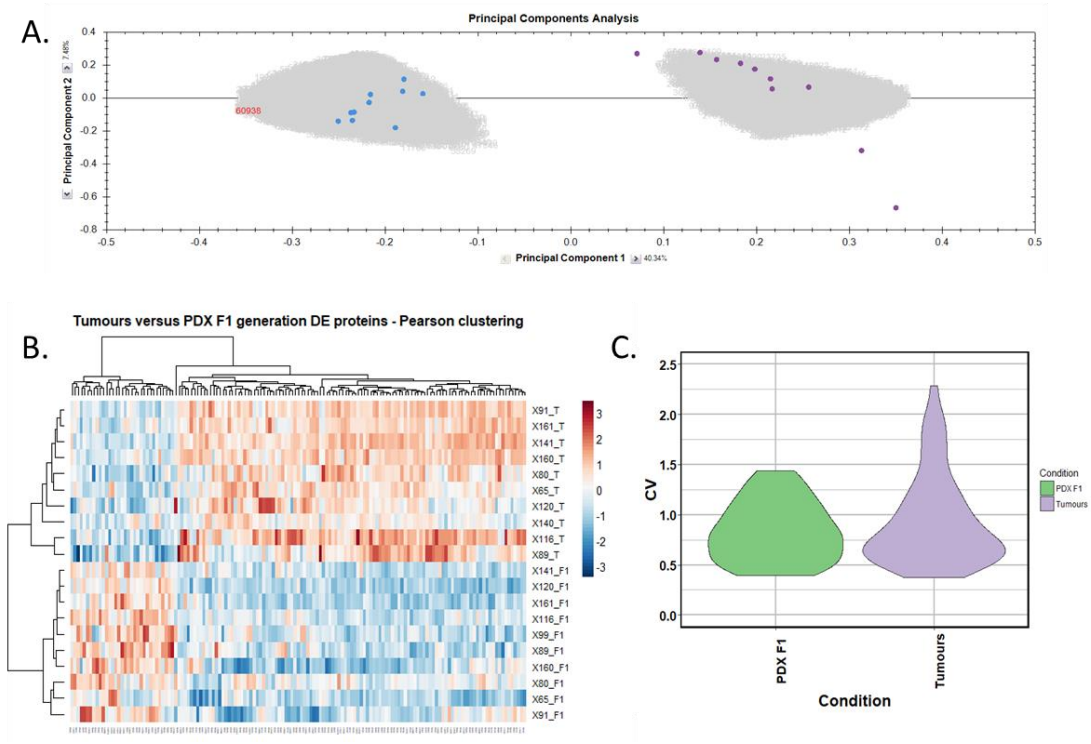


Figure 3.5 Proteomic clustering of pancreatic cancer tumours and PDX F1 tissues. A. Principal components analysis from Progenesis QI for Proteomics shows a clear separation of PDX F1 (blue, n=10) and the primary PDAC tissues (purple, n=10). B. Heat map using unsupervised Pearson clustering of all human-specific differentially expressed proteins visualises their expression levels for each sample. C. The CV was calculated for each differentially expressed protein in the sample groups and those values are displayed as violin plots.

These human specific proteins are potentially compelling biomarkers and targets for pancreatic cancer therapy as they are tumour cell-associated. Through the PDX engraftment process the tumours undergo selection pressure, whereby proteins linked to proliferation and responsible for driving growth may increase in expression. The cohort of differentially expressed human proteins with higher abundance levels in the F1 PDX generation thus potentially represent key regulators of PDAC disease progression.

The 32 human proteins with significantly increased expression in the PDX F1 tumour tissues compared to the patient tumours are outlined in Table 3.2. Of these, the most elevated expression levels, with values of 3-fold or higher, were established for Periplakin, Serpin B5, Hydroxymethylglutaryl-CoA synthase, Chloride intracellular channel protein 3, Integrin beta-4, Protein S100-A16, Annexin A13, Protein S100-A14 and Aldo-keto reductase family 1-member B10.

Table 3.2 The 32 human-specific proteins identified with significantly increased expression in PDX F1 tumours compared to PDAC tumour tissues as determined by quantitative label-free mass spectrometric analysis

Accession	Protein Name	Gene	Peptides	ANOVA (p)	Fold change
O60437	Periplakin	PPL	2	3.34E-08	7.85
P36952	Serpin B5	SERPINB5	7	2.37E-03	4.86
P54868	Hydroxymethylglutaryl-CoA synthase, mitochondrial	HMGCS2	3	4.29E-03	4.79
O95833	Chloride intracellular channel protein 3	CLIC3	2	6.99E-03	3.62
P16144	Integrin beta-4	ITGB4	2	6.03E-04	3.35
Q96FQ6	Protein S100-A16	S100A16	3	9.17E-04	3.28
P27216	Annexin A13	ANXA13	6	2.72E-03	3.27
Q9HCY8	Protein S100-A14	S100A14	4	5.20E-03	3.09
O60218	Aldo-keto reductase family 1-member B10	AKR1B10	3	2.04E-02	3.03
P11047	Laminin subunit gamma-1	LAMC1	2	9.87E-03	2.97
Q12864	Cadherin-17	CDH17	4	2.27E-02	2.81
P09327	Villin-1	VIL1	3	5.26E-04	2.68
Q9H190	Syntenin-2	SDCBP2	2	1.15E-03	2.67
P05787	Keratin, type II cytoskeletal 8	KRT8	5	1.51E-02	2.63
P15311	Ezrin	EZR	2	2.79E-03	2.6
P56470	Galectin-4	LGALS4	7	1.09E-02	2.58
Q14764	Major vault protein	MVP	5	2.31E-03	2.42
Q8NFV4	Alpha/beta hydrolase domain-containing protein 11	ABHD11	4	6.45E-03	2.4
P42765	3-ketoacyl-CoA thiolase, mitochondrial	ACAA2	2	7.25E-04	2.14
P99999	Cytochrome c	CYCS	2	6.30E-03	2.14

P08727	Keratin, type I cytoskeletal 19	KRT19	9	2.03E-02	2.11
P17931	Galectin-3	LGALS3	2	1.92E-02	2.1
P55011	Solute carrier family 12 member 2	SLC12A2	2	2.33E-02	2.04
P22307	Non-specific lipid-transfer protein	SCP2	2	2.01E-02	1.99
P09972	Fructose-bisphosphate aldolase C	ALDOC	2	1.83E-03	1.94
P12830	Cadherin-1	CDH1	3	1.57E-03	1.88
Q14126	Desmoglein-2	DSG2	4	7.28E-05	1.84
Q96I24	Far upstream element-binding protein 3	FUBP3	2	1.23E-02	1.83
Q02218	2-oxoglutarate dehydrogenase, mitochondrial	OGDH	2	1.60E-03	1.82
Q9NR45	Sialic acid synthase	NANS	3	1.48E-02	1.61
P14618	Pyruvate kinase PKM	PKM	3	3.90E-02	1.56
P55072	Transitional endoplasmic reticulum ATPase	VCP	2	1.50E-02	1.51

To investigate the potential clinical relevance of the highest differentially expressed proteins in F1, we analysed the Badea 2008 microarray gene expression data set and the Ramaswamy 2001 multiclass cancer microarray gene expression set using Oncomine software (Figure 3.6) (Badea et al. 2008, Ramaswamy et al. 2001). Periplakin (PPL), which is over 7-fold higher in our cohort of F1 tumours compared to primary tumours, is also significantly overexpressed in the Badea gene expression set of PDAC tumours (n = 39) compared to normal pancreas tissue (n=39). PPL also showed the highest gene expression in PDAC compared to 11 other cancer types in the Ramaswamy 2001 multi-cancer gene expression set. This result illustrates that our proteomic data validates in a larger, separate PDAC cohort and that targets identified here have strong potential as therapeutic targets for PDAC.

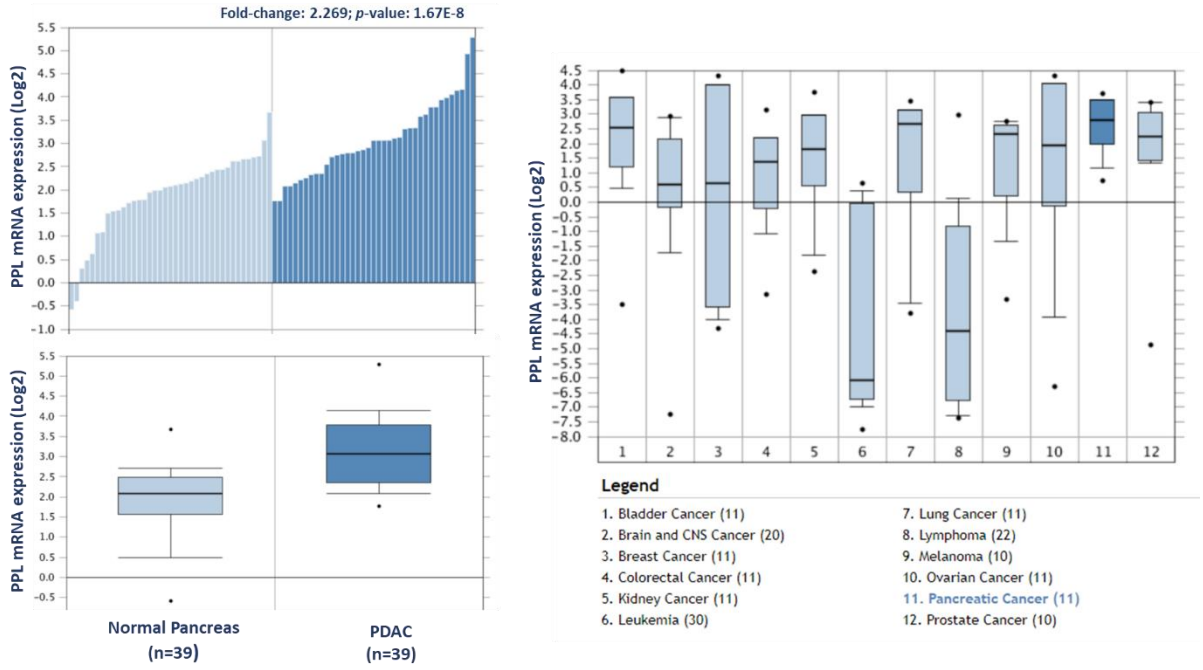


Figure 3.6 Expression of Periplakin in a larger PDAC cohort and across multiple cancer types. mRNA expression of Periplakin (PPL) is significantly higher in PDAC tumours (n=39) than in normal pancreas (n=39). Similarly, PPL mRNA expression is significantly highest in pancreatic cancer compared to 11 other cancer types. The Oncomine database (www.oncomine.org, Life Technologies) was used to analyse our candidates' expression in the Badea 2008 PDAC, and Ramaswamy multi-cancer microarray expression data sets.

3.3.3.3 Species-indistinguishable Proteins

The human specific proteins identified above have high potential as PDAC tumour biomarkers. However, two-thirds of the significantly differential PSMs between PDX F1 and primary PDAC tumours could not be assigned definitively to a single species because the MS/MS analysis and search software did not detect species-unique peptide(s) which could determine whether those proteins are derived from mouse (stromal) cells or human (cancer) cells. A total of 599 differential proteins were identified by peptides common to both human and mouse sequences and are therefore species indistinguishable identifications. Of these, 330 proteins were overexpressed and 269 were under expressed in the PDX F1 tissues compared to the original PDAC tumours. The accession number, fold-change, number of peptides and p-value of these proteins are outlined in Supplementary Table 3.

Identifications using a dual database for xenograft studies for which PSMs report two matching proteins that are not species-unique may be false positive identifications. For example, of the 269 proteins with higher abundance in the PDAC primary tumours compared to the mouse PDX tumours, 206 are assigned to the human species whereas 63 are identified as mouse. The latter group represents an impossibility, as the original patient tumours contain no mouse proteins, and these are false-positive identifications that should have been assigned as human. These candidate markers are classified as species-indistinguishable proteins. Because this large significantly differential indistinguishable cohort potentially presents an interesting set of proteins that may play a role in tumour progression, they deserve closer inspection.

330 indistinguishable proteins were identified with a higher abundance in the PDX F1 compared to primary PDAC tumours. This cohort of differential proteins represents a persistent aggressive subset of proteins which have the capabilities to survive tumour engraftment and proliferate. Some of the most differentially expressed proteins with a fold-change of 10 or more include Platelet glycoprotein 4, Probable ATP-dependent RNA helicase DDX17, Xanthine dehydrogenase/oxidase, Tropomyosin alpha-4 chain, von Willebrand factor A domain-containing protein 5A, Thymosin beta-4, Serum albumin, Granulins, Annexin A11, NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 8 and Lysosomal protective protein.

3.3.4 Stability of Protein Expression over PDX generations

The stability of PDX models of pancreatic cancer for biomarker discovery was evaluated by profiling the tumour proteome over various PDX generations. Pancreatic tumours harvested from the first generation in mice were expanded into further generations. Tumour fragments were stored at each generation and subjected to label-free proteomic analysis. Supplementary Table 1 illustrates the success of the pancreatic PDX models generated in-house with most PDX F1 tumours progressing to successful engraftment and expansion in the F2 PDX generation of mice. Ten F1 generation and ten F2 generation PDX tumours were analysed, with 9 out of 10 of these tumours being patient-matched. Quantitative label-free differential analysis of the F1 and F2 tumours identified a small number of differentially expressed proteins, with 8 human-specific

proteins (Table 3.3) and 41 proteins of indistinguishable species (Supplementary table 3) as significantly differential in abundance levels.

Table 3.3 The 8 significantly differentially expressed human-specific proteins in a comparison of F1 and F2 tumours. The first 5 proteins have an increased expression in the PDX F2 generation compared to F1 tumours, whereas the last 3 proteins listed have higher expression levels in the F1 tumours compared to the F2 tumour group.

Accession	Protein Name	Highest in	Gene	Peptides	ANOVA (p)	Fold change
O75874	Isocitrate dehydrogenase [NADP] cytoplasmic	F2	IDH1	2	1.22E-02	1.6
P07339	Cathepsin D	F2	CTSD	2	2.06E-02	1.62
P30041	Peroxiredoxin-6	F2	PRDX6	2	2.18E-02	1.81
P00352	Retinal dehydrogenase 1	F2	ALDH1A1	4	2.36E-02	1.76
P35237	Serpin B6	F2	SERPINB6	2	2.85E-02	1.87
P02768	Serum albumin	F1	ALB	3	3.26E-02	5.15
P04083	Annexin A1	F1	ANXA1	3	3.92E-02	2.41
P08133	Annexin A6	F1	ANXA6	2	4.19E-02	2.58

The 8 altered human-specific proteins include 5 proteins with an increased expression in the F2 generation compared to the F1 PDX generation. These proteins potentially signify an aggressive cohort of proteins whose expression increased over two stages of engraftment. From this cohort, 4 out of 5 of the proteins (IDH1, CTSD, PRDX6 and SERPINB6) also showed a statistically significant increase in abundance (12, 2.3, 2.6 and 3-fold, respectively) in the patient tumour samples when compared to the PDX F1 generation in the species-indistinguishable study. Although their expression is not significantly increased in the F1 generation they are still present, and it is not until the second engraftment (F2) that their expression levels increase.

Species-indistinguishable proteins were also profiled in the F1 versus F2 PDX generations to investigate proteomic changes over the PDX generations. In this analysis,

41 proteins were identified as significantly differentially expressed, with 22 proteins having an increased expression in F2, and 19 having a decreased expression in F2 compared to F1.

This analysis demonstrates the stability of the human tumour proteome in the PDX mice models of pancreatic cancer. Of the human-specific proteins, only 5 show a significant change in their abundance levels within the F2 generation, and of those, none surpass a 2-fold difference, indicating the change is minimal and unlikely to be phenotypic. This analysis is the first to our knowledge that profiles the human-specific proteomic changes of patient derived tumours from the point of surgical resection through to the different PDX generations (Siolas and Hannon 2013).

3.3.5 Western blot analysis of CD55

Complement decay-accelerating factor (CD55) was identified as significantly differentially expressed between primary PDAC tumours and matched adjacent-normal tissues, with a 6.77-fold increase in expression in PDAC tissue. Western blot analysis was used to verify the presence of CD55 across representative samples from adjacent-normal, primary PDAC, PDX F1, and PDX F2 tumours (Figure 3.7). CD55 is expressed in all 5 of the primary PDAC tumours and in 4/5 of the PDX F1 and F2 tumour tissues. It is absent in all 5 adjacent-normal tissue analysed, suggesting it has strong specificity for PDAC. Coomassie stained gels to show loading are shown in supplementary figure 3. When analysed across the Badea microarray gene expression set (figure 3.3), CD55 was found to have a 4.065-fold increase in the PDAC tumours (n=39) compared to normal pancreas. These results taken together provide evidence that CD55 has potential as a target for PDAC.

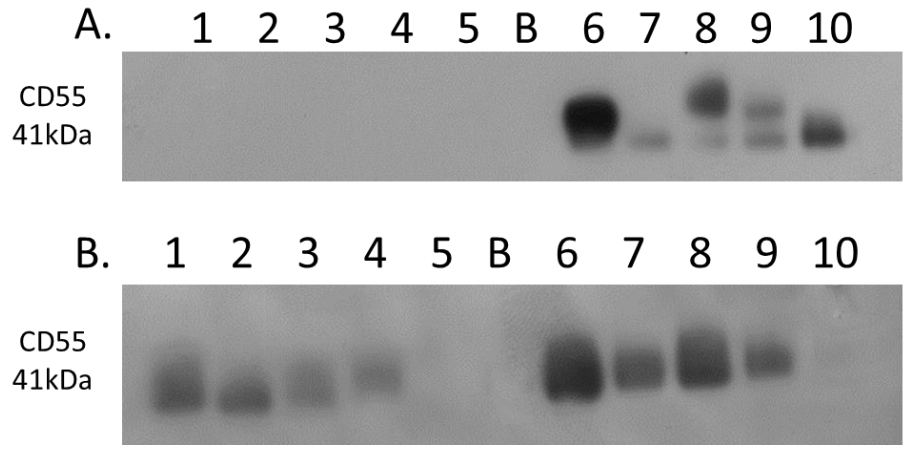


Figure 3.7: Western blot analysis of CD55 expression. A. Expression of CD55 across 5 representative samples from adjacent-normal and primary PDAC tumours (Lanes 1-5: 89N, 91N, 116N, 120N and 161N, Lane B: Blank, Lanes 6-10: 65T, 120T, 140T, 160T and 161T). CD55 is absent from the adjacent-normal tissue set and present in all 5 of the primary PDAC membrane-enriched fractions. B. Expression of CD55 across 5 representative membrane-enriched fractions from PDX F1 and F2 tumours (Lane 1-5:89F1, 116F1, 141F1, 160F, 161F1, Lane B: Blank, Lanes 6-10: 65F2, 80F2, 89F2, 99F2 and 160F2). CD55 is expressed in 4 out of 5 tumours for both F1 and F2 sets.

3.4 Discussion

Pancreatic cancer has traditionally been characterised in the clinic through expression of CA19-9 in the blood despite the median sensitivity of CA 19-9 for diagnosis being 79% and median specificity 82% (Goonetilleke and Siriwardena 2007). In addition, several factors influence CA19-9 levels, including the presence of jaundice, which elevates CA19-9, leading to false positives and also masking disease progression. The lack of sensitivity and specificity shown by CA19-9 demonstrates the need for novel diagnostic or prognostic biomarkers for PDAC.

The development of effective chemotherapies also remains a challenge in PDAC. Gemcitabine is the standard chemotherapy for PDAC and many other cancer types; its incorporation into the DNA leads to inhibition of DNA synthesis and ultimately cell death. Although it has been the first-line treatment in patients of PDAC its use alone achieves only modest survival benefits, with the 1-year survival rate being 18% in those treated with gemcitabine (Min et al. 2002). As most approved drugs target proteins, the evaluation of pancreatic cancer proteomes holds promise for identifying PDAC specific drug targets. The proteomic profiling of PDAC tumours in this study has established

discrete changes in the abundance and expression pattern of many proteins resulting from the disease process. Differential expression of such targets potentially could be manipulated using novel drug schemes, thus improving the treatment of pancreatic cancer, and/or could be employed as novel biomarkers for diagnosis or monitoring disease progression and the efficacy of treatment. Here, we have analysed the proteomic profiles of 9 adjacent-normal patient-matched control pancreatic specimens, 10 PDAC tumours and 20 matched pancreatic cancer patient-derived xenograft tumours. The quantitative label-free LC-MS-based proteomic analysis of membrane protein changes in these samples identified protein targets of PDAC which could potentially be used in a targeted therapeutic drug scheme such as ADC therapy. The sample preparation method used enriches for membrane proteins, but is not an absolute fractionation, so that proteins from other cellular compartments, in particular abundant cytoplasmic proteins such as enzymes, also appear in our lists. Although these proteins are not of interest as candidates for membrane-targeted therapeutics such as ADCs, some of the cytosolic proteins identified warrant future analyses to determine their role in PDAC development. Importantly, this study has established that PDX tumours recapitulate the proteomic signatures of human pancreatic cancers. This profile is narrowly altered in the subsequent F2 generation in PDX mice, which highlights their value as both therapeutic targets and for generating larger masses of tumour material for study than is typically available from the primary tumour sample.

From our analysis of the patient-matched adjacent-normal and pancreatic tumours, 238 proteins were identified as significantly differentially expressed with a minimum fold change of 1.5-fold between the experimental groups. Over 10 of these proteins exhibit a substantial increase in abundance levels with greater than 5-fold level of overexpression in tumours compared to the adjacent-normal group. Of these, integrin beta-6 a member of the integrin family, which is well-studied for its function of attaching the cell cytoskeleton to the extracellular matrix, thus explaining its presence in the membrane-enriched fraction of the pancreatic adenocarcinomas (Hood and Cheresh 2002, Calderwood, Shattil and Ginsberg 2000). Integrin beta-6 specifically functions as a receptor for fibronectin, which justifies the high levels of this protein that functions at the cell membrane in adhesion and motility (Calderwood, Shattil and Ginsberg 2000).

The identification and role of this integrin receptor as a fibronectin-binding protein was first characterised in a pancreatic carcinoma cell line FG-2 in 1992, highlighting its importance in this cancer (Busk, Pytela and Sheppard 1992). Thrombospondin 1 and 2, also overexpressed in PDAC tumours, are adhesive glycoproteins that mediate cell-to-cell and cell-to-matrix interactions (Lawler 2000). Complement decay-accelerating factor (CD55) has been identified by immunohistochemistry in pancreatic cancer and its elevated levels correlate positively with tumour aggressiveness, vascular invasion and prognosis (He et al. 2015). The increased levels of CD55 in the pancreatic adenocarcinoma specimens suggests a protective role to prevent tumour cells from bystander killing by complement (Morgan, Spendlove and Durrant 2002). We found a 5.58-fold increase in abundance of CD55 in the PDAC samples compared to the normal-adjacent samples from the proteomic analysis. This increased expression in the PDAC patients was validated by Western blotting across adjacent-normal and tumour samples and was also confirmed to maintain increased expression in the PDX models. TGF beta-1 induced transcript 1 protein (Hic-5) has not been reported to play a role in pancreatic cancer to date; however, it has been long studied in prostate cancer and its expression is suggested to correlate with tumorigenesis (Li, X. et al. 2011, Heitzer, M. D. and DeFranco 2006). More recently the expression of Hic-5 has been shown to be involved in the modulation of the stromal matrix in prostate and breast cancer as well as other models of the stromal matrix (Deakin and Turner 2011, Heitzer, Marjet D. and DeFranco 2007, Heitzer, M. D. and DeFranco 2006). These proteins, which exhibit a strongly significant and considerable increased abundance specific to the cohort of PDAC tumours, have potential as both novel drug targets and biomarkers for PDAC. In an attempt to validate our findings over a larger, separate cohort of PDAC data, we used the bioinformatic tool Oncomine, which houses microarray gene datasets from various studies worldwide. Four differentially expressed proteins were analysed against the Badea 2008 PDAC gene microarray dataset, which contains differential expression information for 400 genes across normal pancreas (n=39) and PDAC (n=39). In that data set, all four upregulated proteins (ITGB6, FN1, THBS2 and CD55) were similarly found to be overexpressed in PDAC compared to normal pancreas. The retrieval of these potential markers, which have been found previously by microarray analysis of a larger

PDAC cohort, shows the value of quantitative proteomic expression analysis to identify markers for pancreatic cancer that represent candidate proteins for drug targets of PDAC disease.

Proteomic analysis of the PDX tumours we generated proved challenging in terms of identifying the protein targets that are undeniably derived from the original patient PDAC tumour. By using human specific peptide sequences to identify patient tumour marker proteins, we were able to identify 32 significantly differentially expressed proteins having an increased abundance in the PDX F1 tumours compared to primary PDAC tumours. These proteins represent an exceptionally unique cohort of the PDAC proteome that has survived engraftment and also adapted successfully to the murine tumour microenvironment with higher expression levels than in the original tumour. Most interestingly, Periplakin (PPL) had the highest change in expression levels, with a 7.85-fold increase in PDX F1 compared to primary PDAC tumours. PPL is a component of desmosomes, which are present on the cell membrane and allow for cell-cell attachment. Although periplakin has been previously identified in pancreatic cancer, it has never been the focus of study (Kirby et al. 2016). PPL has been extensively studied as a binding partner of Akt (van den Heuvel et al. 2002), which is well established for its role in cell proliferation, migration, apoptosis and various cancers. This may indicate a role for PPL in PDAC tumour progression through Akt, and its increased expression in the PDX F1 tumour cohort suggests it may be driving tumour growth by increasing proliferation. This significantly increased expression in the PDX F1 tumours suggests an important role for this protein in PDAC tumorigenesis, which warrants future investigation.

This is the first study to profile the PDAC PDX tumour proteome through 2 generations after initial engraftment. This analysis identified 8 human-specific proteins and 41 species-indistinguishable proteins with significant differential expression. Of those 49 altered proteins, 30 were overexpressed in the F2 tumours compared to the F1 cohort. This analysis demonstrates that the PDAC PDX model used here largely retains the primary patients' PDAC proteomic profile over time and various generations and provides great encouragement for applying PDX models for successful PDAC research.

The comparison of adjacent normal tissues to primary PDAC tumours has the advantage that all proteins identified are human, but the limitation is that the tumours contain variable amounts of admixed normal cells (stromal cells, immune system cells etc) and this tends to dampen the discriminatory power of the differential expression analysis. The comparison of primary tumours vs xenograft tumours, when confined to "human-specific" may be the richer list for candidate target identification. These differentially expressed proteins are from a comparison of the primary tumour biopsy sample and the pure tumour cells as during the xenograft process, the human stromal cells are lost and replaced by mouse stroma and the latter does not contribute to the human-specific protein list. We the authors believe these proteins could be more reliable PDAC targets for future investigations.

3.5 Conclusions

Our investigation has identified proteins that are significantly differentially expressed in primary PDAC tumours compared to patient-matched adjacent-normal tissues. We present 129 proteins that are overexpressed in PDAC that have potential as therapeutic targets or biomarkers of disease. Importantly, we validated the specificity of 4 differentially expressed proteins across a microarray gene expression set from an independent cohort of PDAC patients. For the first time, proteomic profiling of PDAC PDX models, as described in this study, demonstrates that PDX tumours recapitulate the primary patients PDAC tumour over time and various generations. Through IHC, we established that much of the stromal component of the tumours is overcome by mouse cells once engrafted, and only human tumour cells are retained. This occurrence is advantageous and allowed us to identify tumour cell associated proteins in PDAC upon PDX engraftment. These proteins represent strong candidates for novel therapeutic targeting such as ADC therapy. Of course, the xenografts are not perfect models of human pancreatic cancer because they differ in anatomical location, host species differences and the presence of murine stroma, but they nevertheless provide a rich source of human-derived pancreatic cancer cells, *in situ* in a stromal environment, for molecular studies. This study is the first comparative proteomic analysis of PDAC that

employs PDX models to identify patient tumour cell-associated proteins to identify potential targets for novel therapeutic treatment of PDAC.

Author Contributions: Conceptualization, Paula Meleady and Martin Clynes; Formal analysis, Michael Henry; Investigation, Orla Coleman; Methodology, Fiona O'Neill, Sandra Roche, Ninfa L. Straubinger and Robert M. Straubinger; Resources, Niall Swan, Lorraine Boyle, Jean Murphy, Justine Meiller, Neil T. Conlon, Justin Geoghegan, Kevin Conlon, Vincent Lynch, Gerard McVey and Michael Moriarty; Supervision, Paula Meleady and Martin Clynes.

Funding: This publication has emanated from research supported in part by a research grant from Science Foundation (SFI) under the US-Ireland R&D Partnership Programme Grant number SFI/14/US/B2997. Also, with funding support from The Collaborative research programme on Radiation Biology and Proteomics of pancreatic cancer funded from St Luke's Institute for Cancer Research, Dublin.

3.6 References

Badea L, Herlea V, Dima SO, Dumitrascu T, Popescu I. Combined gene expression analysis of whole-tissue and microdissected pancreatic ductal adenocarcinoma identifies genes specifically overexpressed in tumor epithelia. Hepatogastroenterology. 2008 Nov-Dec;55(88):2016-27.

Balaban EP, Mangu PB, Yee NS. Locally advanced unresectable pancreatic cancer: American Society of Clinical Oncology clinical practice guideline summary. Journal of oncology practice. 2016;13(4):265-9.

Beck A, Goetsch L, Dumontet C, Corvaïa N. Strategies and challenges for the next generation of antibody–drug conjugates. Nature reviews Drug discovery. 2017;16(5):315.

Berberat PO, Friess H, Wang L, Zhu Z, Bley T, Frigeri L, et al. Comparative analysis of galectins in primary tumors and tumor metastasis in human pancreatic cancer. J Histochem Cytochem. 2001 Apr;49(4):539-49.

Burris HA,3rd, Moore MJ, Andersen J, Green MR, Rothenberg ML, Modiano MR, et al. Improvements in survival and clinical benefit with gemcitabine as first-line therapy for patients with advanced pancreas cancer: a randomized trial. J Clin Oncol. 1997 Jun;15(6):2403-13.

Busk M, Pytela R, Sheppard D. Characterization of the integrin alpha v beta 6 as a fibronectin-binding protein. J Biol Chem. 1992 Mar 25;267(9):5790-6.

Calderwood DA, Shattil SJ, Ginsberg MH. Integrins and actin filaments: reciprocal regulation of cell adhesion and signaling. J Biol Chem. 2000 Jul 28;275(30):22607-10.

Chen R, Brentnall TA, Pan S, Cooke K, Moyes KW, Lane Z, et al. Quantitative proteomics analysis reveals that proteins differentially expressed in chronic pancreatitis are also frequently involved in pancreatic cancer. Mol Cell Proteomics. 2007 Aug;6(8):1331-42.

Chen R, Eugene CY, Donohoe S, Pan S, Eng J, Cooke K, et al. Pancreatic cancer proteome: the proteins that underlie invasion, metastasis, and immunologic escape. Gastroenterology. 2005;129(4):1187-97.

Chevet E, Fessart D, Delom F, Mulot A, Vojtesek B, Hrstka R, et al. Emerging roles for the pro-oncogenic anterior gradient-2 in cancer development. Oncogene. 2013;32(20):2499.

Coleman O, Henry M, McVey G, Clynes M, Moriarty M, Meleady P. Proteomic strategies in the search for novel pancreatic cancer biomarkers and drug targets: recent advances and clinical impact. Expert review of proteomics. 2016;13(4):383-94.

Conroy T, Hammel P, Hebbar M, Ben Abdelghani M, Wei AC, Raoul J, et al. Unicancer GI PRODIGE 24/CCTG PA.6 trial: A multicenter international randomized phase III trial of adjuvant mFOLFIRINOX versus gemcitabine (gem) in patients with resected pancreatic ductal adenocarcinomas. JCO. 2018 06/20; 2018/08;36(18):LBA4001.

Delitto D, Pham K, Vlada AC, Sarosi GA, Thomas RM, Behrns KE, et al. Patient-derived xenograft models for pancreatic adenocarcinoma demonstrate retention of tumor

morphology through incorporation of murine stromal elements. The American journal of pathology. 2015;185(5):1297-303.

Deakin NO, Turner CE. Distinct roles for paxillin and Hic-5 in regulating breast cancer cell morphology, invasion, and metastasis. Mol Biol Cell. 2011 Feb 1;22(3):327

Dumartin L, Whiteman HJ, Weeks ME, Hariharan D, Dmitrovic B, Iacobuzio-Donahue CA, et al. AGR2 is a novel surface antigen that promotes the dissemination of pancreatic cancer cells through regulation of cathepsins B and D. Cancer Res. 2011 Nov 15;71(22):7091-102.

Goonetilleke K, Siriwardena A. Systematic review of carbohydrate antigen (CA 19-9) as a biochemical marker in the diagnosis of pancreatic cancer. European Journal of Surgical Oncology (EJSO). 2007;33(3):266-70.

Gronborg M, Kristiansen TZ, Iwahori A, Chang R, Reddy R, Sato N, et al. Biomarker discovery from pancreatic cancer secretome using a differential proteomic approach. Mol Cell Proteomics. 2006 Jan;5(1):157-71.

Haun RS, Quick CM, Siegel ER, Raju I, Mackintosh SG, Tackett AJ. Bioorthogonal labeling cell-surface proteins expressed in pancreatic cancer cells to identify potential diagnostic/therapeutic biomarkers. Cancer biology & therapy. 2015;16(10):1557-65.

He Z, Wu H, Jiao Y, Zheng J. Expression and prognostic value of CD97 and its ligand CD55 in pancreatic cancer. Oncol Lett. 2015 Feb;9(2):793-7.

Heinemann V. Gemcitabine: progress in the treatment of pancreatic cancer. Oncology. 2001;60(1):8-18.

Heitzer MD, DeFranco DB. Hic-5/ARA55 a prostate stroma-specific AR coactivator. Steroids. 2007;72(2):218-20.

Heitzer MD, DeFranco DB. Hic-5/ARA55, a LIM domain-containing nuclear receptor coactivator expressed in prostate stromal cells. Cancer Res. 2006 Jul 15;66(14):7326-33.

Hood JD, Cheresh DA. Role of integrins in cell invasion and migration. Nature Reviews Cancer. 2002;2(2):91.

Kirby MK, Ramaker RC, Gertz J, Davis NS, Johnston BE, Oliver PG, et al. RNA sequencing of pancreatic adenocarcinoma tumors yields novel expression patterns associated with long-term survival and reveals a role for ANGPTL4. Molecular oncology. 2016;10(8):1169-82.

Kim Y, Han D, Min H, Jin J, Yi EC, Kim Y. Comparative proteomic profiling of pancreatic ductal adenocarcinoma cell lines. Mol Cells. 2014 Dec 31;37(12):888-98.

Lambert JM, Morris CQ. Antibody–drug conjugates (ADCs) for personalized treatment of solid tumors: A review. Adv Ther. 2017;34(5):1015-35.

Lawler J. The functions of thrombospondin-1 and-2. Curr Opin Cell Biol. 2000;12(5):634-40.

Li X, Martinez-Ferrer M, Botta V, Uwamariya C, Banerjee J, Bhowmick NA. Epithelial Hic-5/ARA55 expression contributes to prostate tumorigenesis and castrate responsiveness. Oncogene. 2011 Jan 13;30(2):167-77.

Linge A, Maurya P, Friedrich K, Baretton GB, Kelly S, Henry M, et al. Identification and functional validation of RAD23B as a potential protein in human breast cancer progression. Journal of proteome research. 2014;13(7):3212-22.

Liu X, Zhang M, Go VLW, Hu S. Membrane proteomic analysis of pancreatic cancer cells. J Biomed Sci. 2010;17(1):74.

Maestranzi S, Przemioslo R, Mitchell H, Sherwood R. The effect of benign and malignant liver disease on the tumour markers CA19-9 and CEA. Annals of Clinical Biochemistry: An international journal of biochemistry in medicine. 1998;35(1):99-103.

Min YJ, Joo KR, Park NH, Yun TK, Nah YW, Nam CW, et al. Gemcitabine therapy in patients with advanced pancreatic cancer. Korean J Intern Med. 2002 Dec;17(4):259-262

Morgan J, Spendlove I, Durrant LG. The role of CD55 in protecting the tumour environment from complement attack. Tissue Antigens. 2002;60(3):213-23.

O'lorcain P, Deady S, Comber H. Mortality predictions for esophageal, stomach, and pancreatic cancer, Ireland, up to 2015. International journal of gastrointestinal cancer. 2006;37(1):15-25.

Orozco CA, Martinez-Bosch N, Guerrero PE, Vinaixa J, Dalotto-Moreno T, Iglesias M, et al. Targeting galectin-1 inhibits pancreatic cancer progression by modulating tumor-stroma crosstalk. Proc Natl Acad Sci U S A. 2018 Apr 17;115(16):E3769-78.

Pan S, Brentnall TA, Kelly K, Chen R. Tissue proteomics in pancreatic cancer study: discovery, emerging technologies, and challenges. Proteomics. 2013;13(3-4):710-21.

Ramaswamy S, Tamayo P, Rifkin R, Mukherjee S, Yeang CH, Angelo M, et al. Multiclass cancer diagnosis using tumor gene expression signatures. Proc Natl Acad Sci U S A. 2001 Dec 18;98(26):15149-54.

Ryan DP, Hong TS, Bardeesy N. Pancreatic adenocarcinoma. N Engl J Med. 2014;371(11):1039-49.

Siolas D, Hannon GJ. Patient-derived tumor xenografts: transforming clinical samples into mouse models. Cancer Res. 2013 Sep 1;73(17):5315-9.

Shen J, Person MD, Zhu J, Abbruzzese JL, Li D. Protein expression profiles in pancreatic adenocarcinoma compared with normal pancreatic tissue and tissue affected by pancreatitis as detected by two-dimensional gel electrophoresis and mass spectrometry. Cancer Res. 2004 Dec 15;64(24):9018-26.

Takahashi Y, Takeuchi T, Sakamoto J, Touge T, Mai M, Ohkura H, et al. The usefulness of CEA and/or CA19-9 in monitoring for recurrence in gastric cancer patients: a prospective clinical study. Gastric cancer. 2003;6(3):142-5.

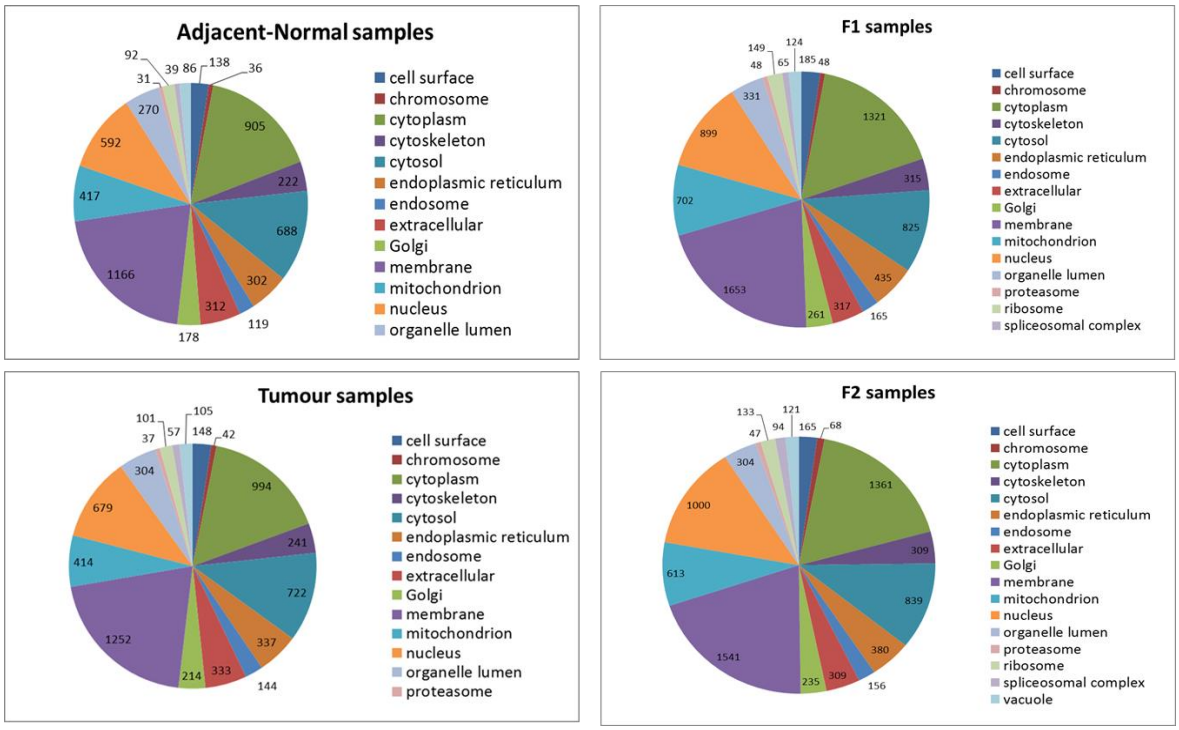
van den Heuvel AP, de Vries-Smits AM, van Weeren PC, Dijkers PF, de Bruyn KM, Riedl JA, et al. Binding of protein kinase B to the plakin family member periplakin. J Cell Sci. 2002 Oct 15;115(Pt 20):3957-66.

Wang W, Zhang T, Zhao W, Xu L, Yang Y, Liao Q, et al. A Single Talent Immunogenic Membrane Antigen and Novel Prognostic Predictor: voltage-dependent anion channel 1 (VDAC1) in Pancreatic Cancer. Scientific reports. 2016;6:33648.

3.7 Appendix A

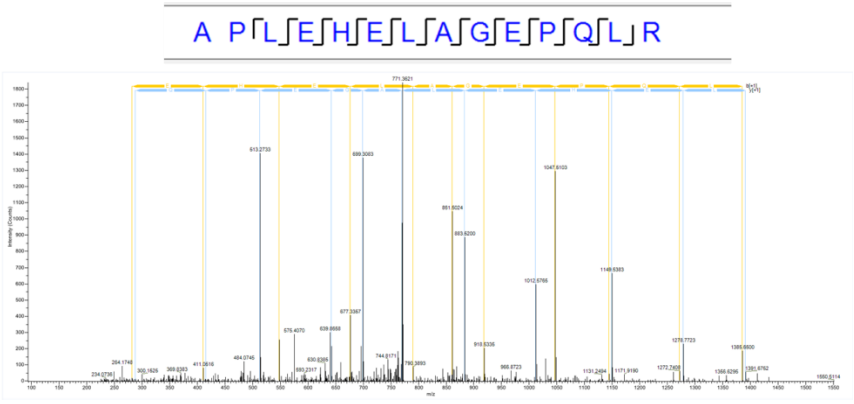
Supplementary tables S2 and S3 are provided on the disk attached to the back cover of the thesis.

Supplementary Figure 1. Cellular component annotation using Gene Ontology analysis through the ProteinCenter node in Proteome Discoverer for each of the sample sets.

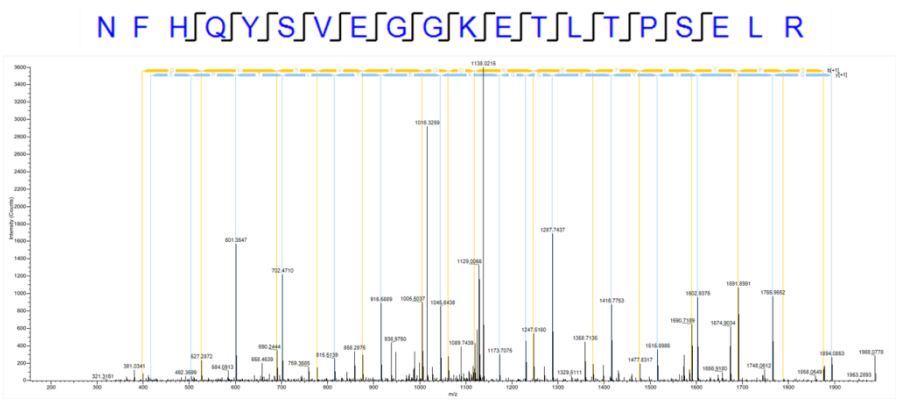


The efficiency of membrane protein enrichment was assessed using gene ontology annotation by the ProteinCenter software. Supplementary figure 1 shows the number of proteins from each sample set in relation to their cellular component annotation. Membrane-associated proteins relate to 48.8%, 66.5%, 47.8% and 59.5% of the total proteins identified for adjacent-normal, tumour, PDX F1 and PDX F2 samples.

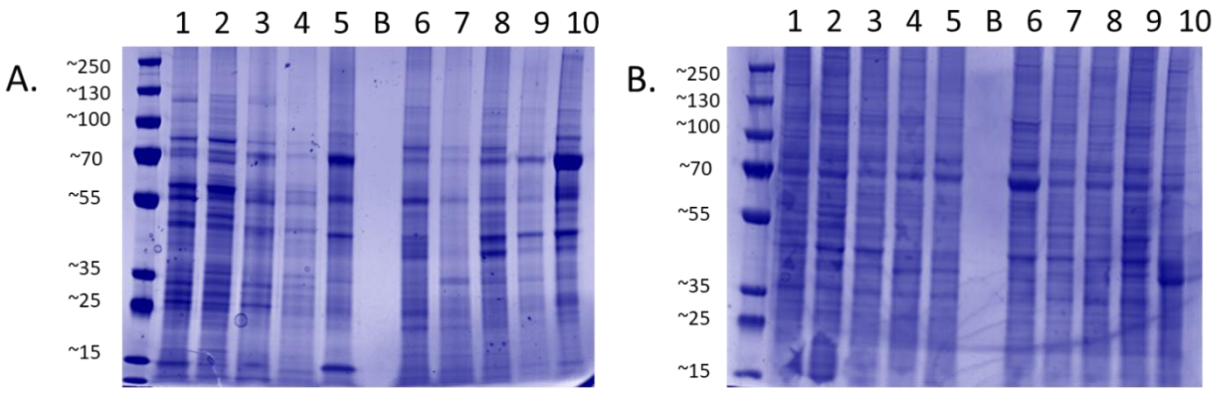
CLIC-3 (77% identity)
Human: APLEHELAGEPQL
Mouse: APLDHELAQEPHL



Protein S100-A14 (80% identity)
Human: NFHQYSVEGGKETLTPSELR
Mouse: NFHKYSVAGKKETLTPAELR



Supplementary figure 2: MS/MS spectra of two human-specific peptides identified from the PDAC PDX F1 generation. Sequence alignments using BLASTp of the identified human peptides to their corresponding mouse peptides are shown. The protein sequence homology between the two species is indicated.



Supplementary figure 3: Coomassie stained gels validating loading of lanes prior to the Western blot of CD55. Coomassie stained gel of (A) Five representative adjacent-normal tissues and primary PDAC tumours membrane protein enriched fractions (Lanes 1-5: 89N, 91N, 116N, 120N and 161N, Lane B is a Blank, Lanes 6-10: 65T, 120T, 140T, 160T and 161T). Coomassie stained gel of (B) 5 representative membrane-enriched fractions from PDX F1 and F2 tumours (Lane 1-5:89F1, 116F1, 141F1, 160F, 161F1, Lane B: Blank, Lanes 6-10: 65F2, 80F2, 89F2, 99F2 and 160F2).

				Type of tissue specimen analysed			
Patient I.D.	**Original Patient**	**Gender**	**Age at surgical resection**	**Adjacent-Normal**	**Tumour**	**F1**	**F2**
PIN 65	Tumour & normal invasive ductal adenocarcinoma	F	76	✓	✓	✓	✓
PIN 80	Tumour & normal invasive ductal adenocarcinoma	M	70	✓	✓	✓	✓
PIN 89	Tumour & normal invasive ductal adenocarcinoma	F	45	✓	✓	✓	✓
PIN 91	Tumour & normal invasive ductal adenocarcinoma	M	64	✓	✓	✓	✓
PIN 99	Tumour & normal invasive ductal adenocarcinoma	F	63	-	-	✓	✓
PIN 116	Tumour & normal invasive ductal carcinoma	M	62	✓	✓	✓	-
PIN 120	Invasive ductal adenoca–HOP (head of Pancreas)	M	65	✓	✓	✓	✓
PIN 140	Tumour & normal invasive ductal carcinoma	M	63	✓	✓	-	✓
PIN 141	Tumour & normal invasive ductal adenocarcinoma	M	82	✓	✓	✓	✓
PIN 160	Tumour & normal invasive ductal adenocarcinoma	F	77	✓	✓	✓	✓
PIN 161	Tumour & normal invasive ductal adenocarcinoma	M	61	✓	✓	✓	✓

Supplementary Table S1: Characteristics of patients and tissue samples used for LC-MS/MS analysis

4 Chapter Four: Filter-Aided Sample Preparation (FASP) for improved proteome analysis of recombinant Chinese hamster ovary cells

Published in *Methods in Molecular Biology: Heterologous Protein Production in CHO cells* May 2017

DOI: 10.1007/978-1-4939-6972-2_12

Authors: Orla Coleman, Paula Meleady, Martin Clynes and Michael Henry

For this book chapter, I adapted and optimised the FASP method from the original method for its use in CHO proteomic analyses. The chapter was prepared, written and submitted by me.

Abstract

Sample solubilisation is a key step in the extraction of protein from any host cell prior to downstream proteomic analyses. Solubilising detergents often interfere with enzymatic digestions and hamper HPLC-MS experimental runs thus making their removal necessary. Conventional detergent removal involves crude methods using organic compounds to precipitate the protein from the detergent solution however; significant sample loss is associated with these methods. The introduction of FASP has circumvented the use of deleterious reagents and is reported to improve proteome coverage. FASP cleverly uses a common ultrafiltration device whereby the membrane pores are small enough to allow contaminating detergents to pass through but proteins are too large and are retained and concentrated in the filter unit. This method of sample preparation and digestion is universally applicable and can be easily employed in any proteomics facilities as standard everyday laboratory reagents and equipment are used. The following chapter provides a comprehensive protocol describing the step-by-step method employed in our proteomics facility for the FASP preparation of recombinant CHO cells.

4.1 Introduction

Chinese hamster ovary (CHO) cells are the most commonly used host cell line for the production of recombinant bio-therapeutics. A greater understanding of the biology of these cells is required in order to improve the efficiency of production with the overall aim to reduce the costs of therapeutics. Systems biology approaches are being increasingly used to characterise recombinant CHO cells in order to achieve this goal. This chapter outlines a recently described method, Filter-Aided Sample Preparation (FASP) (Wiśniewski et al., 2009) which enables protein purification using a standard ultrafiltration device for efficient detergent removal prior to downstream proteomic analyses.

Detergents are the reagents of choice for solubilisation of cells and tissue; however detergents used for proteomic applications can supress enzymatic digestions, contaminate HPLC instrumentation, interfere with column binding and column elution, and because they are readily ionisable can dominate mass spectra. In any proteomics experiment the choice of detergent is a balance of the need to solubilise and stabilise the protein without deleteriously affecting downstream processing and analysis. The detergent should not interfere with downstream techniques but often denaturing detergents used for protein extractions do, therefore their removal is required prior to further processing.

In this chapter we describe a recently introduced method, filter-aided sample preparation (FASP) which can be used to purify protein samples (Wiśniewski et al., 2009). This method cleverly uses a common ultrafiltration device for detergent removal to efficiently purify the proteome for subsequent analysis. The use of a filter device removes the need for protein precipitation as a means of detergent removal which often results in significant sample loss. Sample solubilisation is achieved using SDS; following this the protein is retained, washed and concentrated in the filter device. The filter device then acts as a “proteomic reactor” for detergent removal, buffer exchange,

chemical modification and protein digestion. The underlying principle of this method is that low-molecular-weight detergents and impurities are small enough to pass through the filter membrane while high-molecular-weight substances such as proteins are retained. After protein digestion the digested peptides are small enough to pass through the filter membrane and be collected for mass spectrometry analysis. The efficient removal of impurities and detergents allows for a more complete coverage of the proteome. The FASP method has been successfully applied to various diverse proteomic analyses such as FFPE cancer tissues, brain phosphoproteome, *Escherichia coli* and modified versions for glycoproteomic studies (Deeb et al., 2014; Erde et al., 2014; Wiśniewski et al., 2010, 2011). It has also been used to maximize protein recovery during the CHO proteome analysis carried out by Baycin-Hizal *et al*. who identified a total of 6164 proteins from both glycoproteome and proteome analysis (Baycin-Hizal et al., 2012). This chapter provides a step-by-step comprehensive protocol for improved proteome analysis of recombinant CHO cells using a combination of FASP preparation and strong cation exchange (SCX) for efficient fractionation and identification of the complex proteome.

4.2 Materials

4.2.1 Equipment

1. Microcon-10kDa Centrifugal Filter Unit with Ultracel-10 membrane (Merck Millipore).
2. Microcentrifuge capable of spinning 1.7 mL microcentrifuge tubes at 14,000 x*g* at 20°C.
3. Pierce® C18 100 µL tips (Thermo Fisher Scientific).
4. Sonicating probe.
5. Microplate Spectrophotometer reading 96-well plate format at 595 nm (e.g. Multiskan™ GO, Thermo Fisher Scientific).
6. SpeedVac Vacuum Dryer.
7. Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific).

8. Pierce™ Strong Cation Exchange Spin Column, Mini (Thermo Fisher Scientific).
9. Dry Heating Block.
10. pH meter.

4.2.2 Reagents

1. TRIZMA® HCl.
2. TRIZMA® Base.
3. Sodium dodecyl sulfate (SDS).
4. DL-Dithiothreitol (DTT).
5. Iodoacetamide.
6. Urea.
7. LC-MS grade water.
8. Sequencing-grade modified trypsin (e.g. Promega, Sigma Aldrich).
9. ProteaseMax^{TM} Surfactant Trypsin Enhancer (Promega).
10. Phosphate Buffered Saline.
11. Trifluoroacetic acid (TFA).
12. Acetonitrile (ACN).
13. Sodium Chloride (NaCl).
14. Ammonium Bicarbonate.
15. Potassium Phosphate Monobasic (KH_2PO_4).
16. Potassium Chloride (KCl).

4.2.3 Buffer Preparation for FASP

1. Tris-HCl stock solutions: The following Tris-HCl mixing table is used for the preparation of 0.1 M stock solutions of varying pH's in 50 mL which is used to prepare the lysis and wash buffers for FASP. The precise blend of Trizma HCl and Trizma Base will bring the solutions to the desired pH, however this must be confirmed using a pH meter.

Table 4.1 Preparation of Tris-HCl solutions with varying pH

pH	Trizma-HCl (g)	Trizma Base (g)	LC-MS Water (mL)
7.6	0.606	0.139	50
7.9	0.488	0.230	50
8.5	0.221	0.436	50

2. Lysis Buffer: 4% SDS, 0.1 M DTT solution. To make a 10 mL solution of the lysis buffer add 0.4 g of high purity SDS and 0.0154 g of DTT to 10 mL of 0.1 M Tris-HCl stock solution at pH 7.6. Take caution when weighing SDS by using a fume hood or a dust mask. Use 1 mL of this lysis buffer per 2 x 10^6 CHO cells.
3. Wash buffer 1: A solution of 8 M urea in 0.1 M Tris-HCl pH 8.5. Add 0.4 g of urea to 1 mL of the 0.1 M Tris-HCl pH 8.5 stock solution.
4. Wash buffer 2: A solution of 8 M urea in 0.1 M Tris-HCl pH 7.9. Add 0.4 g of urea to 1 mL of the 0.1 M Tris-HCl pH 7.9 stock solution.
5. Alkylation solution: 0.05 M Iodoacetamide in 8 M Urea, 0.1 M Tris-HCl pH 8.5. To prepare 1 mL of this solution add 0.4 g of urea and 0.009 g of iodoacetamide to 1 mL of the 0.1 M Tris-HCl pH 8.5 stock solution.
6. Rinse solution: Add 0.029 g of NaCl to 1 mL of LC-MS grade water.

4.2.4 Protein Digestion

1. 50 mM Ammonium Bicarbonate (must be freshly prepared and used within 24 hours).
2. Trypsin solution: Use sequence grade trypsin (*see* **Note 1**).
3. ProteaseMax™ Surfactant Enhancer (*see* **Note 2**).

4.2.5 Peptide purification

1. Sample acidification buffer: Prepare a solution of 2.5% TFA in LC-MS grade water.
2. Wetting solution: Prepare 50% ACN in LC-MS grade water.
3. Equilibration and rinse solution: 0.1% TFA, 2% ACN
4. Elution solution: 0.1% TFA, 70% ACN.

4.2.6 Peptide Fractionation using SCX spin cartridges

1. 10 mM KH_2PO_4 in 25% acetonitrile pH 3.0 stock (diluent) solution: To prepare 1 L of this stock (diluent) solution add 1.36 g of KH_2PO_4 and 250 mL of acetonitrile to 750 mL of LC-MS grade water and adjust the pH to 3.0.
2. Prepare the following 10 mL Potassium Chloride (KCl) elution buffers using the 10 mM KH_2PO_4 in 25% acetonitrile pH 3.0 diluent.

Table 4.2 Overview of preparation of KCl elution buffers

Elution Buffer	KCl (g)	10 mM KH_2PO_4 in 25% ACN (mL)	Final KCl concentration
Elution buffer step 1	0.007	10	10 mM
Elution buffer step 2	0.019	10	25 mM
Elution buffer step 3	0.037	10	50 mM
Elution buffer step 4	0.056	10	75 mM
Elution buffer step 5	0.075	10	100 mM
Elution buffer step 6	0.093	10	125 mM
Elution buffer step 7	0.112	10	150 mM
Elution buffer step 8	0.149	10	200 mM
Elution buffer step 9	0.224	10	300 mM
Elution buffer step 10	0.373	10	500 mM

4.3 Methods

4.3.1 Cell Lysis

1. Harvest CHO cells and wash with sterile Phosphate Buffered Saline three times. Centrifuge at 1000 x *g* for 5 min and remove the supernatant. Snap freeze the cell pellet in a microcentrifuge tube using liquid nitrogen (take extreme caution when doing this). Store at -80°C until cell lysis is performed.
2. Cell pellets corresponding to 2 x 10^6 cells (approximately 1 mg pellet of cells) are lysed with 1 mL of lysis buffer. Suspend the cell pellet in the lysis buffer and mix the solution well using a pipette. Further disrupt the cells using a sonicating probe while maintaining the cell lysate at 4°C.
3. Heat the lysate for 20 min at 56°C using a heating block to denature the protein.

4. Determine protein concentration using a BCA assay as per manufacturer's instructions.
5. Transfer 100 µg protein lysate aliquots into microcentrifuge tubes and freeze at -80ºC until required.

4.3.2 Filter Aided Sample Preparation (FASP)

1. Mix 100 µg of protein lysate with 200 µL of 8 M Urea
2. Transfer the solution to a Microcon-10kDa Centrifugal Filter Unit and spin using a microcentrifuge set to 14,000 x *g* at 20ºC for 40 min.
3. Discard the filtrate (flow-through) and dilute the concentrate with 200 µL of Wash buffer 1 and spin using a microcentrifuge set at 14,000 x *g* at 20ºC for 40 min.
4. Discard the filtrate and alkylate the concentrate with 100 µL of the alkylation solution.
5. Centrifuge the device at 14,000 x *g* at 20ºC for 40 min.
6. Discard the filtrate and dilute the concentrate with 100 µL of Wash buffer 2 and centrifuge at 14,000 x *g* at 20ºC for 40 min.
7. Repeat step 6
8. Proceed directly to protein digestion in **subsection 4.3.3**.

4.3.3 Protein Digestion

1. Dilute the concentrate with 100 µL of 50 mM Ammonium Bicarbonate.
2. Add 5 µg of sequence-grade trypsin (1:20 enzyme: protein).
3. Add 1 µL of 1% ProteaseMax^{TM} Surfactant.
4. Digest at 37 ºC for 3 h, *see* **Note 5.**
5. Transfer the Microcon-10kDa Centrifugal Filter Unit to a new collection tube and centrifuge at 14,000x*g* at 20ºC for 40 min. Retain the filtrate in the collection tube.
6. The FASP ultrafiltration device must be rinsed post-digestion to elute residual peptides that may be bound to the filter membrane or walls of the device. Rinse

the filter unit with 50 µL of the 0.5 M NaCl rinse solution and spin at 14,000 x *g* at 20ºC for 20 min.

7. Store the filtrate which contains the peptides at -80ºC until required.

4.3.4 Peptide purification

1. Add 6 µL of sample acidification buffer to the peptide filtrate to give a final concentration of 0.1% TFA.
2. Aspirate the 100 µL C18 tip using 100 µL of wetting solution and discard the solvent flow-through.
3. Repeat step 2.
4. Equilibrate the C18 tip using 100 µL of the equilibration buffer and discard the solvent flow-through.
5. Repeat step 4.
6. Aspirate and dispense the peptide sample into the C18 tip ten times to allow peptide binding (*see* **Note 8**).
7. Wash the C18 tip using 100 µL of the rinse solution and discard rinse wash.
8. Repeat step 7.
9. Slowly elute the peptide sample using 50 µL of elution buffer into a new microcentrifuge tube.
10. Repeat step 9 to give a final volume of 100 µL.
11. Concentrate the peptide sample to dryness in a SpeedVac vacuum dryer.
12. Freeze peptide sample at -80ºC or proceed directly to Strong Cation Exchange preparation.

4.3.5 Strong Cation Exchange (SCX) using Pierce™ Mini Spin Columns

1. Condition SCX spin column with 400 µL of 10mM KH_2PO_4 in 25% ACN pH 3.0 and centrifuge at 2,000 x *g* for 5 min. Discard the flow-through.
2. Suspend the dried peptides from the final step of **subsection 4.3.4** in 200 µL of 10 mM KH_2PO_4 in 25% ACN pH 3.0. Apply this to the conditioned SCX column, centrifuge at 2,000 x*g* for 5 min.

3. Add 200 µL of elution buffer step 1, centrifuge at 2,000 x *g* for 5 min and collect the flow-through.
4. Repeat this procedure to elute the peptides in a step-wise manner using elution buffer step 2 through to elution buffer step 10, collecting each flow-through.
5. Proceed directly to desalting using peptide purification in **subsection 4.3.4** or freeze at -80 ºC Once desalting is complete the peptide samples are ready for LC-MS analysis.

Notes

1. Use sequencing-grade or mass spectrometry grade trypsin. Reconstitute the lyophilised trypsin in 50 mM acetic acid to give a final concentration of 1 µg/µL and store aliquots at -20ºC until required. In our lab we use Promega's Trypsin Gold Mass Spectrometry grade or Pierce™ Trypsin Protease MS grade for protein digestion.
2. Add 100 µL of 50 mM ammonium bicarbonate to a 1 mg vial of ProteaseMax™ Surfactant to a give a 1% solution. Store aliquots at -20ºC until required. Ammonium Bicarbonate must be freshly prepared and used within 24 hours.
3. From our use of this method we advise to prepare the protein lysate and determine the protein concentration on day 1. Perform FASP and digestion the following day and peptide purification and fractionation on day 3.
4. In our experience, to reduce sample loss due to evaporation during sample digestion we recommend securing the lids of the Microcon-10kDa Centrifugal Filter Units using Parafilm wrap.
5. All solvents and water used must be LC-MS grade. All chemicals must be of the highest purity.
6. For peptide purification using C18 tips, ensure that air is not drawn into the tip and the C18 sorbent does not dry during sample processing.
7. We use ProteaseMax™ Surfactant Trypsin Enhancer to ensure efficient protein digestion with trypsin during the 3-hour digestion period.

8. For larger volumes of peptide sample when purifying using the C18 tips, repeat step 3.4.6 as many times as necessary to process the entire sample and ensure peptide binding.
9. For long term storage of the 0.1 M Tris-HCl stock solutions, store at 4ºC when not in use.

4.4 References

Baycin-Hizal, Deniz, David L. Tabb, Raghothama Chaerkady, Lily Chen, Nathan E. Lewis, Harish Nagarajan, Vishaldeep Sarkaria, et al. 2012. "Proteomic Analysis of Chinese Hamster Ovary Cells." Journal of Proteome Research 11 (11): 5265–76. https://doi.org/10.1021/pr300476w.

Deeb, Sally J., Juergen Cox, Marc Schmidt-Supprian, and Matthias Mann. 2014. "N-Linked Glycosylation Enrichment for In-Depth Cell Surface Proteomics of Diffuse Large B-Cell Lymphoma Subtypes." Molecular & Cellular Proteomics 13 (1): 240–51. https://doi.org/10.1074/mcp.M113.033977.

Erde, Jonathan, Rachel R. Ogorzalek Loo, and Joseph A. Loo. 2014. "Enhanced FASP (EFASP) to Increase Proteome Coverage and Sample Recovery for Quantitative Proteomic Experiments." Journal of Proteome Research 13 (4): 1885–95. https://doi.org/10.1021/pr4010019.

Wiśniewski, Jacek R., Nagarjuna Nagaraj, Alexandre Zougman, Florian Gnad, and Matthias Mann. 2010. "Brain Phosphoproteome Obtained by a FASP-Based Method Reveals Plasma Membrane Protein Topology." Journal of Proteome Research 9 (6): 3280–89. https://doi.org/10.1021/pr1002214.

Wiśniewski, Jacek R., Pawel Ostasiewicz, and Matthias Mann. 2011. "High Recovery FASP Applied to the Proteomic Analysis of Microdissected Formalin Fixed Paraffin Embedded Cancer Tissues Retrieves Known Colon Cancer Markers." Journal of Proteome Research 10 (7): 3040–49. https://doi.org/10.1021/pr200019m.

Wiśniewski, Jacek R., Alexandre Zougman, Nagarjuna Nagaraj, and Matthias Mann. 2009. "Universal Sample Preparation Method for Proteome Analysis." Nature Methods 6 (5): 359.

5 Chapter Five: Depletion of endogenous miRNA-378-3p increases peak cell density of CHO DP12 cells and is correlated with elevated levels of ubiquitin carboxyl-terminal hydrolase 14

Published in *Journal of Biotechnology* December 2018

DOI: 10.1016/j.jbiotec.2018.10.008

Authors: Orla Coleman, Alan Costello, Nga T. Lao, Michael Henry, Paula Meleady, Niall Barron and Martin Clynes.

I share joint primary authorship for this manuscript with Alan Costello. For this manuscript I was responsible for all proteomic analyses. This included sample collection, subcellular protein enrichment, protein preparation for MS analysis and bioinformatic data analyses. Jointly we refined a list of candidate proteins for further analysis. I wrote and prepared the proteomic subsections of the manuscript.

Abstract

miRNAs are potent molecular regulators of cellular behaviour. The manipulation of these small non-coding RNAs has been used to enhance industrially relevant phenotypes in Chinese Hamster Ovary (CHO) cells. We investigated the stable depletion of six miRNAs; miR-204-5p, 338-3p, 378-3p, 409-3p, 455-3p and 505-3p, robustly associated with cell growth rate from a previous profiling study. Inhibition of endogenous miR-378-3p function by miRNA-sponge-decoy improved peak cell density by 59%. Quantitative label free LC-MS/MS proteomic analysis of the fractionated cell cultures at day 4 and 8 of batch culture found 216 cytosolic and 114 membrane-associated proteins differentially expressed with stable miR-378-3p depletion. qRT-PCR of 8 genes; Clic4, Hnrnpa1, Prdx1, Actn4, Usp14, Srxn1, Canx and Gnb1, with unidirectional differential protein expression over the two time points of analysis was carried out. In-silico predictive algorithms; TargetScan and miRDB, were used to decipher possible direct targets of miR-378-3p. The Ubiquitin carboxyl-terminal hydrolase 14 (Usp14) protein was identified in the cytosolic fractions at both time-points as differentially expressed with an increased abundance of 1.58-fold in the miR-378-3p depleted cells on day 8. Usp14 is a deubiquitinase (DUB) with previous reports of its up-regulation leading to increased proliferation of cancer cells. Overexpression of Usp14 in CHO cells had significant effects on cell growth supporting a role of Usp14 in the increased peak cell density seen with miR-378-3p depletion. This study highlights miR-378-3p as a novel engineering candidate for improving CHO cell growth. The use of sub-cellular fractionation also improved proteome coverage in the identification of novel miRNA targets.

5.1 Introduction

Targeted genetic engineering of Chinese Hamster Ovary (CHO) cells has become more attainable since publication of the Chinese Hamster and CHO-K1 genomes (Xu et al., 2011; Brinkrolf et al., 2013; Lewis et al., 2014). The use of micro-RNA (miRNA) manipulation is one approach which has been promising. The potency of miRNAs comes from their small ~7-8nt recognition sequence (Bartel et al., 2009) enabling them to post-transcriptionally regulate the expression of multiple protein coding genes or entire pathways (Hackl et al., 2012). Constitutive miRNA manipulations have come in the form of exogenous over-expression of endogenous pri-miRNA hairpins (Klanert et al., 2014) or depletion of endogenous species by miRNA sponge-decoy (Ebert et al., 2007). miRNA engineering of CHO cells has been seen to alter; cell cycle (Sanchez et al., 2014), metabolism (Kelly et al., 2015), productivity (Sanchez et al., 2014; Jadhav et al., 2014; Loh et al., 2014; Fischer et al., 2015; Emmerling et al., 2015; Klanert et al., 2016; Scheollhorn et al., 2017), and inhibit apoptosis (Druz et al., 2013; Griffith et al., 2018). There are currently only a few examples of stable miRNA engineering resulting in enhanced CHO cell growth (Druz et al., 2013; Fischer et al., 2014), and only two miRNA profiling studies have focused on cell growth rate (Clarke et al., 2012; Klanert et al., 2016).

The present study aimed to validate whether prioritised miRNAs from a previous profiling study would impact cell specific growth rate (Clarke et al., 2012) and subsequently explore the biological function of miRNAs where depletion improves cell growth. Numerous challenges still exist in prediction of miRNA targets, chief among which is the lack of computational resources supporting the Chinese Hamster and CHO-K1 genomes (Fischer et al., 2015). Those that do exist fail to consider non-canonical binding events (Helwak et al., 2013). miRNA function to post-transcriptionally regulate the translation of protein coding mRNA. Unlike plant miRNA, which function to promote endonucleolytic cleavage of target mRNA through near perfect ~22nt base pairing (Filipowicz et al., 2008; Jones-Rhoades et al., 2006), miRNA mediated translational inhibition in mammalian cells is rarely through Argonaut 2 (AGO2)

cleavage (Liu et al., 2004; Meister et al., 2004) but rather imperfect miRNA – mRNA binding resulting in translational repression or deadenylation (Wu et al., 2006) via AGO1 – 4 complexing (Pillai et al., 2004). The unpredictable nature of mammalian miRNA – mRNA interactions therefore limits the use of transcriptomic approaches. To overcome this, subcellular fractionation of miRNA depleted cultures in combination with quantitative label-free liquid chromatography mass spectrometry (LC-MS/MS) proteomic analysis was used to identify significantly differentially expressed proteins. Subcellular fractionation prior to high-resolution mass spectrometry allowed us to improve the limit of detection and thus identify changes between low abundant proteins. To our knowledge this is the first report of subcellular proteomic analysis to investigate miRNA function in CHO cells.

5.2 Materials and methods

5.2.1 Cell culture and transfection

CHO DP12 (ATCC) cells were cultured in chemically defined Balan CD (Irvine Scientific) supplemented with 2 % (v/v) polyvinyl alcohol (PVA) and 4 mM L-glutamine (25030081 Gibco). Cells were seeded at 2 x 10^5 cells / mL and sub-cultured in a routine 3-4-3-day manner in 50 mL spin tubes (87050T Helena-BioSciences). Cultures were maintained in an ISF1-XC climo-shaker (Kühner), at 37°C, 80 % humidity and 5 % CO_2. For parental cells, the viable cell density (VCD/ mL) was monitored using the ViaCount^{TM} assay on a Guava® easyCyte benchtop cytometer (Merck Millipore). Selective pressure was applied periodically to maintain recombinant protein expression in DP12 with 200 nM Methotrexate (MTX) (Sigma-Aldrich). The formulae used to calculate growth rate, IVCD, AIVCD and Qp have been reporter previously (Sanchez et al., 2014). Transient transfection of siRNA was carried out using MIRUS TransIT X2 (MIRUS) by complexing 30 nM of siRNA with 1 µL of reagent for 30 minutes in 100 µL of CHO S SFM II (Gibco). siRNAs in this study included a negative control (siNC - Cat # 51-01-14-03 – Integrated DNA Technologies), a positive control targeting Vasolin-containing protein (siVCP) (Doolan et al., 2010) and a Usp14 specific siRNA (Cat. No. 4390824 – Thermo Fisher Scientific). Transient over-expression of human USP14 was carried out with USP14 - NM_005151. (GenScript).

5.2.2 Vector construction and cloning

Construction of the sponge vectors and the stable mixed pools generated for this study has been described previously (Costello et al., 2017). The miRNA sequences sponges used in this study can be found in Supplementary Table 1. miRNA sponge specificity was evaluated in-silico using the following online software http://genie.weizmann.ac.il/pubs/mir07/mir07_prediction.html (Kertesz et al., 2007) algorithm, with results for the NC and miR-378-3p sponges found in Supplementary Table 2. The 3' UTRs of CHO-K1 Actn4 and Usp14 were aligned with human, mouse and rat variants using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo). The target region of each UTR was synthesised by Integrated DNA Technologies, hybridized and cloned downstream of a d2eGFP reporter gene. This was previously described in detail by Costello et al., 2017.

5.2.3 Quantitative reverse-transcription polymerase chain reaction (qRT-PCR)

Cells were harvested by centrifugation at 1000 rpm for 5 minutes. Total RNA was isolated from 1-5x10^6 cells using Tri-reagent (Ambion) following the manufacturer's protocol. RNA quantification and quality were evaluated by NanoDrop (Thermo Fisher Scientific). RNA samples were treated with DNase I (Sigma Aldrich) prior to cDNA synthesis and stored at -80°C. Reverse transcription of total RNA was performed with the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems), on a (G-Storm) thermocycler, in accordance with the manufacturer's protocol. qRT-PCR was run using a 7500 (Applied Biosystems) with Fast SYBR Green Master Mix (Applied Biosystems). The 2X SYBR master mix was combined with 20 ng of cDNA, 200 nM Forward and Reverse primers and water, and made up to a 20 µL final reaction volume. Each biological replicate sample was run in technical triplicate wells. Primer sequences can be found in Supplementary Table 3. Relative quantification was measured by the ddCt method with Gapdh as an endogenous control. For miRNA analysis, the TaqMan miRNA Assay® system (Applied Biosystems) was used. Reverse transcription of specific mature miRNA from total RNA was done using the Taqman® miRNA Reverse Transcription Kit (Applied Biosystems). Relative miRNA abundance was determined by qRT-PCR of three biological replicates run in technical triplicates using the ddCt method with U6 snRNA as an endogenous control.

5.2.4 IgG quantification

CHO DP12 secretes an IgG (anti-IL8) which is quantified by ELISA. Supernatant is harvested simultaneously with RNA and protein sampling, by centrifugation of the cell suspension at 91 x g for 5 minutes. Nunc-Immuno™ MicroWell™ 96 well solid plates (M9410 Sigma-Aldrich) were coated with capture antibody (80-104), (BETHYL Laboratories), diluted 1:100 in coating buffer (C3041-50CAP) (Sigma-Aldrich) overnight at 4°C. Plate washing consisted of 3 x 100 µL /well of ELISA wash buffer (T9039) (Sigma-Aldrich). Blocking was carried out for 1 hr at room temperature with Blocking Buffer (T6789) (Sigma-Aldrich). Standards of Human serum standard (RS10-110) were diluted to range from 0-1000 ng/mL. Samples were diluted to fall within the range of the standard curve and incubated on the plate at room temperature for 1 hr. Detection antibody, HRP antibody (A80-104P) (BETHYL Laboratories), was diluted 1: 100,000 and 100 µL was added to each well and incubated for 1hr at room temperature. To each well, 100 µL of TMB substrate was added and incubated in the dark for 15 minutes. The reaction was stopped using 0.18 M H_2SO_4 and the plate was analysed on a MultiSKAN GO plate reader (Thermo Fisher Scientific) at 450nm.

5.2.5 Protein extraction and in-solution protein digestion

Triplicate biological replicates for day 4 and day 8 of culture following stable mir-378 depletion and NC-spg controls were collected for proteomic analysis. Cells were harvested by centrifugation at 100 x *g* for 5 minutes, at 4°C. Cell pellets were subjected to fractionation using the Mem-PER Plus membrane protein extraction kit (89842 - Thermo Fisher Scientific). The extraction was carried out according to the manufacturer's instructions to yield membrane and cytosolic enriched fractions. Protein concentrations were determined using the Quick Start Bradford protein assay kit (BioRad). Protein samples were purified and digested for mass spectrometry analysis as previously described by Coleman et al., 2017. In short, 100 µg of protein from each sample was purified using filter-aided sample preparation and digested using sequencing grade modified trypsin (Promega) at a ratio of 1:50 (protease: protein) overnight at 37°C. The peptides were purified using Pierce C18 spin columns (Thermo Fisher Scientific), then dried using vacuum centrifugation and re-suspended in an appropriate volume of loading buffer (2% acetonitrile).

5.2.6 LC–MS/MS and Quantitative label-free data analysis

LC–MS/MS analysis of the membrane and cytosolic fractions of mir-378-spg and NC-spg for day 4 and 8 was carried out as previously described (Henry and Meleady, 2017). In short, peptides were separated using an Ultimate 3000 nanoHPLC over 180 min which was coupled in-line to a hybrid linear LTQ Orbitrap (Thermo Fisher Scientific). Quantitative label-free data analysis was performed using Progenesis QI for Proteomics as previously described (Henry et al., 2018). The MS files were searched against the Cricetulus griseus CRIGR database from NCBI containing 44,065 sequences (fasta file downloaded November 2015). Differentially expressed (DE) proteins between the NC-spg and mir-378-spg have a statistical ANOVA p-value <0.05 and a minimum fold change of 1.25 in relative abundance between the NC-spg and mir-378-spg. Data visualization was achieved as before (Henry et al., 2017) using the updated ggplot2 package in R.

5.3 Results

5.3.1 Functional validation of miRNAs associated with cellular growth rate

A previous profiling study conducted by our group identified genes, proteins and miRNAs whose expression was robustly associated with cell specific growth rate (Clarke et al., 2012). Sister clones from the same original transfected pool were evaluated for growth phenotypes over 40 passages. Clones with similar cell specific productivity [24 (± 3) pg/ cell/ day] were characterized by growth rate as "fast" (≥ 0.025 cells/ mL/ hr) or "slow" (≤ 0.023 cells/ mL/ hr). The study found 93 miRNAs of the 667 analysed to be differentially expressed between the two subgroups, 17 of which were anti-correlated with growth. miRNAs whose expression was consistently high across CHO-K1 derived cell lines used by Hackl et al. (Hackl et al. 2011) were prioritized for miRNA sponge mediated knockdown to functionally validate their role in regulating cell specific growth rate. This resulted in a panel of six miRNAs to be tested for phenotypic impact; miR-204-5p, miR-338-3p, miR-378-3p, miR-409-3p, miR-455-3p and miR-505-3p. For each candidate miRNA a specific miRNA-sponge-decoy (Fig. 5.1A) was designed as carried out in previous studies (Ebert et al., 2007; Kluvier et al., 2012). Each sponge sequence contained 10 miRNA responsive elements (MRE) in the 3'

untranslated region (UTR) of a reporter gene. A destabilized GFP (d2eGFP) was used as a reporter due to its short half-life. To assess the effect of candidate miRNA depletion on cell growth the sponge vectors were used to generate stable mixed pools in a model IgG producer line, CHO DP12. The relative quantification (RQ) of mature target miRNAs in stable mixed pools were assessed with respect to a non-specific negative control sponge (NC-spg) (Sanchez et al., 2014) (Fig. 5.1B). Statistically significant miRNA knockdown was observed for five of the six stable mixed pools; miR-204-5p ($p = 0.003107$), miR-338-3p ($p = 0.00227$), miR-378-3p ($p = 0.00202$), miR-455-3p ($p = 0.00284$) and miR-505-3p (0.00075). There was no significant knockdown of miR-409-3p ($p = 0.08679$). The assessment of cell growth following stable depletion of specific candidate miRNAs revealed three of the six prioritized candidates to have a significant effect on cell growth in this cell line (Fig. 5.1C). Depletion of miR-378-3p, miR-455-3p and miR-505-3p caused significant increases in peak viable cell density on day 8 of culture of 60%, 25% and 24% respectively. From this initial screen we identified miR-378-3p depletion as having a biologically significant impact on CHO cell growth and prioritized this candidate for subsequent investigation.

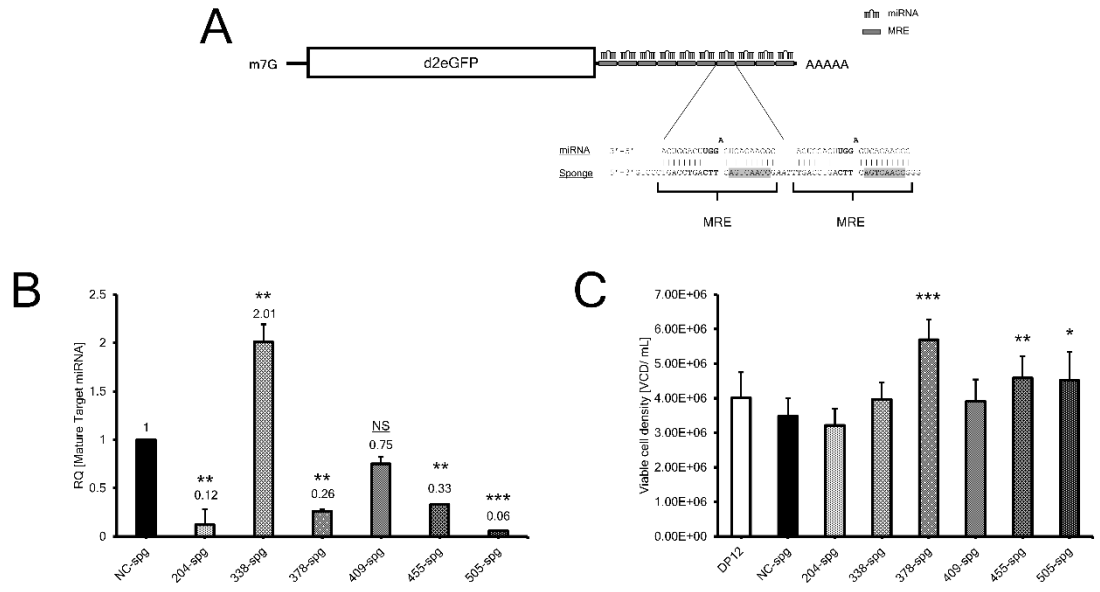


Figure 5.1. Functional assessment of growth associated miRNAs in stable mixed pools. (A) A schematic representation of the miRNA sponge design used in this study. miRNA sponge sequences were cloned downstream of a destabilized GFP (d2eGFP) reporter gene. miRNA seed regions are highlighted and a 4-nucleotide mismatch or "bulge" sequence in (bold). (B) Stable mixed pools of cells expressing a specific miRNA sponge decoy were generated in CHO DP12 cells for miRNAs; 204-5p, 338-3p, 378-3p, 409-3p, 455-3p, 505-3p and a non-specific negative control (NC-spg). miRNA sponge efficacy was evaluated for all candidates by analysing the relative quantification of specific mature miRNA in sponge cell lines compared to a non-specific negative control sponge. (C) Viable cell density [VCD/ mL] of CHO DP12 and stable specific miRNA sponge expressing mixed pools for miRNAs with respect to the NC-spg mixed pool.

The batch culture performance of the 378-spg cells was evaluated in a direct comparison to the NC-spg cells. Phenotypic assessment was conducted at a 5 mL scale in 50 mL spin tube reactors. Stable depletion of miR-378 in CHO DP12 cells achieved a 59% ($p=4.33x10^{-5}$) increase in viable cell density [VCD /mL] over the NC-spg on day 8 of culture with no compromise of culture viability (Fig. 5.2A). Stable mixed pools of NC-spg and miR-378-spg cells were scaled to 60 mL cultures and the strong growth phenotype was stable with this scale up (Fig. 5.2B). The cellular growth rate of the 60 mL cultures during exponential growth was improved from 0.0177 cells/ mL/ hr in the NC-spg cells to 0.0215 cells/ mL/ hr in the miR-378-3p depleted cells (p=0.02256) (Fig. 5.2C). The integrated viable cell density (IVCD) (Fig. 5.2D) and accumulated IVCD (AIVCD) (Fig. 5.2E) were both increased significantly in the 378-spg cells on day eight by 61% ($p = 1.92 \times 10^{-5}$) and 44% ($p = 1.16 \times 10^{-5}$) respectively. There was no difference in volumetric titre between the NC-spg and 378-spg cells in the 60 mL scale

(Fig. 5.2F). The cell specific productivity (Qp) was significantly reduced in the 378-spg cells on days 4, 6 and 8 of culture by; 32% (p = 0.025), 23% (p = 0.0139) and 35% (p = 0.00289) respectively (Fig. 5.2G).

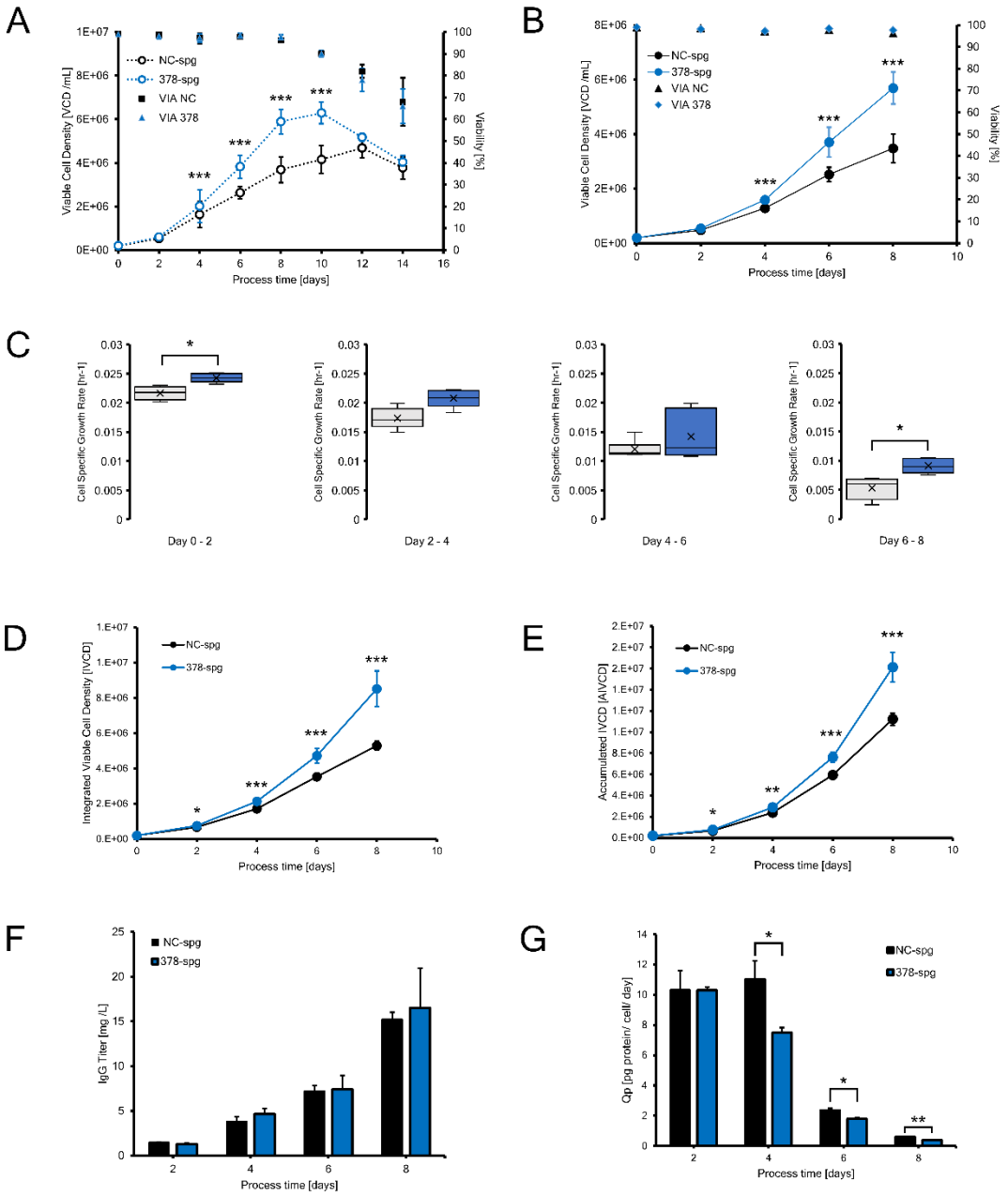


Figure 5.2. Batch analysis of stable miR-378-3p depletion in CHO DP12 cells. (A) The Negative control sponge (NC-spg) was grown in parallel with the miR-378-3p-sponge cell line (378-spg). Viable cell density (VCD) /mL is seen on the y-axis (left), viability (right), and culture duration (Days) on the x-axis. (B) [VCD/ mL] and culture viability of mixed pools scaled to 60mL scale. (C) Cell specific growth rate of NC-spg and 378-spg cells at 60mL scale. (D) Integrated VCD (IVCD) of NC-spg and 378-spg cells at 60mL. (E) Accumulated IVCD of NC-spg and 378-spg cells at 60mL scale. (F) IgG production (mg /L) for NC-spg and 378-spg cells grown at 60mL scale. (G) Cell specific productivity (Qp) (pg/ cell/ day) for NC-spg and 378-spg cells grown at 60mL scale.

5.3.2 Differential proteomic analysis of CHO DP12 cells following miR-378 depletion

Proteomic analysis for this study was carried out using a subcellular enrichment strategy to yield two fractions, membrane and cytosolic (Fig. 5.3A), to improve proteome coverage. Sample complexity negatively affects the ability to detect, identify and quantify low-abundance proteins by MS because low-abundance peptides are beyond the limit of detection by MS in whole proteome analyses. By fractionating the complex cell proteome, the dynamic range of each subcellular fraction is increased in comparison to a total proteome analysis thus proteins at low abundance have a better chance of being detected. By fractionating our mir-378-spg and NC-spg stable mixed pools we improve signal-to-noise and proteome coverage of our samples within the LTQ-OrbiTrap XL mass spectrometer.

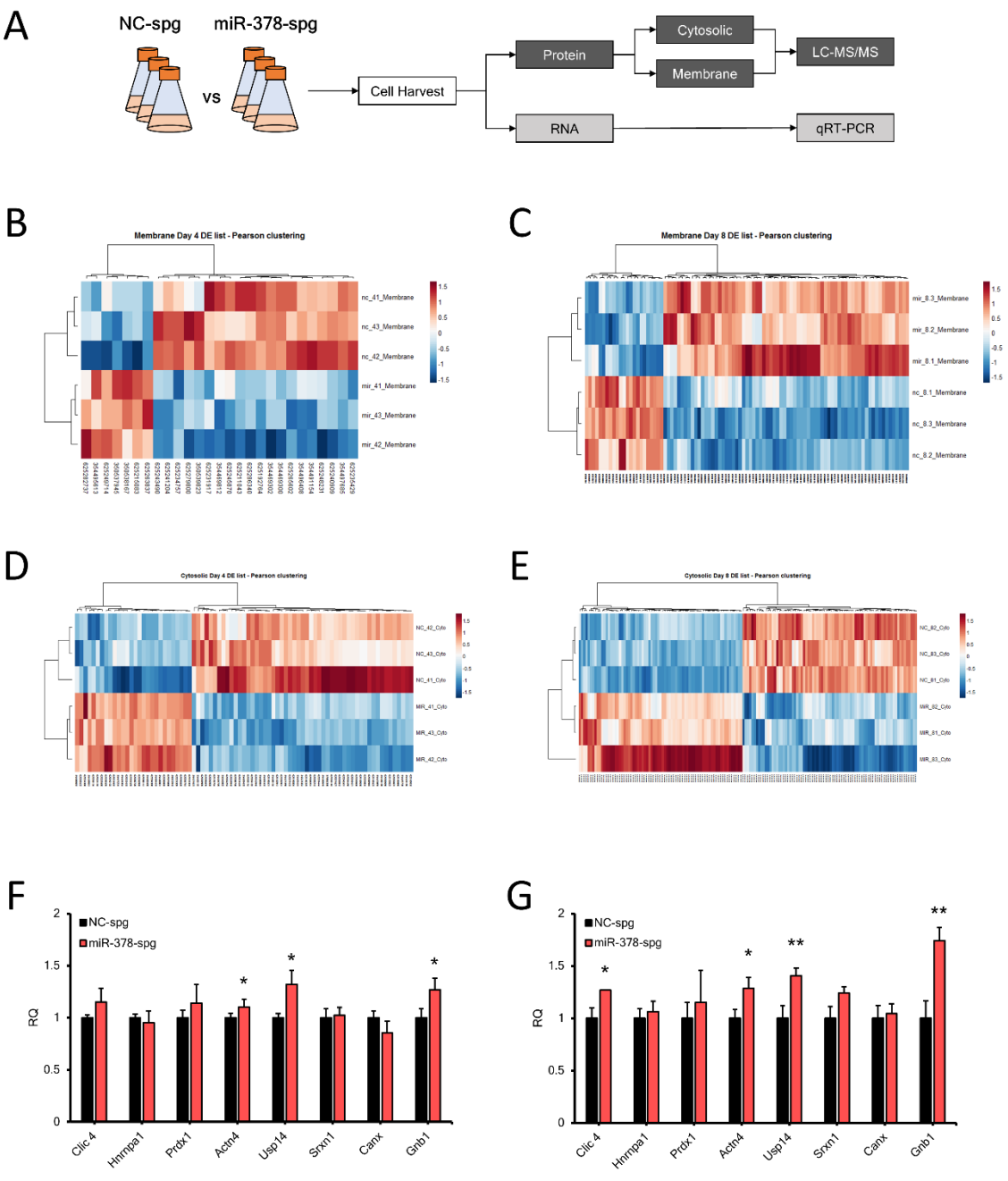


*Figure 5.3. Differential expression analysis of NC-spg and 378-spg cells. (A) Schematic representation of experimental set up. Three biological replicates of NC-spg and 378-spg cells were grown at 60mL with cells harvested on days 4 and 8 of culture. Cells were used for proteomic and transcript analysis. Relative protein abundance visualized by heat map for differentially expressed membrane proteins on day 4 (B) and day 8 (C), and cytosolic proteins day 4 (D) and day 8 (E). Transcript analysis was undertaken for matched differentially expressed proteins on both days of analysis. qRT-PCR was carried out in technical triplicates for each gene on day 4 (F) and day 8 (G). Relative quantification was done using Gapdh as an endogenous control. Statistical analysis consisted of a two-tailed homoscedastic student t-test, ($p \leq 0.05$ *, $p \leq 0.01$**, $p \leq 0.001$ ***), values based on triplicate biological RNA samples.*

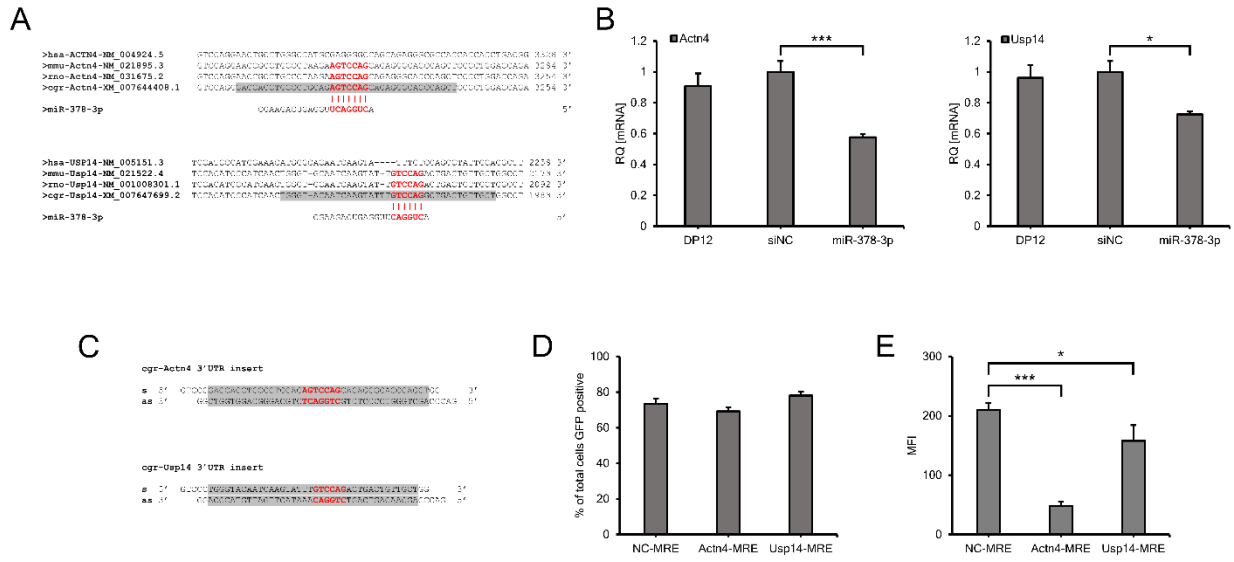


*Figure 5.4 Functional validation of Actn4 and Usp14 mRNA interaction with miR-378-3p. (A) A multiple alignment Actn4 and Usp14 mRNAs from human, mouse, rat and CHO-K1 revealed conservation of miR-378-3p binding sites in rodents but not human. The predicted miRNA responsive element (MRE) is highlighted in bold. (B) miR-378-3p was transiently over-expressed and the relative quantification (RQ) of Actn4 and Usp14 mRNA was compared to a negative control siRNA (siNC). (C) The predicted MRE for Actn4 and Usp14 with 30bp up and downstream flanking sequence was cloned downstream of a GFP reporter with one copy of the MRE in each case. DP12 cells were transfected with one of an; negative control (NC-MRE), Actn4-MRE or Usp14-MRE plasmids. The % of total cells GFP positive (D) and corresponding median fluorescence intensity (MFI) (E) for each construct. Statistics were carried out by a two-tailed homoscedastic student t-test, (p≤0.05 *, p≤0.01**, p≤0.001 ***), values based on three biological replicates.*

Proteomic analysis identified an average of 700 proteins per cytosolic fraction and 590 proteins per membrane fraction. Label-free quantitative data analysis was carried out using Progenesis QI for Proteomics to identify proteins whose abundance significantly changed in comparison to a non-specific negative control as a result of sponge depletion of miRNA-378. Differentially expressed (DE) proteins with significantly increased or decreased expression between samples were determined by using a fold-change cut-off of ±1.25-fold and an ANOVA *p*-value of <0.05 between the two stable mixed pools for each fraction and the time point analysed. Sequestering miR-378-3p using sponge decoy engineering allows the endogenous targets of miR-378-3p to have repealed regulation in the cells thus exhibiting an increased protein abundance level in the miR-378-spg sample sets. To achieve a better understanding of the effect of miR-378-3p on the

proteome, samples were analysed at day 4 and day 8 of culture to capture the mid-exponential and stationary phases of growth.

Unsupervised Pearson hierarchical clustering of the differentially expressed proteins shows clear separation of the cell lines, NC-spg and miR-378-spg, into distinct sample groups based on protein abundance differences resulting from miR-378 depletion as visualised by heat maps for both fractions and time points (Fig. 5.3). Quantitative analysis of day 4 of culture identified 27 (Fig. 5.3B) and 81 (Fig. 5.3D) differentially expressed proteins for the membrane and cytosolic fractions, respectively, with a minimum of 1.25-fold change between the NC-spg line and the miR-378-spg line. Of those significant proteins 7 from the membrane fraction and 28 from the cytosolic showed higher abundance in the mir-378-spg sample set which suggests they are potential targets of miR-378. Late stage culture at day 8 identified 95 (Fig. 5.3C) and 151 (Fig. 5.3E) significantly differentially expressed proteins for the membrane and cytosolic fractions respectively. Of those proteins, 72 from the membrane fraction and 73 from the cytosolic fractions exhibited a higher abundance level in the mir-378-spg stable cell line.

A cohort of differentially expressed proteins were found to have consistently higher abundance in the miR-378-spg cell line at both stages of culture analysed; 6 proteins in the cytosolic fraction and 2 proteins in the membrane fraction. All eight of these proteins were investigated at the transcript level by qRT-PCR for day 4 (Fig. 5.3F) and day 8 (Fig. 5.3G). Relative quantification (RQ) of mRNA for 3 of the 8 genes showed a statistically significant directional correlation between protein and mRNA levels for; Actn4, Usp14 and Gnb1 on both days 4 and 8. The RQ of the remaining genes was unchanged on day 4 and 8, apart from Clic4, whose expression increased on day 8. The molecular function and biological process associated with the 8 genes is shown in Table 5.1. In silico miRNA target prediction software was used to identify potential miR-378 – mRNA interactions of proteins from the differentially expressed list. Hnrnpa1 is a predicted target of miR-378-3p by both TargetScan 7.1 and miRDB algorithms and broadly conserved in vertebrates while Actn4, Usp14 and Srxn1 have conserved target sites in rodents (Table 5.2). The remaining genes had no predicted miR-378-3p target

sites in their 3'UTR, potentially indicative of secondary effects as the result of stable miR-378-3p depletion.

Table 5.1 Proteins with increased peptide abundance on day 4 and day 8

Gene ID	Fraction	FC Day 4	FC Day 8	Anova Day 4	Anova Day 8	Molecular Function	Biological Process
Clic4	Cytosolic	1.37	1.55	2.23E-04	2.26E-02	Anion Channel Activity	Anion Transport
Hnrnpa1	Cytosolic	1.32	1.85	5.72E-04	4.53E-02	RNA Binding	Cellular Response to Glucose Starvation
Prdx1	Cytosolic	1.33	1.85	2.38E-03	2.26E-02	Oxioreductase Activity	Cell Proliferation
Actn4	Cytosolic	1.26	1.28	3.94E-03	3.71E-03	Actin Binding	Protein Transport
Usp14	Cytosolic	1.45	1.58	5.27E-03	5.51E-04	Hydrolase	Ubl Conjugation Pathway
Srxn1	Cytosolic	1.35	1.37	9.24E-03	3.04E-02	Antioxidant	Cellular Response to Oxidative Stress
Canx	Membrane	1.52	1.60	2.35E-02	1.78E-03	Calcium Ion Binding	Exocytosis
Gnb1	Membrane	1.30	1.28	1.83E-02	1.26E-03	GTPase Activity	G-protein Coupled Receptor Signalling Pathway

Table 5.2. TargetScan 7.1 Analysis of miR-378-3p and UTR binding of DE genes

Gene ID	Position in 3'UTR /miRNA	Predicted consequential pairing of target region (top) and miRNA (bottom)	Site type	Context ++ score	Context ++ score percentile	Weighted context ++ score	Conserved branch length
Hnrnpa1	Position 1766-1772 of HNRNPA1 3' UTR mmu-miR-378a-3p	5' ...UAUUGAGCCAAAACUAGUCCAGU... ||||||| 3' GGAAGACUGAGGUUCAGGUCA	7mer-m8	-0.22	88	-0.21	3.822
Hnrnpa1	Position 850-857 of HNRNPA1 3' UTR mmu-miR-378a-3p	5' ...UUAAUGCCACCUAUAAGUCCAGA... ||||||| 3' GGAAGACUGAGGUUCAGGUCA	8mer	-0.44	98	-0.42	0
Actn4	Position 368-374 of ACTN4 3' UTR mmu-miR-378a-3p	5' ...CCGCCUGCCCUAAGAAGUCCAGC... ||||||| 3' GGAAGACUGAGGUUCAGGUCA	7mer-m8	-0.13	74	-0.13	0.403
Usp14	Position 444-450 of USP14 3' UTR mmu-miR-378a-3p	5' ...GUGCAAUCAAGUAUUGUCCAGAC... |||||| 3' GGAAGACUGAGGUUCAGGUCA	7mer-A1	-0.16	79	-0.12	0.403
Srxn1	Position 1428-1434 of SRXN1 3' UTR mmu-miR-378a-3p	5' ...CUUCUGCAAACCUAGAGUCCAGG... |||||| |||||| 3' GGAAGACUGAGGU---UCAGGUCA	7mer-m8	-0.14	75	-0.07	0

5.3.3 Determining direct targets of miR-378-3p from LC-MS/ MS data

Actn4 and Usp14 were up-regulated at both mRNA and protein levels in the miR-378-spg cell line. These two genes also contain a single predicted miR-378-3p target site in their 3' UTR (Fig. 5.4A). To investigate direct interaction between miR-378-3p and the two genes transient over-expression of miR-378-3p was performed using a miRNA mimic in the original DP12 cell line (Fig. 5.4B). Exogenous expression of miR-378-3p significantly reduced the abundance of Actn4 ($p = 0.00083$) and Usp14 ($p = 0.04309$) mRNA. The miR-378-3p target region of each 3' UTR was cloned with flanking sequence downstream of a d2eGFP reporter gene (Fig. 5.4C). The effects of these UTR motifs on reporter gene translation were monitored with respect to a negative control (NC). Each construct showed no significant difference in the total cells GFP positive post transfection (Fig. 5.4D). There was a significant reduction in the median fluorescent intensity of the constructs containing UTR elements from Actn4 ($p = 3.056 \times 10^{-5}$) and Usp14 ($p = 0.0353$) with respect to the NC (Fig. 5.4E).

5.3.4 Functional analysis of Usp14 validates LC-MS/ MS data

Of the 4 differentially expressed proteins with predicted miR-378 binding sites in their UTR, only two have been reported to effect cell proliferation; Hnrnpa1 (He et al., 2005) and Usp14 (Wang et al., 2014). However, He et al., (2005) demonstrated that RNAi mediated knockdown of single heterogenous nuclear ribonucleoproteins had no effect on cell proliferation. Only by targeting a combination of Hnrnpa1 and Hnrnpa2 or Hnrnpa3 cellular proliferation rate was reduced. As neither of these proteins was up-regulated in the 378-spg cell line we postulated that this may not be the source of enhanced growth. The Ubiquitin carboxyl-terminal hydrolase 14 (Usp14) had the strongest p-value of the differentially expressed proteins which is clearly visible when analysed by volcano plot (Fig. 5.5A). Usp14 has also previously been associated with cellular proliferation in ovarian cancer (Wang et al., 2014).

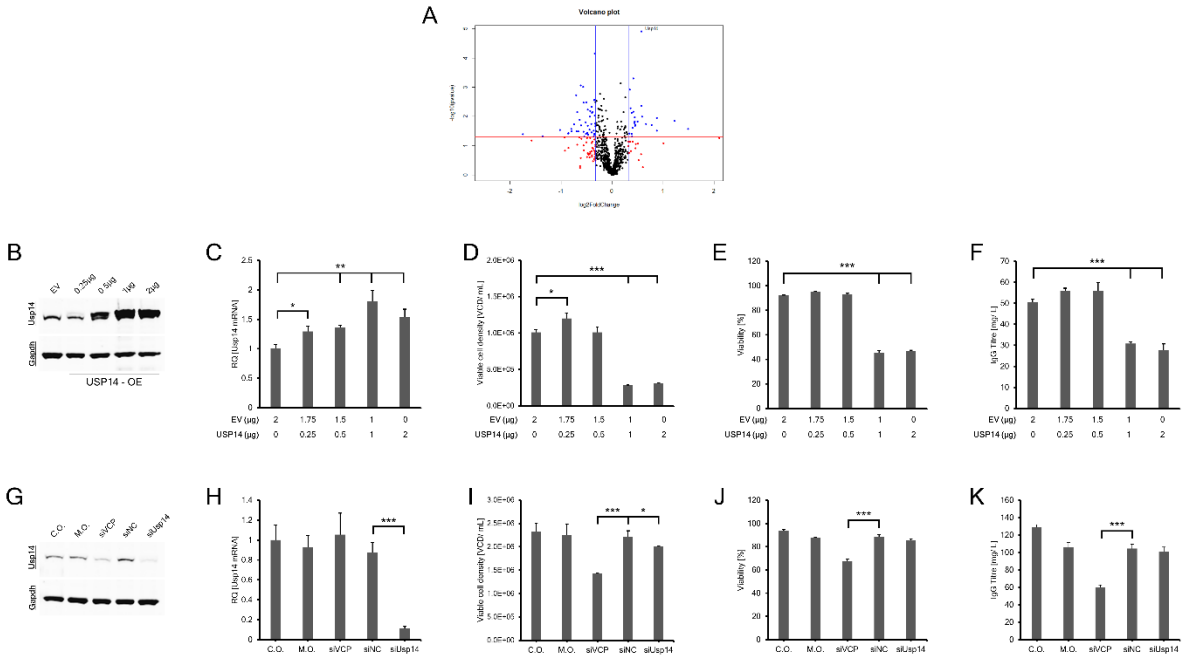


Figure 5.5 Functional validation of LC-MS/ MS and in silico predicted miR-378-3p target Usp14 in CHO DP12 cells. (A) The distribution of differentially expressed proteins based on p*-value and fold-change between mir-378-spg and NC-spg from the cytosolic fraction of day 8 of culture is shown by volcano plot. USP14 was transiently over-expressed (OE) in CHO DP12 cells at different levels. To achieve this USP14 expression vector was diluted with an empty vector (EV). (B) Protein expression 48hrs post transfection was analysed by Western blot. (C) USP14 over expression was evaluated at transcript level by qRT-PCR. Cells transfected with USP14 over-expression vector were monitored for viable cell density (D), viability (E) and volumetric titre (F). Transient knockdown of endogenous Usp14 was done with siRNA. Knockdown was monitored with respect to cells only (C.O.) and vehicle only (M.O.) controls. An siRNA targeting VCP was used as a positive control. (G) The expression of Usp14 was assessed at protein level 72hrs after transfection by Western blot. (H) Endogenous transcript levels of Usp14 was measured by qRT-PCR. The effects of Usp14 knockdown on viable cell density (I), viability (J) and volumetric productivity (K) were monitored 72hrs post transfection.*

To functionally validate the differential expression of Usp14 by depletion of miR-378 via miRNA-sponge decoy, as a source of enhanced cell growth, exogenous expression and repression of Usp14 in CHO DP12 cells was evaluated (Fig. 5.5). Usp14 is highly conserved between human and Chinese hamster, with 96% homology (Sup.Fig. 1). The human USP14 open reading frame (ORF) was transiently over-expressed in CHO DP12 cells. Transient expression of the human USP14 in DP12 cells was assessed with respect to an empty vector (EV) control. To evaluate whether there was a correlation between the level of USP14 up-regulation and phenotype, the USP14 plasmid was used at different concentrations, 0.25, 0.5, 1 and 2 µg, made up to a total of 2 µg with empty vector in each case. After 48 h transfected cells were lysed and over expression (OE) of USP14 protein levels assessed by Western blot (Fig. 5.5B). The level of up-regulation

increased with over-expression plasmid concentration as anticipated. Over-expression of USP14 was also monitored at the transcript level by qRT-PCR (Fig. 5.5C). Cells were monitored for VCD (Fig. 5.5D) and viability (Fig. 5.5E) 48h post transfection. There was a significant ($p=0.016$) increase in cell growth of 20% at the lowest level of Usp14-OE evaluated (Fig. 5.5D). At high levels of Usp14-OE (1µg) cells showed a significant reduction in VCD of 70% (Fig. 5.5D) and viability (55%) (Fig. 5.5E). Increased growth with 0.25 µg of USP14-OE plasmid did not significantly improve product titre (Fig. 5.5F).

To elucidate the effect on cell growth of reduced Usp14 activity siRNA mediated knockdown of endogenous CHO DP12 Usp14 was conducted. CHO DP12 cells were transiently transfected with 30 nM siRNA. Knockdown at the protein level was evaluated by Western blot 72 h after transfection (Fig. 5.5G). There was a significant (89%) knockdown of Usp14 mRNA 24 h post transfection (Fig. 5.5H), with no significant alteration in Usp14 mRNA levels between all controls. Knockdown of Usp14 resulted in a significant ($p= 0.0315$) 15% reduction in viable cell density compared to the cells only control, and a 10% reduction versus the negative control ($p= 0.0455$) (Fig. 5.5I). Knockdown of Usp14 had no significant effect on cell viability (Fig. 5.5J) or volumetric titer (Fig. 5.5K).

Given the role of Usp14 in the un-folded protein (UPR) response we investigated the expression of UPR marker genes with stable miR-378-3p depletion. Six genes, three involved in PERK signalling; Atf4, Atf6 and Chop, and three involved in protein folding; Grp78, Grp94 and Xbp1 (spliced variant) were evaluated at transcript level on days four and eight of culture (Fig. 5.6). There was a significant reduction in Chop ($p = 0.0237$), Grp78 ($p = 0.0454$) and Xbp1 ($p = 0.01118$) mRNA on day four of culture. These genes were not found to be DE in the LC-MS/ MS data suggesting this lower mRNA abundance did not significantly alter them at the protein level.

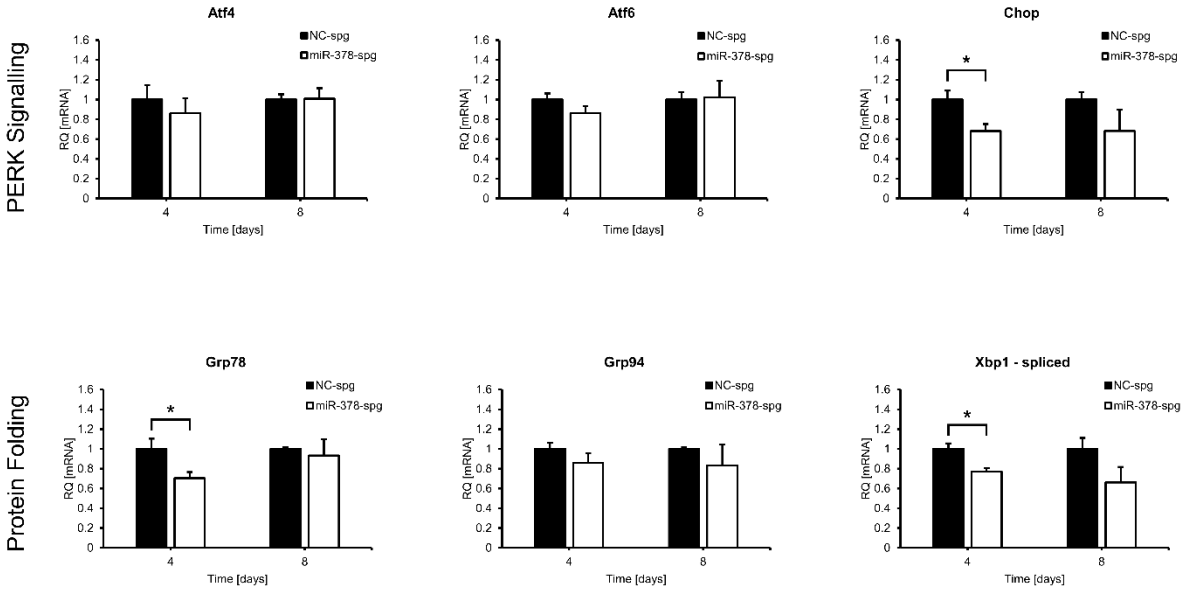


*Figure 5.6 Evaluation of UPR marker genes with miR-378-3p depletion. Marker genes of the un-folded protein response were evaluated by qRT-PCR. The relative quantification (RQ) of gene mRNA was compared on days four and eight of culture with miR-378-3p depletion to the NC-spg line. Three genes; Aft4, Atf6 and Chop are involved in the PERK signalling pathway and the three involved in protein folding; Grp78, Grp94 and Xbp1. Statistics were carried out by a two-tailed homoscedastic student t-test, ($p \leq 0.05$ *, $p \leq 0.01$**, $p \leq 0.001$ ***), values based on three biological replicates.*

5.4 Discussion

The notion of multi-miRNA engineering of CHO cells is not new, yet the catalogue of miRNAs with known robust phenotypes required to implement such strategies does not yet exist (Costello et al., 2018). As it currently stands there are reports of several promising miRNA candidates to stably enhance CHO cell specific productivity (Sanchez et al., 2014; Jadhav et al., 2014; Loh et al., 2014; Fischer et al., 2015; Emmerling et al., 2015; Klanert et al., 2016; Scheollhorn et al., 2017) and inhibit apoptosis (Druz et al., 2013; Griffith and Kelly et al., 2018). Currently there are very few studies showing significantly enhanced CHO cell growth through miRNA manipulation (Druz et al., 2013; Fischer et al., 2014). The present study evaluated the effect of stable depletion of miRNAs previously associated with CHO cell growth rate (Clarke et al., 2012). Of the six miRNAs evaluated, 50% significantly improved peak viable cell density of CHO DP12 cells in batch culture (Fig. 5.1C). Perhaps unsurprisingly the three miRNA that evoked a growth phenotype had higher basal

expression than those that did not show any effect (Sup. Fig. 2). Target miRNA knockdown was observed for five of the candidates evaluated (Fig.5.1B). The theory of target mediated miRNA protection (TMMP) (Chatterjee, et al. 2011), a proposed phenomenon wherein the miRNA is protected from exoribonuclease activity by target association, could explain the apparent increase in miR-338 abundance with stable sponge expression (Fig. 5.1B). TMMP has been reported with a previous miRNA knockdown in CHO cells (Sanchez et al., 2014). From this screen we identified the 59% improvement in peak viable cell density via miR-378-3p depletion as a biologically significant impact. The improved growth phenotype remained statistically significant in the scale up from 5 to 60 mL (Fig. 5.2B). The increase in VCD did not lead to any improvement in volumetric titer (Fig. 5.2D). This reduction in cell specific productivity could be the cause of the improved growth. miR-378-3p has been described as an onco-miR and is associated with tumorigenesis and proliferation in nasopharyngeal carcinoma (Yu et al., 2014), prostate (Nguyen et al., 2013; Avgeris et al., 2014; Chen et al., 2015), colorectal (Wang et al., 2014; Zhang et al., 2014) and gastric (Fei et al., 2013) cancers.

Eight proteins; Clic4, Hnrnpa1, Prdx1, Actn4, Usp14, Srxn1, Canx and Gnb1, had the highest abundance in the miR-378-spg cells for both time points of LC-MS/MS analysis. These 8 proteins were assessed at the transcript level by qRT-PCR, 3 of which; Actn4, Usp14 and Gnb1, showed a significant directional correlation between mRNA and protein levels at both points of culture analysed. This cohort of proteins are potentially direct targets of miR-378 as our data proves their protein expression levels significantly increase upon depletion of miR-378-3p. However due to the complexity of eukaryotic miRNA – mRNA interactions, translational repression associated with miRNA targeting may not involve mRNA turnover. Four of the eight proteins; Hnrnpa1, Actn4, Usp14 and Srxn1, contained a predicted miR-378-3p binding site in the 3'UTR. Transient over expression of miR-378-3p in DP12 cells resulted in significant knockdown of Actn4 and Usp14 (Fig. 5.4B) and presence of their predicted binding region in the 3' UTR of a GFP reporter caused significant reduction in median fluorescence intensity (Fig. 5.4D & E). Direct targets or secondary effects, these proteins are very likely to contribute to the growth phenotype seen with miR-378-3p depletion. Clic4 has previous associations with cell proliferation (Tung et al., 2009), as does Prdx1 (Cai et al., 2018), USP14 (Wang et

al., 2015) and Hnrnpa1 (He et al., 2005). Two proteins from the list are also associated with cellular un-folded protein response (UPR); Usp14 and Canx. Calr was identified as significantly up-regulated on day 8 of culture only. Usp14 (Lee et al, 2010), Canx and Calr (Pieren et al., 2005) are all involved in cellular UPR and may contribute to the reduced specific productivity of the miR-378-3p depleted cells.

Ubiquitin carboxyl-terminal hydrolase 14 (Usp14) had the strongest statistically differentially expressed protein abundance between the NC-spg and miR-378-spg cells (ANOVA, $p = 5.51x10^{-4}$). Elevated levels of cellular Usp14 has previously been linked to proliferation and tumorigenesis in breast (Zhu et al., 2015), leukemic (Ishiwata et al., 2001), colorectal (Shinji et al., 2006), lung (Wu et al., 2013) and ovarian (Wang et al., 2014) cancers. Usp14 plays a major role in the regulation of the proteasome (Lee et al, 2010) and is one of three proteasome-associated deubiquitinating (DUB) enzymes (Hanna et al, 2006; Bashore et al, 2015; Lee et al, 2016). The proteasome is an essential mechanism in eukaryotes, responsible for non-lysosomal intracellular protein degradation, regulating many aspects of cell physiology and responses including; cell division, proliferation and apoptosis (Finley, 2009; Schrader et al., 2009; Wang et al., 2014). Proteins are tagged for proteasome mediated degradation by the covalent attachment of ubiquitin or ubiquitin-like proteins (Ub or UBLs). Three classes of proteins (E1, E2 and E3) act in a sequential manner to ubiquitinate proteins destined for proteasome mediated turn-over. The process starts with the activation of the highly conserved 78-amino acid protein, Ub, by the E1 (activator) class of enzymes in an ATP dependant reaction. Followed by transfer of the activated Ub to the E2 (conjugase) class and finally covalent tagging of target proteins by E3 (ligase) enzymes. Targeted proteins may be mono or poly-ubiquitinated, (Jin et al, 2009). The tagged protein is finally degraded by the proteasome. Cellular ubiquitination is balanced by deubiquitination (Zhu et al, 2016), this is achieved by DUBs. Mammalian proteasomes are associated with three predominant DUBs; RPN11, UCH37 and USP14 (Verma et al, 2002; Yao et al, 2002; Koulich et al, 2008; Anderson et al, 2005). Unlike RPN11, which is a stoichiometric subunit of the proteasome, UCH37 and USP14 can work independently to disassemble or trim but not degrade Ub chains, thus impeding the tagged protein's

commitment to proteasome degradation (Lam et al, 1997; Hanna et al, 2006; Hedge et al, 2007).

Over expression of human USP14 in the original CHO DP12 cell line showed a dynamic response to different levels of up-regulation. With a minor increase in USP14 protein abundance VCD was improved by 20%, yet at high levels of protein expression there was a 70% reduction in VCD and 55% reduction in viability. Knockdown of endogenous Usp14 by siRNA significantly reduced VCD. Exogenous expression and repression of Usp14 had anticipated positive and negative effects on CHO DP12 cell growth respectively, providing confidence in the LC-MS/ MS differential expression analysis. However, the effect of this single gene manipulation was less impactful on cell growth than knockdown of miR-378-3p.

5.5 Conclusions

This data represents the first subcellular proteomic approach to understanding the mechanism of action of microRNAs in CHO cells. Forced expression and repression of the strongest statistically differentially expressed protein from the LC-MS/ MS data gave anticipated inverse phenotypes. Stable depletion of miR-378-3p enhanced CHO DP12 cell density by 59%, a much greater impact that transient over-expression of the top candidate from the LC-MS/ MS data. This confirms the known benefit of miRNA engineering strategies, in that a single miRNA manipulation can affect multiple genes simultaneously. Finally, while the cell specific productivity was reduced with miR-378-3p depletion, its beneficial impact on cell growth could make miR-378-3p a candidate for multi-miRNA engineering strategies.

Acknowledgments

This work was conducted under the financial support of Scientific Foundation of Ireland (SFI) grant numbers [13/IA/1963] and [13/IA/1841]. The authors would also like to thank Dr. Colin Clarke of the National Institute for Bioprocessing Research and training miRNA candidates used in this study.

5.6 References

Anderson, C., Crimmins, S., Wilson, J. A., Korbel, G. A., Ploegh, H. L., & Wilson, S. M. (2005) Loss of Usp14 results in reduced levels of ubiquitin in ataxia mice. *Journal of neurochemistry. 95*, 724-731.

Avgeris, M., Stravodimos, K., & Scorilas, A. (2014) Loss of miR-378 in prostate cancer, a common regulator of KLK2 and KLK4, correlates with aggressive disease phenotype and predicts the short-term relapse of the patients. *Biological chemistry. 395*, 1095-1104.

Bartel, D. P. (2009) MicroRNAs: Target recognition and regulatory functions. *Cell. 136*, 215-233.

Bashore, C., Dambacher, C. M., Goodall, E. A., Matyskiela, M. E., Lander, G. C., & Martin, A. (2015) Ubp6 deubiquitinase controls conformational dynamics and substrate degradation of the 26S proteasome. *Nature structural & molecular biology. 22*, 712-719.

Brinkrolf, K., Rupp, O., Laux, H., Kollin, F., Ernst, W., Linke, B., Kofler, R., Romand, S., Hesse, F., & Budach, W. E. (2013) Chinese hamster genome sequenced from sorted chromosomes. *Nature biotechnology. 31*, 694-695.

Cai, A., Zeng, W., Cai, W., Liu, J., Zheng, X., Liu, Y., Yang, X., Long, Y., & Li, J. (2018) Peroxiredoxin-1 promotes cell proliferation and metastasis through enhancing Akt/mTOR in human osteosarcoma cells. *Oncotarget. 9*, 8290.

Chen, Q., Zhou, W., Han, T., Du, S., Li, Z., Zhang, Z., Shan, G., & Kong, C. (2016) MiR-378 suppresses prostate cancer cell growth through downregulation of MAPK1 in vitro and in vivo. *Tumor Biology. 37*, 2095-2103.

Clarke, C., Henry, M., Doolan, P., Kelly, S., Aherne, S., Sanchez, N., Kelly, P., Kinsella, P., Breen, L., Madden, S. F., Zhang, L., Leonard, M., Clynes, M., Meleady, P., & Barron, N. (2012) Integrated miRNA, mRNA and protein expression analysis reveals the role of post-transcriptional regulation in controlling CHO cell growth rate. *BMC genomics. 13*, 656-2164-13-656.

Coleman, O., Henry, M., Clynes, M., & Meleady, P. (2017) Filter-aided sample preparation (FASP) for improved proteome analysis of recombinant chinese hamster ovary cells. *Heterologous Protein Production in CHO Cells: Methods and Protocols*, 187-194.

Costello, A., Lao, N., Clynes, M., & Barron, N. (2017) Conditional knockdown of endogenous MicroRNAs in CHO cells using TET-ON-SanDI sponge vectors. *Heterologous Protein Production in CHO Cells: Methods and Protocols*, 87-100.

Costello, A., Lao, N. T., Gallagher, C., Capella Roca, B., Julius, L. A., Suda, S., Ducrée, J., King, D., Wagner, R., & Barron, N. (2018) Leaky expression of the TET-On system hinders control of endogenous miRNA abundance. *Biotechnology journal*, 1800219.

Doolan, P., Meleady, P., Barron, N., Henry, M., Gallagher, R., Gammell, P., Melville, M., Sinacore, M., McCarthy, K., & Leonard, M. (2010) Microarray and proteomics expression profiling identifies several candidates, including the valosin-containing protein (VCP), involved in regulating high cellular growth rate in production CHO cell lines. Biotechnology and bioengineering. 106, 42-56.

Druz, A., Son, Y., Betenbaugh, M., & Shiloach, J. (2013) Stable inhibition of mmu-miR-466h-5p improves apoptosis resistance and protein production in CHO cells. *Metabolic engineering. 16*, 87-94.

Ebert, M. S., Neilson, J. R., & Sharp, P. A. (2007) MicroRNA sponges: Competitive inhibitors of small RNAs in mammalian cells. *Nature methods. 4*, 721-726.

Emmerling, V. V., Fischer, S., Stiefel, F., Holzmann, K., Handrick, R., Hesse, F., Hörer, M., Kochanek, S., & Otte, K. (2015) Temperature-sensitive miR-483 is a conserved regulator of recombinant protein and viral vector production in mammalian cells. *Biotechnology and bioengineering*

Fei, B., & Wu, H. (2013) MiR-378 inhibits progression of human gastric cancer MGC-803 cells by targeting MAPK1 in vitro. *Oncology Research Featuring Preclinical and Clinical Cancer Therapeutics. 20*, 557-564.

Filipowicz, W., Bhattacharyya, S. N., & Sonenberg, N. (2008) Mechanisms of post-transcriptional regulation by microRNAs: Are the answers in sight?. *Nature reviews.Genetics. 9*, 102.

Finley, D. (2009) Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annual Review of Biochemistry. 78*, 477-513.

Fischer, S., Buck, T., Wagner, A., Ehrhart, C., Giancaterino, J., Mang, S., Schad, M., Mathias, S., Aschrafi, A., Handrick, R., & Otte, K. (2014) A functional high-content miRNA screen identifies miR-30 family to boost recombinant protein production in CHO cells. *Biotechnology journal. 9*, 1279-1292.

Fischer, S., Handrick, R., & Otte, K. (2015) The art of CHO cell engineering: A comprehensive retrospect and future perspectives. *Biotechnology Advances. 33*, 1878-1896.

Griffith, A., Kelly, P. S., Vencken, S., Lao, N. T., Greene, C. M., Clynes, M., & Barron, N. (2017) miR-CATCH identifies biologically active miRNA regulators of the Pro-Survival gene XIAP, in chinese hamster ovary cells. Biotechnology journal

Hackl, M., Jakobi, T., Blom, J., Doppmeier, D., Brinkrolf, K., Szczepanowski, R., Bernhart, S. H., zu Siederdissen, C. H., Bort, J. A. H., & Wieser, M. (2011) Next-generation sequencing of the chinese hamster ovary microRNA transcriptome: Identification, annotation and profiling of microRNAs as targets for cellular engineering. Journal of Biotechnology. 153, 62-75.

Hanna, J., Hathaway, N. A., Tone, Y., Crosas, B., Elsasser, S., Kirkpatrick, D. S., Leggett, D. S., Gygi, S. P., King, R. W., & Finley, D. (2006) Deubiquitinating enzyme Ubp6 functions noncatalytically to delay proteasomal degradation. *Cell. 127*, 99-111.

He, Y., Brown, M. A., Rothnagel, J. A., Saunders, N. A., & Smith, R. (2005) Roles of heterogeneous nuclear ribonucleoproteins A and B in cell proliferation. Journal of cell science. 118, 3173-3183.

Hegde, A. N., & Upadhya, S. C. (2007) The ubiquitin–proteasome pathway in health and disease of the nervous system. *Trends in neurosciences. 30*, 587-595.

Helwak, A., Kudla, G., Dudnakova, T., & Tollervey, D. (2013) Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. Cell. 153, 654-665.

Henry, M., Gallagher, C., Kelly, R. M., Frye, C. C., Osborne, M. D., Brady, C. P., Barron, N., Clynes, M., & Meleady, P. (2018) Clonal variation in productivity and proteolytic clipping of an fc-fusion protein in CHO cells: Proteomic analysis suggests a role for defective protein folding and the UPR. *Journal of Biotechnology*

Henry, M., & Meleady, P. (2017) Clinical proteomics: Liquid Chromatography–Mass spectrometry (LC–MS) purification systems. In: Anonymous , Protein Chromatography. Springer, pp. 375-388

Henry, M., Power, M., Kaushik, P., Coleman, O., Clynes, M., & Meleady, P. (2017) Differential phosphoproteomic analysis of recombinant chinese hamster ovary cells following temperature shift. *Journal of proteome research. 16,* 2339-2358.

Ishiwata, S., Katayama, J., Shindo, H., Ozawa, Y., Itoh, K., & Mizugaki, M. (2001) Increased expression of queuosine synthesizing enzyme, tRNA-guanine transglycosylase, and queuosine levels in tRNA of leukemic cells. *The Journal of Biochemistry. 129,* 13-17.

Jadhav, V., Hackl, M., Klanert, G., Hernandez Bort, J. A., Kunert, R., Grillari, J., & Borth, N. (2014) Stable overexpression of miR-17 enhances recombinant protein production of CHO cells. *Journal of Biotechnology. 175*, 38-44.

Jones-Rhoades, M. W., Bartel, D. P., & Bartel, B. (2006) MicroRNAs and their regulatory roles in plants. *Annu.Rev.Plant Biol. 57,* 19-53.

Kelly, P. S., Breen, L., Gallagher, C., Kelly, S., Henry, M., Lao, N. T., Meleady, P., O'Gorman, D., Clynes, M., & Barron, N. (2015) Re-programming CHO cell metabolism using miR-23 tips the balance towards a highly productive phenotype. *Biotechnology journal. 10*, 1029-1040.

Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U., & Segal, E. (2007) The role of site accessibility in microRNA target recognition. *Nature genetics. 39*, 1278.

Klanert, G., Jadhav, V., Chanoumidou, K., Grillari, J., Borth, N., & Hackl, M. (2014) Endogenous microRNA clusters outperform chimeric sequence clusters in chinese hamster ovary cells. Biotechnology journal. 9, 538-544.

Klanert, G., Jadhav, V., Shanmukam, V., Diendorfer, A., Karbiener, M., Scheideler, M., Bort, J. H., Grillari, J., Hackl, M., & Borth, N. (2016) A signature of 12 microRNAs is robustly associated with growth rate in a variety of CHO cell lines. Journal of Biotechnology

Kluiver, J., Gibcus, J. H., Hettinga, C., Adema, A., Richter, M. K., Halsema, N., Slezak-Prochazka, I., Ding, Y., Kroesen, B. J., & van den Berg, A. (2012) Rapid generation of microRNA sponges for microRNA inhibition. *PloS one. 7*, e29275.

Koulich, E., Li, X., & DeMartino, G. N. (2008) Relative structural and functional roles of multiple deubiquitylating proteins associated with mammalian 26S proteasome. *Molecular biology of the cell. 19*, 1072-1082.

Lam, Y. A., Xu, W., DeMartino, G. N., & Cohen, R. E. (1997) Editing of ubiquitin conjugates by an isopeptidase in the 26S proteasome. *Nature. 385*, 737.

Lee, B., Lee, M. J., Park, S., Oh, D., Elsasser, S., Chen, P., Gartner, C., Dimova, N., Hanna, J., & Gygi, S. P. (2010) Enhancement of proteasome activity by a small-molecule inhibitor of USP14. *Nature. 467*, 179-184.

Lee, B. H., Lu, Y., Prado, M. A., Shi, Y., Tian, G., Sun, S., Elsasser, S., Gygi, S. P., King, R. W., & Finley, D. (2016) USP14 deubiquitinates proteasome-bound substrates that are ubiquitinated at multiple sites. *Nature. 532*, 398-401.

Lewis, N. E., Liu, X., Li, Y., Nagarajan, H., Yerganian, G., O'Brien, E., Bordbar, A., Roth, A. M., Rosenbloom, J., & Bian, C. (2013) Genomic landscapes of chinese hamster ovary cell lines as revealed by the cricetulus griseus draft genome. *Nature biotechnology. 31*, 759-765.

Liu, J., Carmell, M. A., Rivas, F. V., Marsden, C. G., Thomson, J. M., Song, J. J., Hammond, S. M., Joshua-Tor, L., & Hannon, G. J. (2004) Argonaute2 is the catalytic engine of mammalian RNAi. *Science (New York, N.Y.). 305*, 1437-1441.

Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., & Tuschl, T. (2004) Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Molecular cell. 15*, 185-197.

Nguyen, H. C. N., Xie, W., Yang, M., Hsieh, C., Drouin, S., Lee, G. M., & Kantoff, P. W. (2013) Expression differences of circulating microRNAs in metastatic castration resistant prostate cancer and low-risk, localized prostate cancer. *The Prostate. 73*, 346-354.

Pieren, M., Galli, C., Denzel, A., & Molinari, M. (2005) The use of calnexin and calreticulin by cellular and viral glycoproteins. *The Journal of biological chemistry. 280*, 28265-28271.

Pillai, R. S., Artus, C. G., & Filipowicz, W. (2004) Tethering of human ago proteins to mRNA mimics the miRNA-mediated repression of protein synthesis. *RNA (New York, N.Y.). 10*, 1518-1525.

Sanchez, N., Kelly, P., Gallagher, C., Lao, N. T., Clarke, C., Clynes, M., & Barron, N. (2014) CHO cell culture longevity and recombinant protein yield are enhanced by depletion of miR-7 activity via sponge decoy vectors. *Biotechnology journal. 9*, 396-404.

Schoellhorn, M., Fischer, S., Wagner, A., Handrick, R., & Otte, K. (2017) miR-143 targets MAPK7 in CHO cells and induces a hyperproductive phenotype to enhance production of difficult-to-express proteins. *Biotechnology progress*

Schrader, E. K., Harstad, K. G., & Matouschek, A. (2009) Targeting proteins for degradation. *Nature chemical biology. 5*, 815-822.

Shinji, S., Naito, Z., Ishiwata, S., Ishiwata, T., Tanaka, N., Furukawa, K., Suzuki, H., Seya, T., Matsuda, A., & Katsuta, M. (2006) Ubiquitin-specific protease 14 expression in colorectal cancer is associated with liver and lymph node metastases. *Oncology reports. 15*, 539-543.

Tung, J. J., Hobert, O., Berryman, M., & Kitajewski, J. (2009) Chloride intracellular channel 4 is involved in endothelial proliferation and morphogenesis in vitro. *Angiogenesis. 12*, 209-220.

Verma, R., Aravind, L., Oania, R., McDonald, W. H., Yates, J. R.,3rd, Koonin, E. V., & Deshaies, R. J. (2002) Role of Rpn11 metalloprotease in deubiquitination and degradation by the 26S proteasome. *Science (New York, N.Y.). 298*, 611-615.

Wang, K., Ma, J., Zhang, F., Yu, M., Xue, J., & Zhao, J. (2014) MicroRNA-378 inhibits cell growth and enhances l-OHP-induced apoptosis in human colorectal cancer. *IUBMB life. 66*, 645-654.

Wang, Y., Wang, J., Zhong, J., Deng, Y., Xi, Q., He, S., Yang, S., Jiang, L., Huang, M., & Tang, C. (2015) Ubiquitin-specific protease 14 (USP14) regulates cellular proliferation and apoptosis in epithelial ovarian cancer. *Medical oncology. 32*, 379.

Wang, Y., Wang, J., Zhong, J., Deng, Y., Xi, Q., He, S., Yang, S., Jiang, L., Huang, M., & Tang, C. (2015) Ubiquitin-specific protease 14 (USP14) regulates cellular proliferation and apoptosis in epithelial ovarian cancer. *Medical oncology. 32*, 379.

Wu, N., Liu, C., Bai, C., Han, Y., Cho, W., & Li, Q. (2013) Over-expression of deubiquitinating enzyme USP14 in lung adenocarcinoma promotes proliferation through the accumulation of β-catenin. *International journal of molecular sciences. 14*, 10749-10760.

Wu, N., Liu, C., Bai, C., Han, Y., Cho, W., & Li, Q. (2013) Over-expression of deubiquitinating enzyme USP14 in lung adenocarcinoma promotes proliferation through the accumulation of β-catenin. *International journal of molecular sciences. 14*, 10749-10760.

Wu, L., Fan, J., & Belasco, J. G. (2006) MicroRNAs direct rapid deadenylation of mRNA. *Proceedings of the National Academy of Sciences of the United States of America. 103*, 4034-4039.

Xu, X., Nagarajan, H., Lewis, N. E., Pan, S., Cai, Z., Liu, X., Chen, W., Xie, M., Wang, W., & Hammond, S. (2011) The genomic sequence of the chinese hamster ovary (CHO)-K1 cell line. *Nature biotechnology. 29*, 735-741.

Yao, T., & Cohen, R. E. (2002) A cryptic protease couples deubiquitination and degradation by the proteasome. *Nature. 419*, 403.

Yu, B., Peng, X., Zhao, F., Liu, X., Lu, J., Wang, L., Li, G., Chen, H., & Li, X. (2014) MicroRNA-378 functions as an onco-miR in nasopharyngeal carcinoma by repressing TOB2 expression. *International journal of oncology. 44*, 1215-1222.

Zhang, G., Zhou, H., Xiao, H., Li, Y., & Zhou, T. (2014) MiR-378 is an independent prognostic factor and inhibits cell growth and invasion in colorectal cancer. *BMC cancer. 14*, 109.

Zhu, L., Yang, S., He, S., Qiang, F., Cai, J., Liu, R., Gu, C., Guo, Z., Wang, C., & Zhang, W. (2016) Downregulation of ubiquitin-specific protease 14 (USP14) inhibits breast cancer cell proliferation and metastasis, but promotes apoptosis. *Journal of molecular histology. 47*, 69-80.

5.7 Appendix B

Supplementary Table 1: miRNA sponge sequences

Sequence Name	miRNA Target	miRBase ID	miRNA Sponge Sequence (5'-3')
NC-spg s	-	-	GTCCCAAGTTTTCAGAAAGCTAACACCGGAAGTTTTCAGAAAGCTAACAGG
NC-spg as	-	-	GACCCTGTTAGCTTTCTGAAAACTTCCGGTGTTAGCTTTCTGAAAACTTGG
miR-204-5p-spg s	miR-204-5p	MI0020454	GTCCCAGGCATAGGACTGAAAGGGAAAATTAGGCATAGGACTGAAAGGGAAGG
miR-204-5p-spg as	miR-204-5p		GACCCTTCCCTTTCAGTCCTATGCCTAATTTTCCCTTTCAGTCCTATGCCTGG
miR-338-spg s	miR-338	MI0020499	GTCCCTCAACAAAATCGAGATGCTGGAAATTTCAACAAAATCGAGATGCTGGAGG
miR-338-spg as	miR-338		GACCCTCCAGCATCTCGATTTTGTTGAAATTTCCAGCATCTCGATTTTGTTGAGG
miR-378-spg s	miR-378-3p	MI0020515	GTCCCGCCTTCTGACGTTAGTCCGATAATTGCCTTCTGACGTTAGTCCGATGG
miR-378-spg as	miR-378-3p		GACCCATCGGACTAACGTCAGAAGGCAATTATCGGACTAACGTCAGAAGGCGG
miR-409-3p-spg s	miR-409-3p	MI0020520	GTCCCAGGGGTTCACTTTCAACATTCAATTAGGGGTTCACTTTCAACATTCGG
miR-409-3p-spg as	miR-409-3p		GACCCGAATGTTGAAAGTGAACCCCTAATTGAATGTTGAAAGTGAACCCCTGG
miR-455-3p-spg s	miR-455-3p	MI0020528	GTCCCAGTGTATATGAAATGGACTGCAATTAGTGTATATGAAATGGACTGCGG
miR-455-3p-spg as	miR-455-3p		GACCCGCAGTCCATTTCATATACACTAATTGCAGTCCATTTCATATACACTGG
miR-505-3p-spg s	miR-505-3p	MI0020538	GTCCCAGAGGAAACCATTTGTGTTGACAATTAGAGGAAACCATTTGTGTTGACGG
miR-505-3p-spg as	miR-505-3p		GACCCGTCAACACAAATGGTTTCCTCTAATTGTCAACACAAATGGTTTCCTCTGG

Supplementary Table 2: Evaluating Sponge – miRNA Interaction and Specificity

Prediction Rank	Sponge	miRNA Target	MRE	Score	Prediction Rank	Sponge	miRNA Target	MRE	Score
1	NC-spg	hsa-miR-769-3p	2	-8.3	**1**	miR-378-spg	hsa-miR-378	10	-30.48
2	NC-spg	hsa-miR-920	2	-7.05	**2**	miR-378-spg	hsa-miR-422a	10	-26.51
3	NC-spg	hsa-miR-1306	10	-6.18	**3**	miR-378-spg	hsa-miR-432	7	-16.91
4	NC-spg	hsa-miR-138	5	-6.14	**4**	miR-378-spg	hsa-miR-383	7	-16.34
5	NC-spg	hsa-miR-637	4	-5.73	**5**	miR-378-spg	hsa-miR-1268	8	-15.81
6	NC-spg	hsa-miR-146b-5p	9	-5.34	**6**	miR-378-spg	hsa-miR-663	3	-13.08
7	NC-spg	hsa-miR-626	9	-5.17	**7**	miR-378-spg	hsa-miR-214	7	-12.72
8	NC-spg	hsa-miR-7	9	-5.14	**8**	miR-378-spg	hsa-miR-665	7	-11.58
9	NC-spg	hsa-miR-571	5	-4.87	**9**	miR-378-spg	hsa-miR-890	8	-11.21
10	NC-spg	hsa-miR-199b-5p	5	-4.46	**10**	miR-378-spg	hsa-miR-1287	3	-10.06
11	NC-spg	hsa-miR-146a	9	-4.38	**11**	miR-378-spg	hsa-miR-650	10	-9.95
12	NC-spg	hsa-miR-492	2	-4.3	**12**	miR-378-spg	hsa-miR-524-3p	4	-9.23
13	NC-spg	hsa-miR-630	9	-4.28	**13**	miR-378-spg	hsa-miR-944	4	-8.83
14	NC-spg	hsa-miR-1263	2	-4.27	**14**	miR-378-spg	hsa-miR-548j	4	-7.73
15	NC-spg	hsa-miR-1299	9	-4.18	**15**	miR-378-spg	hsa-miR-510	7	-7.57

hsa-miR-378-3p: 5′ A**CUGGACUU**GGAGUCAGAAGGC 3′
hsa-miR-422-5p: 5′ A**CUGGACUU**AGGGUCAGAAGGC 3′

Supplementary Table 3: Primer Information (qRT-PCR)

DE - 378	Gene Symbol	Primer Name	Sequence (5'-3')	Tm (°C)	Amplicon Size (bp)
UP	**Clic4**	Clic4 FP	GAGGGGGCTCTTGAAGACAC	60	145
UP	**Clic4**	Clic4 RP	GCAGTCTGCCAGTGTCATCT	60	145
UP	**Hnrnpa1**	Hnrnpa1 FP	GGTCGAGGGGGAAACTTCAG	60	195
UP	**Hnrnpa1**	Hnrnpa1 RP	ATAGCCACTGCCCTGGTTTC	60	195
UP	**Prdx1**	Prdx1 FP	ACTTTTGTGTGCCCCACTGA	60	157
UP	**Prdx1**	Prdx1 RP	TCATGGGTCCCAATCCTCCT	60	157
UP	**Actn4**	Actn4 FP	CAGATCCTCACCCGAGATGC	60	151
UP	**Actn4**	Actn4 RP	TCTCCACATCGTAGCCCAGA	60	151
UP	**Usp14**	Usp14 FP	GAGCTTCAGGGGAAATGGCT	60	193
UP	**Usp14**	Usp14 RP	ACCCAGCACTCGTTAGCATC	60.1	193
UP	**Srxn1**	Srxn1 FP	GTCCTCTGGATCAAAGGGGC	60.1	109
UP	**Srxn1**	Srxn1 RP	GCTTGGCAGGAATGGTCTCT	60	109
UP	**Canx**	Canx FP	TCTGGCAGCGACCTTTGATT	60	182
UP	**Canx**	Canx RP	GGACCAGAGCTCCAAACCAA	59.9	182
UP	**Gnb1**	Gnb1 FP	TGGGATGTCCGAGAAGGGAT	60	114
UP	**Gnb1**	Gnb1 RP	AGCATCGTCTGAACCAGTGG	60	114
DOWN	**S100a4**	S100a4 FP	CTGAAGGAGCTGCTGACCAG	60.4	152
DOWN	**S100a4**	S100a4 RP	ATGGCAATGCAGGACAGGAA	60	152
DOWN	**Calr**	Clar FP	CTCTGGCAGGTCAAGTCTGG	60	199
DOWN	**Calr**	Clar RP	CCTCAGCCTCTTCCTCCTCT	60	199
DOWN	**Hsp90b1**	Hsp90b1 FP	ACCGGGAAGCAACAGAGAAG	60	105
DOWN	**Hsp90b1**	Hsp90b1 RP	CCGTGAGACGCTGAGATACC	60	105
DOWN	**Rpl26**	Rpl26 FP	AGGACACTACAAAGGCCAGC	60	161
DOWN	**Rpl26**	Rpl26 RP	TGTCCAGCTTTAGCCTGGTG	60	161
-	**Gapdh**	Gapdh FP	TGGCTACAGCAACAGAGTGG	59.9	144
-	**Gapdh**	Gapdh RP	GTGAGGGAGATGATCGGTGT	60	144
-	**Xbp1**	Xbp1 FP	CTCCAGAGACGGAGTCCAAG	60	181
-	**Xbp1**	Xbp1F	ACAGGGTCCAACTTGTCCAG	60	181

-	**Grp78**	Grp78 FP	GTGCAGAAACTTCGTCGTGA	60	162
-	**Grp78**	Grp78 RP	GGTAGATCGGAACAGGTCCA	59.9	162
-	**Grp94**	Grp94 FP	ACCGGGAAGCAACAGAGAAG	60	105
-	**Grp94**	Grp94 RP	CCGTGAGACGCTGAGATACC	60	105
-	**Atf6**	ATF6 FP	GCCACTGAAGGAAGACAAGC	60	173
-	**Atf6**	ATF6 RP	TGATGGTTTTTGCTGGAACA	60.1	173
-	**Atf4**	Atf4 FP	TTGCAACCTTTTCCCTGTTC	60.1	167
-	**Atf4**	Atf4 RP	TCTGAGGGGGTGTCTTCATC	60	167
-	**Chop**	Chop FP	TTTTGCCTTGGAGACGGTGT	60	160
-	**Chop**	Chop RP	TAGGGATGCGGAGTCTAGGG	59.9	160
-	-	Usp14 3'UTR MRE s	GTCCCTGGGTACAATCAAGTATTTGTCCAGACTGACTGTTGCTGG	-	-
-	-	Usp14 3'UTR MRE as	GACCCAGCAACAGTCAGTCTGGACAAATACTTGATTGTACCCAGG	-	-
-	-	Actn4 3'UTR MRE s	GTCCCGACCACCTGCCCTGCAGAGTCCAGCAGAGGGCACCCAGCTGG	-	-
-	-	Actn4 3'UTR MRE as	GACCCAGCTGGGTCCCCTCTGCTGGACTCTGCAGGGCAGGTGGTCGG	-	-

6 Chapter Six: A proteomic profiling dataset of recombinant Chinese hamster ovary cells showing enhanced cellular growth following miR-378 depletion

Published in *Data-in-Brief* December 2018

DOI: 10.1016/j.dib.2018.11.115

Authors: Orla Coleman, Alan Costello, Michael Henry, Nga T. Lao, Niall Barron, Martin Clynes and Paula Meleady

I share joint primary authorship for this manuscript with Alan Costello. I was responsible for all data analyses included in this manuscript comprising sample preparation, quantitative data analysis, and bioinformatic analysis. I prepared, wrote and submitted this manuscript to the journal and performed all revisions.

Abstract

The proteomic data presented in this article provide supporting information to the related research article "Depletion of endogenous miRNA-378-3p increases peak cell density of CHO DP12 cells and is correlated with elevated levels of Ubiquitin Carboxyl-Terminal Hydrolase 14". Control and microRNA-378 depleted CHO DP12 cells were profiled using label-free quantitative proteomic profiling. CHO DP12 cells were collected on day 4 and 8 of batch culture, subcellular proteomic enrichment was performed, and subsequent fractions were analysed by liquid chromatography tandem mass spectrometry (LC-MS/MS). Here we provide the complete proteomic dataset of proteins significantly differentially expressed by greater than 1.25-fold change in abundance between control and miR-378 depleted CHO DP12 cells, and the lists of all identified proteins for each condition.

Specifications Table

Subject area	Biotechnology
More specific subject area	Proteomics
Type of data	Excel Tables and Figure
How data was acquired	LTQ-Orbitrap XL mass spectrometer (Thermo Scientific), Progenesis QI for Proteomics (Non-linear Dynamics, Waters) and Proteome Discoverer software (Thermo Scientific)
Data format	Relative quantitation calculated and qualitative analysis
Experimental factors	Subcellular enrichment for of CHO DP12 control cells and cells depleted of microRNA-378
Experimental features	Quantitative proteomic profiling of CHO DP12 cells following microRNA-378 depletion
Data source location	Dublin, Ireland
Data accessibility	Data available in this article
Related research article	Costello, A., Coleman, O., Lao, N.T., Henry, M., Meleady, M., Barron, N. & Clynes, M. Depletion of Endogenous miRNA-378-3p Increases Peak Cell Density of CHO DP12 Cells and is Correlated with Elevated Levels of Ubiquitin Carboxyl-Terminal Hydrolase 14. Journal of Biotechnology (Costello et al., 2018)

Value of the data

- This data reveals protein expression patterns associated with microRNA-378
- Differentially expressed proteins between control and miR-378 depleted CHO cells may serve as indicators of CHO cell growth
- This dataset reports enriched proteins from the cytosolic and membrane subcellular fractions of CHO DP12 cells
- This data provides proteomic profiles for two time-points of CHO DP12 batch culture; exponential and stationary phase

6.1 Data

The data presents a quantitative proteomic profiling of subcellular-enriched protein fractions from day 4 and day 8 cultures of CHO DP12 cells following microRNA-378 stable depletion. Both the cytosolic and membrane protein enriched fractions were analysed to identify significantly differentially expressed proteins between control and miR-378 depleted CHO cells (miR-378-spg) for each time point. Differentially expressed proteins between control and miR-378-spg cells are required to have (i) a *p*-value ≤ 0.05 on the peptide and the protein level and (ii) a minimum of 1.25-fold change in normalized abundance levels.

Tables 6.1 - 6.4 list the differentially expressed proteins with an increased abundance in miR-378 depleted cells when compared to control cells. Proteins with an increased abundance in miR-378-spg cells represent potential direct targets of miR-378 in CHO cells and are of most interest. Tables 6.1 - 6.4 report the accession number, peptide count, number of unique peptides, ANOVA *p*-value, q-value, maximum fold-change and protein name. Supplementary table S1 presents the complete list of all differentially overexpressed and under expressed proteins for each subcellular fraction and time-point. Supplementary table S2 presents the qualitative list of all identified proteins for each condition (control and miR-378-spg), subcellular enriched fraction (cytosolic and membrane protein enriched) and time-point (day 4 and day 8 of culture). Heat maps are shown in figure 6.1 that outlines the clustering of significantly increased versus decreased proteins in miR-378-spg cells, as compared to control cells.

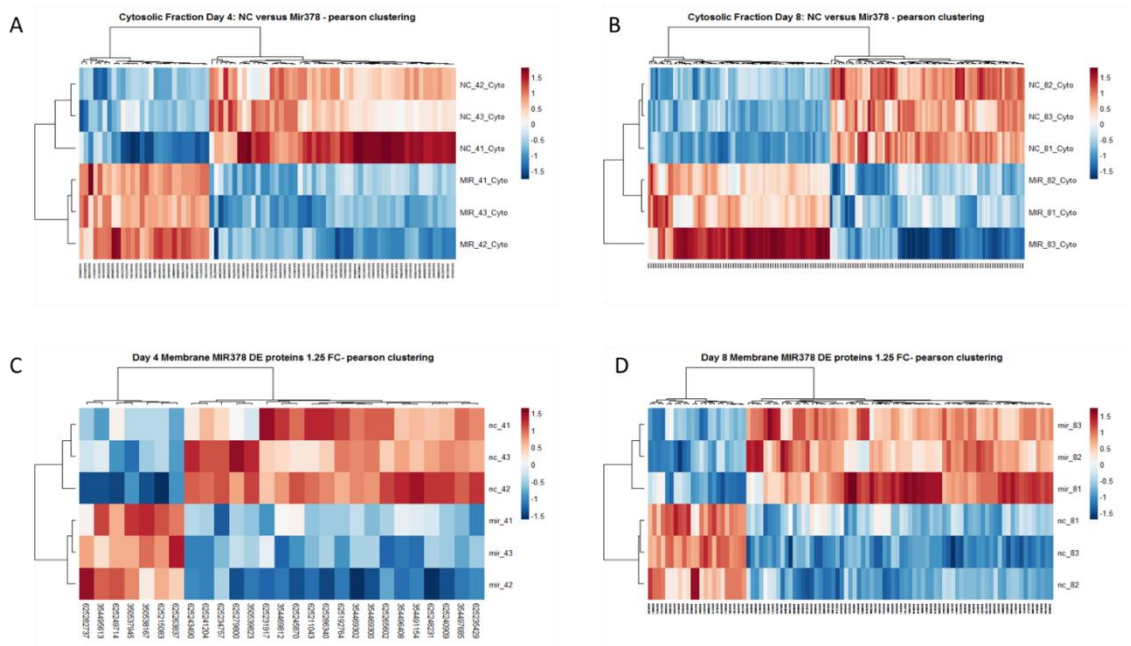


Figure 6.1. Heat maps of differentially expressed proteins in miR-378-spg CHO cells. 1A and 1B show the clustering of significantly increased and decreased proteins identified in the cytosolic enriched fraction of miR-378-spg cells for day 4 and day 8, respectively. 1C and 1D show the clustering of differentially expressed proteins identified in the membrane enriched fraction of miR-378-spg when compared to control on day 4 and day 8 of culture, respectively. The normalised abundance values of differentially expressed proteins were log2 transformed and hierarchical Pearson clustering was performed on Z-score normalised intensity values.

Table 6.1 Mass spectrometric identification of 28 proteins from the cytosolic enriched protein fraction with ≥1.25-fold increase in the miR-378 depleted CHO cells on day 4 of cell culture

Accession	Peptides	Unique peptides	Anova (p)	Q value	Max fold change	Protein name
625285532	1	1	1.08E-02	2.96E-02	1.51	splicing factor 3B subunit 3
625233305	2	2	5.27E-03	1.98E-02	1.45	ubiquitin carboxyl-terminal hydrolase 14
354504493	9	9	1.24E-03	1.13E-02	1.44	6-phosphogluconate dehydrogenase, decarboxylating
625250820	4	4	7.81E-03	2.69E-02	1.43	copine-1 isoform X3
354500682	1	1	2.00E-02	3.88E-02	1.41	cytochrome b5
625231502	1	1	3.25E-02	4.62E-02	1.40	leucine-rich repeat-containing protein 47
625204380	3	3	2.23E-04	7.06E-03	1.37	chloride intracellular channel protein 4
625279800	1	1	5.08E-03	1.95E-02	1.37	caveolin-1 isoform X1
354481364	1	1	3.55E-02	4.65E-02	1.36	crk-like protein isoform X1
625258134	1	1	9.24E-03	2.76E-02	1.35	sulfiredoxin-1 isoform X2
625290509	1	1	3.70E-02	4.65E-02	1.34	T-complex protein 1 subunit beta isoform X2
350537945	9	9	2.38E-03	1.35E-02	1.33	peroxiredoxin-1
625225560	2	2	5.72E-04	7.39E-03	1.32	heterogeneous nuclear ribonucleoprotein A1
625260720	1	1	1.75E-02	3.77E-02	1.31	TAR DNA-binding protein 43 isoform X1
354477234	2	2	2.29E-02	3.97E-02	1.30	F-actin-capping protein subunit alpha-2
354502560	2	2	1.95E-03	1.35E-02	1.30	protein DJ-1
354480001	1	1	1.29E-02	3.06E-02	1.30	T-complex protein 1 subunit delta
354495613	1	1	9.68E-03	2.82E-02	1.30	thrombomodulin
625250988	1	1	2.32E-02	3.97E-02	1.29	inositol-3-phosphate synthase 1 isoform X2

625224185	1	1	1.14E-02	2.99E-02	1.29	spermidine synthase
625280088	10	10	4.60E-03	1.85E-02	1.28	alpha-enolase isoform X3
625234360	1	1	4.51E-02	4.89E-02	1.28	glutaredoxin-3 isoform X2
625280141	2	2	2.09E-03	1.35E-02	1.28	cytosolic acyl coenzyme A thioester hydrolase
625258715	1	1	4.84E-03	1.90E-02	1.26	branched-chain-amino-acid aminotransferase, cytosolic isoform X3
625267589	1	1	3.94E-03	1.70E-02	1.26	alpha-actinin-4 isoform X2
625233493	1	1	4.49E-02	4.89E-02	1.26	26S proteasome non-ATPase regulatory subunit 13 isoform X3
625237309	2	2	1.62E-03	1.30E-02	1.25	adenosylhomocysteinase
625240103	2	2	1.88E-02	3.81E-02	1.25	T-complex protein 1 subunit epsilon

Table 6.2 Mass spectrometric identification of 73 proteins from the cytosolic enriched protein fraction with ≥1.25-fold increase in the miR-378 depleted CHO cells on day 8 of cell culture.

Accession	Peptides	Unique peptides	Anova (p)	Q value	Max fold change	Protein Name
625185523	1	1	3.14E-02	5.71E-03	18.24	60S ribosomal protein L26
625188420	1	1	3.92E-02	5.85E-03	10.57	heterogeneous nuclear ribonucleoprotein A3
350539695	1	1	1.13E-02	5.34E-03	7.67	protein disulfide-isomerase precursor
346421364	1	1	4.09E-02	5.85E-03	7.35	calreticulin precursor
625242946	1	1	4.88E-02	6.22E-03	4.95	protein S100-A4
625223066	1	1	3.69E-02	5.85E-03	4.78	60S ribosomal protein L22
625195560	1	1	4.30E-02	6.00E-03	4.45	protein disulfide-isomerase A6
625229196	2	2	4.03E-02	5.85E-03	4.13	40S ribosomal protein S6
354478978	1	1	2.80E-02	5.71E-03	3.70	protein S100-A6
625203562	1	1	2.20E-02	5.51E-03	2.84	14-3-3 protein beta/alpha

354495666	1	1	3.30E-02	5.71E-03	2.74	60S ribosomal protein L27a
625237172	1	1	3.93E-02	5.85E-03	2.67	cathepsin D
625272649	1	1	4.01E-02	5.85E-03	2.66	nucleophosmin isoform X4
625191956	1	1	3.70E-02	5.85E-03	2.64	glucosidase 2 subunit beta
625221706	5	5	2.41E-02	5.51E-03	2.60	alpha-enolase isoform X2
354499455	2	2	3.84E-02	5.85E-03	2.34	60S ribosomal protein L29
625193837	1	1	4.78E-02	6.22E-03	2.31	60S ribosomal protein L4
346986359	2	2	3.15E-02	5.71E-03	2.28	calreticulin precursor
350539629	1	1	4.93E-02	6.22E-03	2.21	40S ribosomal protein S4
625278207	1	1	2.91E-02	5.71E-03	2.09	transcription elongation factor B polypeptide 2
625234436	3	3	3.74E-02	5.85E-03	2.04	acyl-CoA-binding protein
625290232	1	1	2.66E-02	5.51E-03	1.89	60S ribosomal protein L18 isoform X2
625194917	1	1	1.47E-02	5.34E-03	1.88	14-3-3 protein gamma
350537945	7	7	2.26E-02	5.51E-03	1.85	peroxiredoxin-1
625225560	3	3	4.53E-02	6.15E-03	1.85	heterogeneous nuclear ribonucleoprotein A1
625265794	1	1	1.13E-02	5.34E-03	1.84	60S ribosomal protein L35a
350537423	4	4	3.03E-02	5.71E-03	1.79	78 kDa glucose-regulated protein precursor
354484084	2	2	3.84E-02	5.85E-03	1.77	40S ribosomal protein S3a
354487474	5	5	3.26E-02	5.71E-03	1.75	endoplasmin
625242866	3	3	3.26E-02	5.71E-03	1.74	tropomyosin alpha-3 chain
625271377	1	1	3.10E-02	5.71E-03	1.73	peptidyl-prolyl cis-trans isomerase FKBP4
350538733	1	1	3.20E-02	5.71E-03	1.71	60S ribosomal protein L13
625218325	1	1	1.98E-02	5.45E-03	1.71	Y-box-binding protein 3, partial
625219233	2	2	1.46E-02	5.34E-03	1.71	heterogeneous nuclear ribonucleoprotein D0, partial
354497356	1	1	3.08E-02	5.71E-03	1.69	ADP-ribosylation factor 3

625286340	1	1	4.66E-02	6.16E-03	1.66	annexin A5
354507332	1	1	1.72E-02	5.34E-03	1.63	60S ribosomal protein L8
346227155	3	3	2.10E-02	5.46E-03	1.62	elongation factor 2
625223526	1	1	1.81E-02	5.39E-03	1.59	40S ribosomal protein S3
625233305	3	3	5.51E-04	2.52E-03	1.58	ubiquitin carboxyl-terminal hydrolase 14 isoform X4
625263837	1	1	1.23E-02	5.34E-03	1.55	reticulocalbin-3 isoform X2
625204380	1	1	2.26E-02	5.51E-03	1.55	chloride intracellular channel protein 4
354506476	2	2	1.53E-02	5.34E-03	1.54	glutathione S-transferase Mu 7
625190571	3	3	2.60E-02	5.51E-03	1.47	tropomyosin alpha-4 chain
354497863	1	1	2.57E-02	5.51E-03	1.45	RNA-binding protein FUS isoform X1
625203986	2	2	2.64E-02	5.51E-03	1.44	peptidyl-prolyl cis-trans isomerase A
354475571	1	1	2.89E-02	5.71E-03	1.42	NSFL1 cofactor p47
625198438	1	1	2.48E-02	5.51E-03	1.42	ran-specific GTPase-activating protein
625282303	1	1	1.94E-02	5.45E-03	1.40	serine/threonine-protein kinase SMG1 isoform X3
625223520	2	2	1.48E-02	5.34E-03	1.39	serpin H1 isoform X1
625190791	1	1	2.90E-03	4.00E-03	1.39	60S ribosomal protein L7a-like
354471594	1	1	2.06E-02	5.45E-03	1.38	cathepsin B
625227859	3	3	1.48E-02	5.34E-03	1.37	glutathione S-transferase Mu 6
625258134	1	1	3.04E-02	5.71E-03	1.37	sulfiredoxin-1 isoform X2
625225201	1	1	1.69E-02	5.34E-03	1.35	annexin A2 isoform X1
625262042	2	2	1.34E-02	5.34E-03	1.35	heat shock protein beta-1
625180993	2	2	1.11E-02	5.34E-03	1.34	eukaryotic initiation factor 4A-I-like
350540646	1	1	4.13E-02	5.85E-03	1.32	phosphoglycerate kinase 1
625222844	3	3	2.44E-04	1.80E-03	1.32	septin-11 isoform X1

625240830	1	1	2.66E-02	5.51E-03	1.31	nucleoside diphosphate kinase B
625222011	1	1	4.63E-02	6.16E-03	1.30	inosine-5'-monophosphate dehydrogenase 2 isoform X1
625289462	1	1	3.93E-02	5.85E-03	1.30	calcium-regulated heat stable protein 1
625199022	1	1	1.31E-02	5.34E-03	1.30	m7GpppX diphosphatase
354489619	1	1	2.61E-02	5.51E-03	1.29	isocitrate dehydrogenase [NADP] cytoplasmic
625249460	2	2	1.10E-02	5.34E-03	1.29	src substrate cortactin
625202098	1	1	4.66E-02	6.16E-03	1.29	myosin light polypeptide 6-like
625256794	2	2	1.36E-04	1.60E-03	1.29	fatty acid-binding protein, adipocyte
350538479	2	2	3.11E-02	5.71E-03	1.28	tubulin beta-6 chain
625267589	5	5	3.71E-03	4.07E-03	1.28	alpha-actinin-4 isoform X2
625206697	1	1	2.49E-02	5.51E-03	1.26	ATP-binding cassette sub-family F member 1
354483012	2	2	1.40E-02	5.34E-03	1.25	heterogeneous nuclear ribonucleoprotein R
625249889	1	1	4.29E-03	4.53E-03	1.25	caldesmon isoform X3
354465044	1	1	2.33E-02	5.51E-03	1.25	rab GDP dissociation inhibitor beta

Table 6.3 Mass spectrometric identification of 7 proteins from the membrane protein enriched fraction with ≥1.25-fold increase in the miR-378 depleted CHO cells on day 4 of cell culture.

Accession	Peptides	Unique peptides	Anova (p)	Q value	Max fold change	Protein name
350538167	1	1	2.35E-02	3.75E-02	1.52	calnexin precursor
354495613	2	2	2.29E-02	3.71E-02	1.37	thrombomodulin
625263837	3	3	2.63E-03	2.00E-02	1.34	reticulocalbin-3
350537945	3	3	3.71E-03	2.34E-02	1.33	peroxiredoxin-1
625249714	1	1	3.11E-02	3.99E-02	1.31	perilipin-4 isoform X14
625282737	1	1	4.67E-02	4.88E-02	1.31	protein dpy-30 homolog
625215083	3	3	1.83E-02	3.38E-02	1.30	guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1

Table 6.4Mass spectrometric identification of 72 proteins from the membrane protein enriched fraction with ≥1.25-fold increase in the miR-378 depleted CHO cells on day 8 of cell culture.

Accession	Peptides	Unique peptides	Anova (p)	Q value	Max fold change	Protein Name
625244585	2	2	1.10E-02	2.45E-02	2.52	histone H2A.V isoform X2
354496412	1	1	6.56E-03	2.14E-02	2.45	histone H1.0
354480100	5	5	2.14E-02	3.33E-02	2.17	histone H2B type 1
354494381	1	1	1.63E-04	1.65E-02	2.12	fibronectin isoform X1
354494231	1	1	9.56E-03	2.38E-02	2.08	high mobility group nucleosome-binding domain-containing protein 5
345842361	1	1	3.64E-02	4.43E-02	2.08	high mobility group protein HMG-I/HMG-Y
625206001	7	7	1.31E-02	2.76E-02	2.06	histone H3.1-like
625285909	3	3	1.01E-02	2.38E-02	1.84	histone H2A type 1-H-like
625229196	1	1	4.46E-02	4.84E-02	1.81	40S ribosomal protein S6
625205207	1	1	2.86E-03	1.65E-02	1.77	rRNA 2'-O-methyltransferase fibrillarin
354480104	6	6	1.53E-02	2.89E-02	1.74	histone H1.4 isoform X1
625289934	1	1	3.81E-02	4.46E-02	1.73	calumenin isoform X2

350537403	1	1	2.63E-02	3.77E-02	1.68	DNA topoisomerase 2-alpha
625262546	1	1	3.48E-02	4.36E-02	1.68	replication protein A 14 kDa
625209863	1	1	1.44E-02	2.87E-02	1.67	alpha-parvin
625234125	4	4	1.10E-02	2.45E-02	1.63	elongation factor 1-gamma
350538167	3	3	1.78E-03	1.65E-02	1.60	calnexin precursor
625284147	1	1	8.52E-03	2.30E-02	1.50	legumain
350539823	1	1	4.09E-04	1.65E-02	1.50	heat shock cognate 71 kDa protein
625204124	1	1	4.42E-02	4.82E-02	1.47	platelet glycoprotein 4
625256908	1	1	3.23E-03	1.65E-02	1.47	septin-2
625211254	2	2	2.82E-02	3.86E-02	1.47	plectin isoform X1
625260069	1	1	7.33E-03	2.14E-02	1.44	14-3-3 protein epsilon
354504493	2	2	1.54E-03	1.65E-02	1.44	6-phosphogluconate dehydrogenase, decarboxylating isoform X1
625231575	2	2	7.14E-03	2.14E-02	1.43	eukaryotic initiation factor 4A-II isoform X1
625274484	1	1	4.37E-02	4.80E-02	1.42	serum albumin isoform X3
625262669	1	1	3.56E-02	4.39E-02	1.42	cellular nucleic acid-binding protein isoform X2
625243141	1	1	2.10E-02	3.32E-02	1.41	ATP-dependent RNA helicase DDX39A
625216841	1	1	6.75E-03	2.14E-02	1.41	coronin-1B
625292335	1	1	1.81E-02	3.04E-02	1.40	high mobility group protein B2 isoform X2
354489619	1	1	7.40E-03	2.14E-02	1.40	isocitrate dehydrogenase [NADP] cytoplasmic
625215758	1	1	2.88E-03	1.65E-02	1.39	enoyl-CoA delta isomerase 1, mitochondrial isoform X1
354483223	1	1	2.25E-02	3.44E-02	1.39	prolyl 4-hydroxylase subunit alpha-1 isoform X1
354467247	1	1	2.41E-03	1.65E-02	1.39	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 9
625284339	3	3	1.46E-02	2.87E-02	1.38	succinate dehydrogenase ubiquinone] iron-sulfur subunit, mitochondrial

625231917	1	1	2.18E-02	3.36E-02	1.38	guanine nucleotide-binding protein subunit beta-4
625238921	1	1	4.00E-02	4.53E-02	1.38	EH domain-containing protein 4 isoform X2
354500398	1	1	1.25E-03	1.65E-02	1.37	ubiquitin-like modifier-activating enzyme 1
625190571	1	1	2.57E-02	3.74E-02	1.35	tropomyosin alpha-4 chain
354485048	1	1	3.51E-02	4.37E-02	1.35	polymerase I and transcript release factor
354485701	1	1	4.91E-04	1.65E-02	1.35	stomatin-like protein 2, mitochondrial
354492573	1	1	4.19E-02	4.72E-02	1.35	actin-related protein 3B
354465900	2	2	1.57E-02	2.93E-02	1.35	ATP-dependent RNA helicase DDX3X
354484391	1	1	2.57E-02	3.74E-02	1.35	14-3-3 protein zeta/delta
625190862	1	1	1.66E-02	2.97E-02	1.34	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 10, mitochondrial isoform X1
625240103	1	1	4.16E-03	1.83E-02	1.34	T-complex protein 1 subunit epsilon
625188420	1	1	4.22E-02	4.73E-02	1.33	heterogeneous nuclear ribonucleoprotein A3
625211596	1	1	6.89E-04	1.65E-02	1.33	60S ribosomal protein L7
625248231	1	1	2.96E-03	1.65E-02	1.33	ADP/ATP translocase 1
625251833	1	1	2.86E-02	3.86E-02	1.32	hydroxymethylglutaryl-CoA lyase, mitochondrial
350540646	2	2	1.59E-02	2.93E-02	1.32	phosphoglycerate kinase 1
625254434	1	1	6.32E-03	2.14E-02	1.32	superoxide dismutase [Mn], mitochondrial isoform X2
625249635	1	1	5.33E-03	2.06E-02	1.31	lon protease homolog, mitochondrial
625208910	1	1	4.97E-02	5.03E-02	1.30	septin-7 isoform X1
354486011	1	1	4.29E-02	4.74E-02	1.30	acyl-coenzyme A thioesterase 1 isoform X1
625213146	1	1	9.70E-03	2.38E-02	1.29	integrin beta-1 isoform X1
625232358	3	3	4.03E-04	1.65E-02	1.29	lipoprotein lipase

625279800	1	1	4.90E-02	5.00E-02	1.28	caveolin-1 isoform X1
625235290	1	1	4.73E-02	4.89E-02	1.28	peroxiredoxin-5, mitochondrial
625215083	2	2	1.26E-03	1.65E-02	1.28	guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1 isoform X1
346986359	3	3	2.62E-03	1.65E-02	1.27	elongation factor 1-alpha 1
625201330	1	1	1.81E-02	3.04E-02	1.27	cell division control protein 42 homolog
625288359	2	2	1.73E-02	3.01E-02	1.27	dephospho-CoA kinase domain-containing protein
552953713	1	1	1.43E-02	2.87E-02	1.26	40S ribosomal protein S7
354482483	1	1	2.73E-02	3.82E-02	1.26	vimentin
625243995	2	2	3.76E-02	4.46E-02	1.26	leucine-rich PPR motif-containing protein, mitochondrial isoform X2
625184898	1	1	3.26E-02	4.24E-02	1.25	39S ribosomal protein L12, mitochondrial isoform X1
625236680	1	1	1.11E-02	2.45E-02	1.25	60 kDa heat shock protein, mitochondrial
625183009	2	2	1.37E-02	2.85E-02	1.25	triosephosphate isomerase
354486540	2	2	2.91E-03	1.65E-02	1.25	hydroxymethylglutaryl-CoA synthase, mitochondrial
625224152	1	1	2.50E-02	3.72E-02	1.25	nuclear body protein SP140-like isoform X1
625291524	1	1	3.79E-02	4.46E-02	1.25	mitochondrial import inner membrane translocase subunit Tim13 isoform X3, partial

6.2 Experimental Design, Materials and Methods

6.2.1 Subcellular protein extraction and in-solution protein digestion

Triplicate biological samples for control and miR-378 depleted cells were collected on day 4 and day 8 of batch cultures. Subcellular protein enrichment was achieved using the Mem-Per Plus Membrane protein extraction kit (#89842, Thermo Fisher Scientific) which yielded a cytosolic and membrane protein enriched fraction. Protein concentration was determined using the Quick Start Bradford assay (Bio-rad). Equal concentrations (100 µg) of protein from each sample were purified and trypsin digested for mass spectrometry using the filter-aided sample preparation method as previously described (Coleman et al., 2017). The resulting peptide samples were purified using Pierce C18 spin columns then dried using vacuum centrifugation and suspended in 2% acetonitrile and 0.1% trifluoracetic acid in LC grade water prior to LC-MS/MS analysis.

6.2.2 Label-free liquid chromatography mass spectrometry

Quantitative label-free liquid-chromatography mass spectrometry (LC–MS/MS) analysis of mir-378-spg and NC-spg membrane and cytosolic fractions from day 4 and day 8 was carried out using a Dionex UltiMate™ 3000 RSLCnano system (Thermo Fisher Scientific) coupled to a hybrid linear LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific). LC-MS/MS methods were applied as previously described (Henry and Meleady 2017). A 5 µL injection of each sample was loaded onto a C18 trapping column (PepMap100, C18, 300 µm × 5 mm; Thermo Fisher Scientific). Each sample was desalted for 5 min using a flow rate of 25 µL/min with 2% ACN, 0.1% TFA before being switched online with the analytical column (PepMap C18, 75 µm ID × 250 mm, 3 µm particle and 100 Å pore size; (Thermo Fisher Scientific)). Peptides were eluted using a binary gradient of Solvent A (2% ACN and 0.1% formic acid in LC grade water) and Solvent B (80% ACN and 0.08% formic acid in LC grade water). The following gradient was applied; 6–25% solvent B for 120 min and 25–50% solvent B in a further 60 min at a column flow rate of 300 nL/min. Data was acquired with Xcalibur software, version 2.0.7 (Thermo Fisher Scientific). The LTQ Orbitrap XL was operated in data-

dependent mode with full MS scans in the 400–1200 m/z range using the Orbitrap mass analyser with a resolution of 30,000 (at m/z 400). Up to three of the most intense ions (+1, +2, and +3) per scan were fragmented using collision-induced dissociation (CID) in the linear ion trap. Dynamic exclusion was enabled with a repeat count of 1, repeat duration of 20 s, and exclusion duration of 40 s. All tandem mass spectra were collected using a normalised collision energy of 32%, and an isolation window of 2 m/z with an activation time of 30 ms.

6.2.3 Quantitative Label-free LC-MS/MS Data Analysis

Protein identification was achieved using Proteome Discoverer 2.1 with the Sequest HT and MASCOT algorithm followed by Percolator validation (Käll et al., 2007) to apply a false-discovery rate <0.01. Data was searched against the NCBI Chinese Hamster (*Cricetulus griseus*) protein database containing 44,065 sequences (fasta file downloaded November 2015). The following search parameters were used for protein identification: (1) precursor mass tolerance set to 20 ppm, (2) fragment mass tolerance set to 0.6 Da, (3) up to two missed cleavages were allowed, (4) carbamidomethylation of cysteine set as a static modification and (5) methionine oxidation set as a dynamic modification. The complete lists of all identified proteins from the cytosolic and membrane enriched fractions of day 4 and day 8 cell cultures of the control (NC378-spg) and miR-378-spg are provided in supplementary table S2.

Quantitative label-free data analysis was performed using Progenesis QI for Proteomics (version 2.0; Nonlinear Dynamics, a Waters company) as described by the manufacturer (www.nonlinear.com). To counteract potential drifts in retention time a reference run was assigned to which all MS data files were aligned. The triplicate samples from the two experimental groups (NC-378-spg and miR-378-spg) were set up for differential analysis and label-free relative quantitation was carried out after peak detection, automatic retention time calibration and normalisation to account for experimental variation. The experimental analyses performed compared the three biological replicates for control cells to miR-378-spg triplicates for each time point and subcellular fraction collected. The following settings were applied to filter peptide features (1) peptide

features with a one-way ANOVA p-value <0.05 between experimental groups, (2) mass peaks with charge states from +1 to +3 and (3) greater than one isotope per peptide. The normalised data is transformed prior to statistical analysis, using an arcsinh transformation to meet the assumptions of the one-way ANOVA test. A mascot generic file (mgf) was generated from all exported MS/MS spectra which satisfied the peptide filters, the mgf was used for peptide and protein identification in Proteome Discoverer. Protein identifications were imported into Progenesis and considered differentially expressed if they passed the following criteria: (i) a protein one-way ANOVA p-value <0.05 and (ii) a ≥1.25-fold change in relative abundance between the two experimental groups. All differentially expressed proteins identified between NC378-spg and miR-378-spg cells are reported in supplementary table S1.

Heat maps illustrating protein abundances for statistically significant and differentially expressed proteins were designed using ggplot2 in R-studio. The normalised abundance values of differentially expressed proteins were determined using Progenesis QI for Proteomics and were loaded as a txt file into R-studio and the data was log2 transformed. Hierarchical Pearson clustering was then performed on Z-score normalised intensity values by clustering both samples and proteins.

6.3 References

Costello, A., Coleman, O., Lao, N.T., Henry, M., Meleady, M., Barron, N. & Clynes, M. (2018) Depletion of Endogenous miRNA-378-3p Increases Peak Cell Density of CHO DP12 Cells and is Correlated with Elevated Levels of Ubiquitin Carboxyl-Terminal Hydrolase 14. *Journal of Biotechnology 288* pp 30-40.

Coleman, O., Henry, M., Clynes, M., & Meleady, P. (2017) Filter-aided sample preparation (FASP) for improved proteome analysis of recombinant Chinese hamster ovary cells. *Heterologous Protein Production in CHO Cells: Methods and Protocols*, 187-194.

Henry, M., & Meleady, P. (2017) Clinical proteomics: Liquid Chromatography–Mass spectrometry (LC–MS) purification systems. *Methods in Mol. Bol.* Springer, pp. 375-388

Käll, L., Canterbury, J.D., Weston, J., Noble, W.S. & MacCoss, M.J. (2007) Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nature methods*. Vol. 4(11), pp 923-925.

6.4 Appendix C

Supplementary tables S1 and S2 are provided on the disk attached to the back cover of the thesis.

7 Chapter Seven: Increased growth rate and productivity following stable depletion of miR-7 in a mAb producing CHO cell line causes an increase in proteins associated with the Akt pathway and ribosome biogenesis

Accepted in *Journal of Proteomics* January 2019

Authors: Orla Coleman, Srinivas Suda, Justine Meiller, Michael Henry, Markus Riedl, Niall Barron, Paula Meleady and Martin Clynes

For this publication I was responsible for protein sample preparation, data analysis and manuscript preparation. I performed subcellular protein enrichment, total proteome analysis, protein preparation for MS analysis and all data analyses and bioinformatics. I deposited all MS files to the public repository of proteomic data. I prepared, wrote and submitted the manuscript to *Journal of Proteomics* and performed all revisions.

Abstract

Cell line engineering using microRNAs represents a desirable route for improving the efficiency of recombinant protein production by CHO cells. In this study we generated stable CHO DP12 cells expressing a miR-7 sponge transcript which sequesters miR-7 from its endogenous targets. Depletion of miR-7 results in a 65% increase in cell growth and >3-fold increase in yield of secreted IgG protein. Quantitative label-free LC-MS/MS proteomic profiling was carried out to identify the targets of miR-7 and understand the functional drivers of the improved CHO cell phenotypes. Subcellular enrichment and total proteome analysis identified more than 3000 proteins per fraction resulting in over 5000 unique proteins identified per time point analysed. Early stage culture analysis identified 117 proteins overexpressed in miR-7 depleted cells. A subset of these proteins are involved in the Akt pathway which could be the underlying route for cell density improvement and may be exploited more specifically in the future. Late stage culture identified 160 proteins overexpressed in miR-7 depleted cells with some of these involved in ribosome biogenesis which may be causing the increased productivity through improved translational efficiency. This is the first in-depth proteomic profiling of the IgG producing CHO DP12 cell line stably depleted of miR-7.

Significance

Chinese hamster ovary (CHO) cells are the mammalian cell expression system of choice for production of recombinant therapeutic proteins. There is much research ongoing to characterise CHO cell factories through the application of systems biology approaches that will enable a fundamental understanding of CHO cell physiology, and as a result, a better knowledge and understanding of recombinant protein production. This study profiles the proteomic effects of microRNA-7 depletion on the IgG producing CHO DP12 cell line. This is one of the very few studies that attempt to identify the functioning proteins driving improved CHO cell phenotypes resulting from microRNA manipulation. Using subcellular enrichment and total proteome analysis we identified over 5000 unique proteins in miR-7 depleted CHO cells. This work has identified a cohort of proteins involved in the Akt pathway and ribosome biogenesis. These proteins may drive improved CHO cell phenotypes and are of great interest for future work

Highlights

1. Stable depletion of miR-7 in CHO DP12 is achieved using sponge decoy technology

2. MiR-7 depletion leads to a 65% increase in viable cell density

3. Productivity increases by over 3-fold on day 3 and 8

4. Subcellular enrichment identifies over 5000 proteins per experimental group

5. >100 proteins are identified as potential direct targets of miR-7 on day 3 and 8

6. The Akt pathway and ribosome biogenesis are involved in driving the miR-7 phenotype

7.1 Introduction

Chinese Hamster Ovary (CHO) cells have been the predominant host cell of choice for manufacturing mammalian recombinant therapeutics since their introduction into the biopharmaceutical market. These cell factories are proficient in protein folding, post-translational modifications and scalability which makes them the most frequently used mammalian cell hosts in a multi-billion pharmaceutical industry (Estes and Melville, 2014). Improvement of CHO cells toward maximum product yields, faster cell growth and enhanced product quality in an affordable, efficient and robust platform is a major goal in the biopharmaceutical industry. Historically efforts to achieve these goals have focused on media optimization (Altamirano et al., 2004; E. P. Huang et al., 2007; Kishishita et al., 2015) and process development (Bollati-Fogolín et al., 2005; Huang et al., 2004; Trummer et al., 2006). More recently, genetic engineering of CHO cells has emerged to compare pharmaceutically-relevant phenotypes.

A promising approach for cell line engineering is the manipulation of microRNAs (miRNAs) to positively impact bioprocess relevant CHO cell phenotypes. miRNAs are small, non-coding, regulatory RNA molecules that bind to the 3' untranslated region (UTR) of target mRNAs. These molecules are characterised by their sequence complementarity with multiple gene targets, their highly conserved nature across species and the lack of translational burden placed on the CHO cell (Hackl et al., 2012). Although previous studies have shown functional effects following miRNA manipulations, determining the causative molecules behind phenotypic changes is of upmost importance if these molecules are to be used by the biopharma industry as engineering targets to enhance phenotypic traits. In silico prediction of miRNA targets are largely based on miRNA seed region complementarity to the 3' UTR of mRNAs; however increasing evidence suggests targeting can also be mediated to other sites (Grimson et al., 2007; Shin et al., 2010b). The application of in silico predictions can therefore be hindered and may not fully catalog the targets of miRNAs. With that in mind, proteomics has been employed to accurately identify and quantify protein changes resulting from miRNA manipulation in various studies. Mass spectrometry (MS) -based proteomics is increasingly contributing to our understanding of the dynamics,

interactions, and roles that proteins play, advancing our understanding of biology on a systems wide level for a wide range of applications (Angel et al., 2012). Incorporating omics-based approaches into current bioproduction processes will maximise gains from CHO engineering and bioprocess improvements (Stolfa et al., 2018). Although miRNA engineering of CHO cells has gained popularity in recent years, few studies have performed proteomic profiling to identify miRNA targets. Previous work in our group combined miRNA, mRNA and protein expression data to discern the biological roles of growth-related miRNAs (Clarke et al. 2012). The identification and integrated analysis of 51 differentially expressed (DE) miRNAs, 432 DE genes and 285 DE proteins revealed several proteins central to mRNA processing and protein synthesis to be post-transcriptionally regulated by miRNAs correlated with growth rate in CHO cells. MiR-7 has been identified as a miRNA involved in CHO cell culture and productivity (Barron et al. 2011; Meleady et al. 2012; Sanchez et al. 2014a). MiR-7 was originally identified using TLDA profiling as 8-fold down-regulated in CHO K1 cells following a temperature shift of the culture to 31^{o}C (Barron et al. 2011). Further analyses in this study revealed that transient over-expression of miR-7 significantly arrested cell growth and although it reduced total yield of production, normalized productivity was increased. Follow-up analysis identified the proteomic effect of miR-7 transient overexpression on the CHO K1 cell line and suggested a role for stathmin and catalase as possible direct targets of miR-7 in regulating cell growth (Meleady et al., 2012a). In contrast, a more recent evaluation of miR-7 following stable depletion of the miRNA via sponge decoy technology displayed improved maximum cell density (40%), improved viability and a 2-fold increase in yield of the protein product in CHO K1 cells (Sanchez et al., 2014b).

This study carries on from the work of Sanchez et al., using the same sponge decoy construct for miR-7 in the more industrially relevant monoclonal antibody (mAb) producing CHO DP12 cell line. Although proteomic profiling was previously reported for miR-7 in CHO cells, the major improvements for this study design is proteomic profiling of (i) the more industrially-relevant mAb producing CHO DP12 cell line, (ii) stably rather than transient transfected CHO cells, and (iii) miR-7 depletion which leads to a more attractive biopharmaceutical phenotype than overexpression. The aim of this

study is to analyse changes in protein levels resulting from miR-7 stable depletion to better understand its mode of action. Identification of upregulated proteins may determine direct targets and pathways affected by this miRNA which are causing improved biopharmaceutically relevant CHO cell phenotypes.

7.2 Materials & Methods

7.2.1 Cell Culture

An IgG-producing CHO_DP-12 (ATCC CRL-12445 clone #1934) cell line producing human anti-IL-8 mAb were cultured in chemically defined BalanCD medium (Irvine Scientific) supplemented with 4 mM L-Glutamine, 400 mM methotrexate (MTX, Sigma-Aldrich). Cells were seeded at 2 $x10^5$ cells per mL and incubated at 37˚C in 5% CO_2, in a Kuhner Climo-Shaker ISF1-XC. Viable cell density (VCD) and cell viability was routinely measured using a benchtop flow cytometer, Guava Easy Cyte HT (Millipore) and the ViaCount assay as per the manufacturer's guidelines.

7.2.2 Plasmids and Transfection

Stable knock down of miRNA-7 was achieved using the miR-7 sponge plasmids previously created and validated by our group (Sanchez et al, 2014). Sponge depletion was achieved by transfection of 2 $x10^5$ cells/mL with 3 µg of plasmid DNA (CMV-miR7-Hyg Sponge and NC vectors (control)) and 6 µL of Transit® X2 Delivery System (Mirus). Two separate transfections were performed using CMV-miR7-Hyg, and CMV-miR NC-Hyg. Transfections were carried out in triplicate (n=3). After overnight incubation cells stably integrating the plasmid were selected in media containing DMEM-F12 with 5% FBS, 400 mM MTX + Hygromycin 350 µg/mL final concentration for 6 passages before adapting them to suspension culture in BalanCD medium with 4 mM L-Glutamine, 400 mM MTX, 350 µg/mL Hygromycin.

7.2.3 Cell Phenotype Analysis

Stably selected CHO DP12 miR-7 and miR-NC sponge cells grown in suspension culture were used for phenotype analysis of cell growth and IgG productivity. Cells were passaged and split into Spin Tube bioreactors tubes (n=3) each with 2 $x10^5$ cells

per mL for miR-NC, and miR-7 sponge. Cells were grown at the same culture conditions for several days until the cell viability dropped to 70%. Daily 100 µL was removed from each tube and placed into sterile microtubes for cell counting and product titer analysis. CHO DP12 miR-7 and miR-NC cells produce the human IgG anti-IL-8 mAb which is secreted into the culture medium. IgG titer was quantified using an ELISA assay (Human IgG ELISA Quantitation kit, Bethyl Laboratories Inc., E80-104). CHO DP12 culture supernatants for miR-7-spg and miR-NC-spg were collected on day 3 and 8 by centrifugation of the cultures at 1,000 x *g* for 5 minutes and collection of supernatant. IgG production was quantified using ELISA as per the kit protocol and measured on a Multiskan GO spectrophotometer (Thermo Fisher Scientific) at 450 nm.

7.2.4 Sample preparation for label-free LC-MS/MS analysis

Biological triplicate cell pellets collected from day 3 and day 8 of culture were subjected to protein extraction using the Mem-PER plus membrane protein extraction kit (Thermo Fisher Scientific). This method enriches integral membrane proteins and membrane-associated proteins in one fraction and cytosolic proteins in a second fraction to increase proteome coverage. Separately, cell pellets from day 8 were also collected for whole cell proteome analysis by lysis using a buffer containing 7 M urea, 2 M Thiourea, 4% CHAPS, 30 mM Tris at pH 8.5. Protein concentration for all samples was determined using the QuickStart Bradford assay (Biorad). Equal concentrations (100 µg) of all protein samples were purified and digested using the FASP method (Coleman et al., 2017; Wiśniewski et al., 2009). Peptides were purified using C18 spin columns (Thermo Fisher Scientific), dried under vacuum and suspended in 2% acetonitrile and 0.1% trifluoroacetic acid prior to mass spectrometry.

7.2.5 LC-MS/MS and Label-free quantitative differential analysis

Liquid chromatography-mass spectrometry (LC-MS/MS) was performed on a Dionex UltiMate3000 nanoRSLC coupled in-line with an Orbitrap Fusion Tribrid™ mass spectrometer (Thermo Fisher Scientific). LC-MS/MS parameters were used as previously described (Kaushik et al., 2018) using a 140-minute gradient per sample. MS1 spectra were acquired over m/z 400–1500 in the Orbitrap (120 K resolution at 200 m/z), and automatic gain control (AGC) was set to accumulate 4×10^5 ions with a

maximum injection time of 100 ms. Data-dependent tandem MS analysis was performed using a top-speed approach (cycle time of 3 s), and the normalized collision energy was optimized at 35% for Collision Induced Dissociation (CID). MS2 spectra were acquired in the ion trap. The intensity threshold for fragmentation was set to 5000 and included charge states 2+ to 7+. A dynamic exclusion of 50 s was applied with a mass tolerance of 10 ppm. AGC was set to 1×10^4 with a maximum injection time set at 35 ms.

Protein identification was achieved using Proteome Discoverer v.2.1 (Thermo Fisher Scientific) with the SEQUEST HT algorithm. MS files were searched against the NCBI Chinese Hamster Ovary (*Cricetulus griseus*) protein database downloaded in November 2017, containing 24,906 sequences. The following search parameters were set for protein identifications: (i) MS/MS mass tolerance set at 0.6 Da; (ii) peptide mass tolerance set to 10 ppm; (iii) up to two missed cleavages were allowed; (iv) cysteine carbamidomethylation set as a static modification; (v) methionine oxidation set as a dynamic modification, and (vi) Sequest HT XCorr was set at 1.9 for +1 ions, 2.2 for +2 ions and 3 for +3 ions. Quantitative label-free analysis was performed using Progenesis QI for Proteomics v.3.1 (Non-Linear Dynamics, Waters). Experimental analyses compared mir-7-spg cells to control cells on both culture time-points of day 3 and 8 using all generated protein fractions, i.e. membrane, cytosolic and total proteome. Only peptide ions with charge states +1, +2 and +3 were allowed and normalised. The normalised peptide abundance data is transformed prior to statistical analysis, using an arcsinh transformation to meet the assumptions of the one-way ANOVA test. Peptides with a one-way ANOVA p-value ≤0.05 between experimental groups were exported and identified using Proteome Discoverer as described. Protein identifications were imported into Progenesis and considered differentially expressed proteins if they passed the following criteria: (i) a protein level one-way ANOVA *p*-value ≤0.05 between experimental groups, (ii) a minimum of 2 peptides contributing to a protein identification and (iii) a minimum fold-change in abundance between mir-7-spg and NC samples of ±1.5 fold.

7.2.6 Bioinformatics

In silico prediction of mRNA targets of miR-7 was performed using TargetScan 7.1 (Agarwal et al. 2015; Lewis, Burge, and Bartel 2005) and selecting a human species search for predicted targets of miR-7-5p due to the lack of a CHO option. The STRING database (http://string-db.org/ ; version 10.5) was used to classify the differentially expressed proteins according to enriched biological processes and molecular functions determined by gene ontology functional annotation for input protein lists (Mering et al., 2003; Szklarczyk et al., 2014). KEGG pathway analysis was performed using the DAVID classification tool to identify enriched pathways associated with the input protein lists (D. W. Huang et al., 2007).

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Vizcaíno et al., 2016) partner repository with the dataset identifier PXD011951.

7.3 Results

7.3.1 Stable depletion of miR-7 in CHO DP12 increases cell growth and mAb production

The previous observation by Sanchez et al. (2014) that miR-7 depletion in CHO K1 SEAP-producing cells enhanced cell culture concentration and protein yield led to this analysis of the model IgG-producing CHO DP12 cell line. The same sponge constructs were employed to stably transfect the DP12 cell line with either a miR-7 sponge vector (miR-7-spg) or a control sponge (NC) containing non-specific sequences in place of the miR-7 miRNA response elements of the construct. DP12 cells with sequestered miR-7 achieved a 65% increase in viable cell density (VCD/mL) on day 8 of culture compared to NC cells (Figure 7.1A). Product titre was assessed in DP12 cells using human IgG ELISA quantitation. miR-7 depleted cells yielded 3-fold more secreted IgG than NC cells on day 3 and 8 of batch culture (Figure 7.1 B).

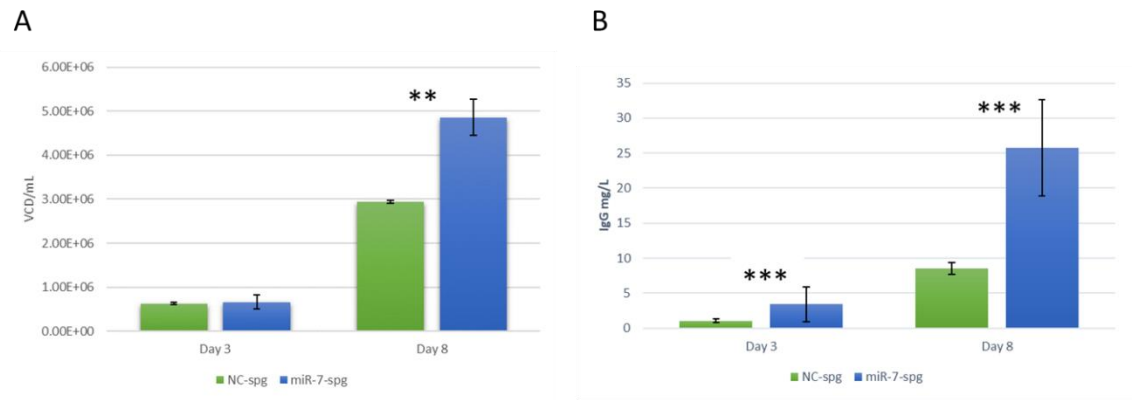


*Figure 7.1 Phenotypic effect of miR-7 depletion on CHO DP12 cell at time points selected for proteomic analysis. (A) Cell growth shown as viable cells per mL of CHO DP12 cells in batch, suspension culture stably transfected with a negative control sponge decoy transcript (green) or miR-7 sponge transcript (blue). (B) Total yield of mAb IgG produced by miR-7 depleted cells (blue) and negative control cells (green). ** p<0.005, *** p<0.0005 (t-test). At least three biologically individual cell cultures from each group were measured in triplicate.*

7.3.2 Proteomic analysis of miR-7 depleted CHO DP12 cells

Samples from three independent batch cultures were collected at day 3 and day 8 for proteomic analysis. Sub-cellular proteomic analysis was carried out for both time-points; additionally, whole proteome analysis was also carried out for day 8 of culture. Fractionation of the cells into membrane-enriched and cytosolic-enriched protein fractions allows an in-depth sub-cellular analysis of the effect of miR-7 depletion on the CHO cell compartments and to improve proteome coverage. Total proteomes from day 8 of culture were extracted and analysed to achieve an understanding of the overall effect miR-7 depletion has on DP12 cells. All fractions (membrane, cytosolic and total) were analysed by quantitative label-free mass spectrometry (LC-MS/MS). A state-of-the-art Orbitrap Fusion Tribrid™ mass spectrometer was used to understand the role of miR-7 in CHO cells. Combining total proteome and subcellular proteome analysis followed by advanced mass spectrometry provides a comprehensive overview of the mode of action of miR-7 to achieve increased cell growth and productivity. We identified over 3000 proteins in each fraction, using the Sequest HT algorithm and applying a false discovery rate of <0.05 (Figure 7.2). Median sequence coverage was not hindered by subcellular enrichment or culture time points, evidenced in Figure 7.2B. Although fewer proteins were identified in the cytosolic fractions the median sequence coverage is increased compared to all other protein fractions suggesting there is a strong

selective enrichment. Label-free data analysis focused on the identification of differentially expressed proteins with an increased expression in the miR-7 sponge cells compared to NC. The sponge decoy technology used here binds to and sequesters endogenous miR-7 in-turn allowing targets of miR-7 to be expressed which had previously subsided. By identifying proteins with an increased expression after the miRNA has been isolated potential targets and thus pathways underpinning the enhanced phenotypic effect can be determined. Use of subcellular enrichment in the sample preparation allowed for an increased proteome coverage with over 1500 unique proteins identified in the membrane fractions of miR-7-spg cells on both days and approximately 800 unique proteins to the cytosolic fractions of both time points (Figure 7.2 C-D). Although over 3000 proteins were identified in each fraction, the number of unique proteins for miR-7-spg cells on day 3 and 8 is 5025 and 5022 respectively (Figure 7.2 C-D). All identified proteins for each subcellular protein fraction, condition and time point are reported in supplementary table S1. This proves subcellular enrichment successfully retrieves cohorts of unique proteins thus increasing protein identifications and expanding the knowledgebase surrounding CHO cells and miRNA effects. This study presents the first in-depth proteomic analysis on the effect of miR-7 stable depletion in CHO cells.

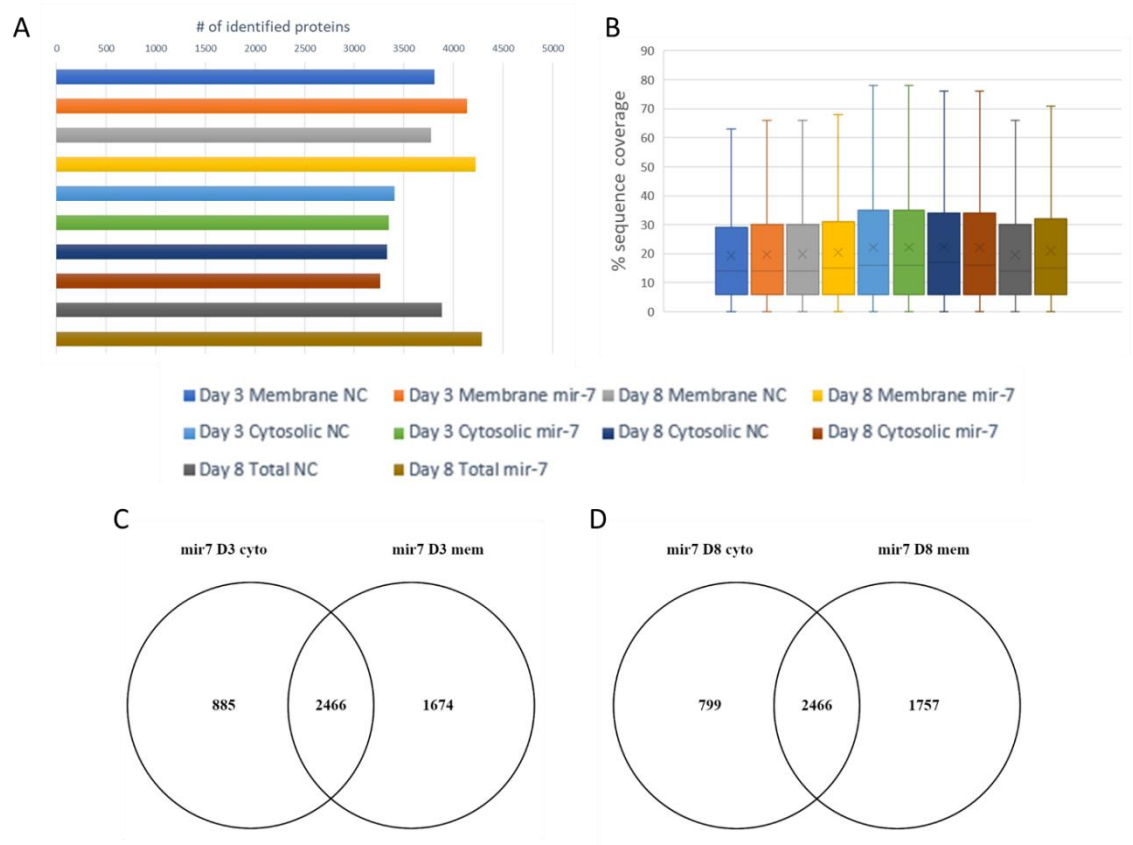


Figure 7.2. Subcellular enrichment and total proteome analysis enables in-depth analysis of miR-7 effects on the proteome of CHO-DP12 cells.(A) Number of protein identifications in the indicated protein fractions and experimental conditions. (B) The box and whisker plots depict the distribution of protein sequence coverage (% coverage of tryptic peptides per protein) for the indicated protein fractions and experimental conditions. (C, D) The Venn diagrams depict the overlap of identified proteins between the cytosolic (cyto) and membrane (mem) enriched protein fractions for (C) day 3 and (D) day 8 cell cultures of miR-7 depleted CHO DP12 cells.

7.3.2.1 Early-stage culture analysis

The effect of miR-7 depletion was assessed during the exponential stage of culture by collecting samples for proteomic analysis on day 3 of growth. Day 3 samples for miR-7-spg and NC were subjected to subcellular enrichment which yielded a membrane protein enriched and cytosolic protein enriched fraction. Label-free quantitative analysis of the membrane and cytosolic fraction was carried out separately in Progenesis QI for Proteomics. Differentially expressed (DE) proteins between miR-7-spg and NC have a minimum of 2 peptides contributing to the protein identification, an ANOVA *p*-value ≤0.05 and a minimum fold-change between experimental groups of 1.5-fold. The membrane fraction contained 130 significantly DE proteins with 89 proteins having a higher expression in the miR-7-spg samples. The cytosolic fraction identified 58 DE

proteins, 29 of which having the highest expression in miR-7-spg. These overexpressed proteins potentially represent direct targets of miR-7. The top 10 overexpressed proteins from both subcellular fractions are shown in table 7.1 (excluding contaminating keratin proteins). All differentially expressed proteins are shown in supplementary table S2

Table 7.1 Top 10 differentially expressed proteins with increased expression in miR-7-spg cells compared to NC cells for the cytosolic and membrane fractions of day 3

Cytosolic

Accession	Peptides	Anova (p)	Fold change	Description
EGW11504.1	5	2.3E-05	2.32	Dihydrofolate reductase
EGV97529.1	2	3.5E-02	2.13	Striatin-3
EGW00013.1	2	1.0E-02	2.10	Hippocalcin-like protein 1
EGW10295.1	12	1.5E-02	1.97	E3 SUMO-protein ligase RanBP2
EGW10493.1	2	1.6E-03	1.84	Protein ariadne-2-like
EGW13185.1	2	2.0E-02	1.83	putative ATP-dependent RNA helicase DHX36
EGV96932.1	2	1.6E-03	1.79	Spermine synthase
EGW12370.1	6	2.2E-04	1.79	Glutathione S-transferase P 2
EGW11950.1	5	4.4E-04	1.72	BAG family molecular chaperone regulator 3
EGW09430.1	5	6.6E-04	1.65	Disks large-associated protein 5

Membrane

Accession	Peptides	Anova (p)	Fold change	Description
EGW14444.1	2	7.0E-03	79.75	DNA damage-binding protein 1
EGV96475.1	2	8.8E-03	47.88	Drebrin
EGV98391.1	2	3.6E-03	36.26	Calpastatin
EGW13056.1	2	1.3E-03	14.16	60S ribosomal protein L19
EGW11553.1	2	3.0E-03	13.95	Importin-4
EGV94774.1	2	3.4E-02	9.10	Histone H3
EGV94642.1	3	2.5E-02	8.40	Erlin-2
NP_001243811.1	3	2.3E-02	6.66	thioredoxin reductase 1, cytoplasmic
EGW07491.1	2	1.4E-02	6.45	Inositol 1,4,5-trisphosphate receptor type 2
EGV99480.1	2	3.1E-03	6.03	Calponin-2

All overexpressed proteins in miR-7-spg from the cytosolic and membrane fraction (117 proteins) were assessed together to provide a comprehensive insight into early-stage culture effects of miR-7. These proteins are controlled either by direct binding with the miRNA or by a secondary downstream effect as a result of a binding partner being directly targeted by miR-7. To elucidate the identities of miR-7 driven phenotypes, early-stage culture proteins were assessed in terms of biological processes and

molecular functions using Gene Ontology (GO) classification within the STRING software (Franceschini et al., 2012) (Table 7.2). The most enriched biological processes include RNA processing, mRNA metabolic process and RNA splicing all with an FDR ≤0.0001 and >10 proteins involved. Similarly, enriched molecular functions include poly(A) RNA binding, RNA binding and nucleic acid binding all with an FDR <0.00001 and >30 proteins involved in each enriched function. Taken together these results indicate there may be an increase in RNA processing molecules and thus translation during early stage culture in the miR-7 depleted cells.

Biological Process

#GO ID	GO description	observed gene count	false discovery rate
GO.0006396	RNA processing	19	5.65E-05
GO.0016071	mRNA metabolic process	17	5.65E-05
GO.0008380	RNA splicing	13	0.000137
GO.0000398	mRNA splicing, via spliceosome	10	0.000356
GO.0016032	viral process	17	0.000356
GO.0019083	viral transcription	8	0.000356
GO.0044260	cellular macromolecule metabolic process	60	0.000356
GO.1901360	organic cyclic compound metabolic process	49	0.000356
GO.0090304	nucleic acid metabolic process	43	0.000365
GO.0019080	viral gene expression	8	0.000411

Molecular Function

#GO ID	GO description	observed gene count	false discovery rate
GO.0044822	poly(A) RNA binding	34	5.32E-13
GO.0003723	RNA binding	36	2.27E-11
GO.0003676	nucleic acid binding	44	5.99E-05
GO.0017166	vinculin binding	3	0.0171
GO.1901363	heterocyclic compound binding	49	0.0171
GO.0097159	organic cyclic compound binding	49	0.0205
GO.0000166	nucleotide binding	27	0.03
GO.0036002	pre-mRNA binding	3	0.03
GO.0036094	small molecule binding	29	0.03
GO.0005220	inositol 1,4,5-trisphosphate-sensitive calcium-release channel activity	2	0.0344

Table 7.2 Overview of the most significantly enriched gene ontology terms for the 117 potential direct targets of miR-7 during early stage culture of CHO DP12 cells

7.3.2.2 Late-stage culture analysis

Depletion of miR-7 in CHO DP12 cells results in a 65% increase in VCD and 3-fold increase in product yield on day 8 of culture (Figure 7.1). This time-point captures the greatest impact of miR-7 depletion on the phenotype of CHO DP12 cells. To gather a complete biological understanding of miR-7's actions, assorted sample preparations were conducted for day 8 culture cells. Proteomic profiling using subcellular enrichment and total proteome analysis of miR-7-spg and NC cells provided 3 distinct proteomic fractions for investigation. The membrane fraction identified 107 differentially expressed proteins with 61 showing the highest abundance in miR-7-spg compared to NC. The cytosolic fraction contained 230 differentially expressed proteins with 74 having an increased expression in miR-7-spg. Lastly, total proteomics identified 45 differentially expressed proteins with 34 having the highest abundance in miR-7-spg. The top 10 differentially expressed proteins with an increased abundance in miR-7-spg compared to NC for each fraction of day 8 are shown in Figure 7.3A. The overlap of all proteins with highest expression in miR-7-spg is shown as a Venn diagram in Figure 7.3B. Not a single protein was identified in all three fractions and only 8 proteins are present in any 2 fractions, as a result 160 unique proteins have been identified as potential direct targets of miR-7 during late-stage cell culture. This result demonstrates the benefit of using multiple proteomic sample preparation techniques to attain a comprehensive biological insight into (i) the mechanism of action of miRNAs and (ii) biopharmaceutical-relevant phenotypes. The full list of differentially expressed proteins is found in Supplementary Table S2.

A

	Accession	Peptides	Anova (p)	Fold change	Description
Cytosolic	EGV98552.1	2	0.01363922	36.50	Protein C10
	EGV96242.1	3	0.04430692	10.24	Eukaryotic peptide chain release factor GTP-binding subunit ERF3B
	EGW13651.1	2	0.02753857	8.02	TBC1 domain family member 20
	EGW00411.1	2	0.034263	3.47	14-3-3 protein eta
	EGV95270.1	2	0.00135996	2.99	Heat shock 70 kDa protein 1A
	EGW00467.1	3	0.02500824	2.90	Coatomer subunit beta'
	EGW05086.1	3	0.04330913	2.90	Nucleophosmin
	EGW09476.1	3	0.02796712	2.88	Cellular nucleic acid-binding protein
	EGW04132.1	2	0.03695607	2.75	AP-1 complex subunit gamma-1
	EGV92125.1	2	0.01981574	2.70	cAMP-dependent protein kinase type II-alpha regulatory subunit
Membrane	EGV93317.1	2	0.01122702	253.71	Small subunit processome component 20-like
	EGW05858.1	2	0.00126103	49.75	Delta-1-pyrroline-5-carboxylate dehydrogenase, mitochondrial
	EGW03989.1	2	0.0177751	42.13	U1 small nuclear ribonucleoprotein 70 kDa
	XP_007610270.1	3	0.00055909	23.83	PREDICTED: tripeptidyl-peptidase 2 isoform X6
	EGW00596.1	2	0.01664015	21.74	Transportin-2
	EGV95912.1	2	0.00021786	16.77	Insulin-like growth factor 2 mRNA-binding protein 2
	EGV97834.1	2	0.00207543	15.82	Cullin-2
	EGW08326.1	2	0.00971507	15.07	Ribosome biogenesis protein BMS1-like
	EGW13771.1	3	0.01041966	14.36	GDP-L-fucose synthetase
	EGW08037.1	2	0.01011474	14.34	Ras-related protein Rab-5C
Total	EGV94774.1	2	0.00607451	285.82	Histone H3
	EGW04525.1	2	0.04876883	24.59	Pyruvate kinase isozymes M1/M2
	XP_007649989.1	2	0.04414292	24.00	PREDICTED: ubiquitin thioesterase OTUB1 isoform X1
	EGW03135.1	3	0.00528372	20.65	Elongation factor 2
	XP_003507020.2	2	0.00557335	12.59	PREDICTED: thioredoxin isoform X1
	EGV93853.1	2	0.01180336	9.88	60S ribosomal protein L9
	EGV98163.1	2	0.01050747	9.42	ADP-ribosylation factor 5
	AAA37026.1	3	0.03262693	8.79	alpha-tubulin III
	EGV91661.1	2	0.02127163	7.06	Ubiquitin carboxyl-terminal hydrolase 14
	XP_003512064.2	2	0.04430572	6.31	PREDICTED: alpha-enolase isoform X2

B

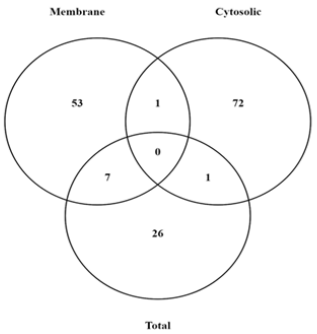


Figure 7.3 Differentially expressed proteins during late-stage culture. A. Top 10 differentially expressed proteins with highest abundance in miR-7-spg cells compared to NC cells for cytosolic, membrane and total proteome fractions on day 8 of culture. B. Venn diagram showing the overlap of miR-7-spg potential target proteins from the three fractions.

The 160 unique differentially expressed late-stage culture proteins were classified according to their biological processes and molecular functions as determined by GO analysis within the STRING software (table 7.3). The most enriched biological processes include intracellular protein transport, mRNA splicing, mRNA metabolic process and macromolecule complex subunit organisation which all have an FDR <0.00001 and over 10 proteins associated with each process. The most enriched molecular functions include poly(A) RNA binding, RNA binding, nucleic acid binding and nucleotide binding all with an FDR <1e-9 and over 50 proteins from the network included in the annotations.

Table 7.3 Overview of the most significantly enriched gene ontology terms for the 160 potential direct targets of miR-7 during late stage culture of CHO DP12 cells.

Biological Process

#GO ID	GO description	observed gene count	false discovery rate
GO.0006886	intracellular protein transport	23	1.89E-05
GO.0043933	macromolecular complex subunit organization	41	1.89E-05
GO.0000398	mRNA splicing, via spliceosome	13	2.55E-05
GO.0016071	mRNA metabolic process	20	2.55E-05
GO.0008380	RNA splicing	15	6.77E-05
GO.0016032	viral process	21	9.18E-05
GO.0006396	RNA processing	20	0.000237
GO.0034641	cellular nitrogen compound metabolic process	65	0.000237
GO.0044265	cellular macromolecule catabolic process	21	0.000237
GO.0051649	establishment of localization in cell	33	0.000237

Molecular Function

#GO ID	GO description	observed gene count	false discovery rate
GO.0044822	poly(A) RNA binding	58	3.95E-29
GO.0003723	RNA binding	58	3.61E-23
GO.0003676	nucleic acid binding	66	5.60E-10
GO.0000166	nucleotide binding	50	1.35E-09
GO.0036094	small molecule binding	52	4.71E-09
GO.1901363	heterocyclic compound binding	75	6.78E-07
GO.0097159	organic cyclic compound binding	75	1.11E-06
GO.0001882	nucleoside binding	37	8.06E-06
GO.0032555	purine ribonucleotide binding	37	9.20E-06
GO.0032550	purine ribonucleoside binding	36	1.29E-05

7.3.2.3 Comparison of early-stage and late-stage culture miR-7 candidate targets

Potential miR-7 targets were determined using state-of-the-art high resolution mass spectrometry to identify proteins significantly differentially expressed showing a minimum of 1.25-fold change in abundance levels. Early-stage culture identified a total of 117 potential miR-7 targets while late-stage culture analysis identified 160 potential targets of miR-7. In order to more specifically determine miR-7 targets driving the improved cellular phenotypes, a comparison of early and late-stage culture was performed. The overlap between both time points is visualised as a Venn diagram in figure 7.4A. A total of 19 proteins were commonly upregulated in CHO DP12 cells

following miR-7 stable depletion. These are: DBN1, RANBP2, HIST1H3A, MTA1, UBE2C, SMS, PRPF8, KPNA2, IPO4, HPCAL1, HNRNPA2B1, DLGAP5, DHFR, CYCS, BAG3, CAST, CTSZ, CUL2 and CEP170. Because of the problem of a lack of commercially available CHO-specific antibodies (Henry et al., 2017; Lin et al., 2015) we were unable to validate the expression of the differentially expressed proteins by Western blotting.

To further determine the biological role of these proteins, KEGG pathway analysis was performed within the DAVID software; however, no statistically significant pathways were identified for this set of 19 proteins. Individual KEGG pathway analysis of the early-stage and late-stage candidate miR-7 targets did however; reveal dysregulated pathways that were common to both stages of culture (Figure 7.4B and 7.4C). These pathways were splicesome, RNA transport and glutathione metabolism. Supplementary Table S3 shows the identified differentially expressed proteins from the early and late stage cultures that were significantly associated with a KEGG pathway.

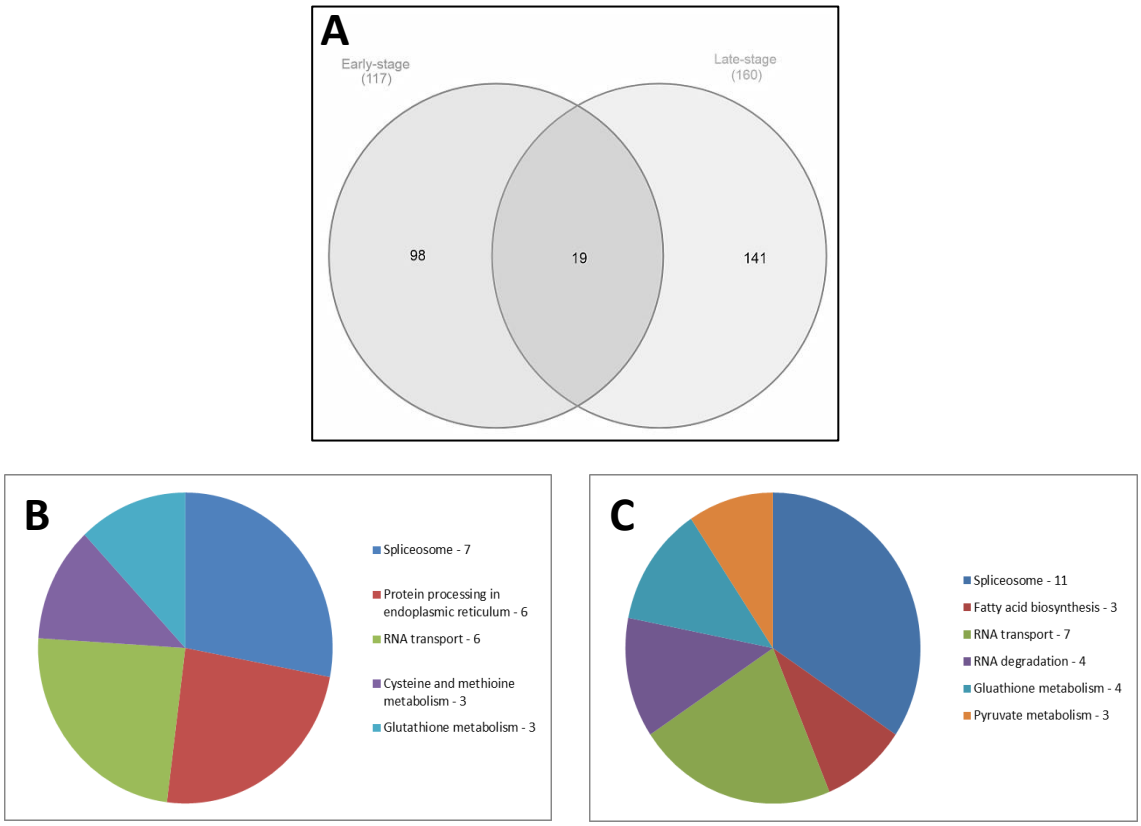


Figure 7.4. Comparative analysis of early and late-stage potential miR-7 targets. (A) Venn diagram showing the overlap between proteins upregulated in miR-7-spg DP12 cells in early stage versus late-stage culture. Nineteen proteins are consistently upregulated throughout the cell culture batch. (B) KEGG pathways analysis of the 117 early-stage culture miR-7 candidate targets. Five significant pathways were identified and the corresponding number of proteins identified from each pathway is listed beside the pathway description. (C) KEGG pathway analysis of the 160 late-stage culture miR-7 candidate targets shows 6 significant pathways were found in this cohort. The pathway description and number of proteins identified within each pathway is listed.

7.3.3 Identification of miR-7 targets using in-silico prediction

miRNAs are known to regulate mRNA expression through complimentary sequence binding of the miRNA seed region to target mRNAs 3' untranslated region (Hammond 2015). In silico tools have been established to predict miRNA targets based on sequence complementarity. Unfortunately, due to the lack of publicly available sequence information and annotation surrounding CHO cells such prediction tools do not house the CHO genome for miRNA: mRNA target analyses. To overcome this challenge in silico prediction of miR-7-5p targets was carried out using the human species. Fortunately, miRNA evolutionary conservation is widely acknowledged (Hackl et al. 2011; Lee, Risom, and Strauss 2007; Friedman et al. 2009) and miRviewer analysis of miR-7 verified it to be broadly conserved in 40/50 annotated animal genomes included in their database (data not shown) (Kiezun et al., 2012).

TargetScan prediction for miR-7-5p identified 558 transcripts with conserved sites in human (Supplementary table S4). Of the 117 proteins overexpressed in miR-7-spg in early stage of culture 3 proteins, CALU, CNN3 and STRN3 were identified as predicted targets of miR-7-5p. Interestingly these proteins have been reported as binding partners to each other in a calcium-dependant manner. Moreover, STRN3 also known as SG2NA, has been reported to be highly expressed during the S to G2 phase of the cell cycle and also involved in transcription activation. STRN3 and CNN3 bind Calmodulin which is involved in cell cycle progression. Calmodulin was previously found to be increased in expression following overexpression of miR-7 in CHO K1 cells (Meleady et al., 2012a) and although it was not identified here as significantly decreased following miR-7 depletion the identification of its binding partners still suggests that it plays a role in the CHO cell cycle. Taken together, the 3 predicted target proteins may be acting in combination to initiate growth improvement during early-stage culture. miR-7 binding to these genes could be limiting cell cycle progression in batch culture of CHO DP12 cells.

Of the 160 overexpressed proteins in miR-7-spg on day 8 of culture, 7 were predicted targets of miR-7-5p. Those proteins are BMS1, CUL5, GLG1, MAP1B, OSBPL11, PLEC and SRGAP2. Downregulation of miR-7 has previously been reported to

upregulate CUL5 in hepatocellular carcinoma cells and in turn facilitate G1/S transition perhaps reflecting a similar mode for miR-7 in this study (Ma et al., 2013).

7.4 Discussion

The appeal of genetically engineering CHO cells by miRNA manipulation is the multi-targeting nature of such a small non-coding RNA molecule; however, this feat could simultaneously be its own hindrance. Reports have suggested miRNA could bind hundreds of mRNA transcripts, though this estimation is dependent on the prediction model employed. Regardless, the small ~22 nucleotide make-up of miRNA infers the possibility to bind multiple mRNA targets through sequence complementarity. Caution must be exercised to avoid detrimental bioprocess effects caused by undesirable genes or pathways; therefore, identifying miRNA target genes to understand their impact on bioprocessing phenotypes is of upmost importance if they are to be employed as engineering targets to improve efficiency of production. A complete understanding of miRNAs with regards to their mode of action, all possible mRNA targets and reliability is essential if their introduction into biopharmaceutical production is sought.

Since the first report of miRNA expression in CHO cells in 2007 (Gammell et al., 2007), numerous studies have been performed to identify and manipulate beneficial miRNAs in CHO cell cultures; however limited attempts have been made to identify the functioning miRNA targets. MiR-143 was found to improve protein production in CHO-SEAP cells without negatively impacting on cell growth and viability (Fischer et al., 2014, p. 30). A follow-up study revealed MAPK7 among other genes as a validated target of miR-143 in human and rat and led to its analysis in CHO. mRNA and protein levels of MAPK7 were shown to decrease throughout the culture period upon miR-143 overexpression, furthermore siRNA knockdown of MAPK7 increases cell-specific productivity in various SEAP and mAb-producing CHO cell lines (Schoellhorn et al., 2017). Although MAPK7 was determined to be a target of miR-143 and found to independently drive productivity, miRNA target identification by surveying evolutionary conserved miRNA binding sites is of limited usefulness in determining miRNA gene targets specific to CHO. Proteomic analysis subsequent to miRNA manipulation is reported as a key

technique for miRNA target identification to unravel how miRNAs function (Huang, Pinto, and Pandey 2013; Thomas, Lieberman, and Lal 2010; Li et al. 2012). A study describing integrated miRNA analysis of CHO cells using miRNA, mRNA and proteomic profiling simultaneously is a key report which establishes a standard for comprehensive miRNA analysis (Clarke et al. 2012). Using LC-MS/MS profiling 285 proteins were found to be differentially expressed between CHO cell line clones with fast and slow growth rate phenotypes, and these proteins were found to be associated with translation, ribosome biogenesis and energy metabolism. Interestingly 25% of the proteins identified as DE were computationally predicted to be targeted by anti-correlated miRNAs identified in the study. This suggests the remaining 75% of proteins may be targeted by miRNAs but not through seed region complementarity or perhaps they are indirect, off-targets which still warrant further analysis. This study signifies the importance of analysing dysregulated proteins in response to miRNA alterations in order to gather a comprehensive insight into the functioning molecules causing phenotypic changes in CHO cell cultures.

MiR-7 has been previously characterised by our group to be involved in CHO cell growth and productivity, and more recently its reduction in expression levels in CHO-K1 cells using a sponge decoy technique displayed 40% improvement in maximum cell density and a 2-fold increase in SEAP production which are phenotypes desirable to the biopharmaceutical industry (Sanchez et al., 2014b). The study presented here employs the same sponge decoy technique but in the mAb producing CHO DP12 cell line to sequester miR-7 and improve CHO cell phenotype. Cell density and productivity was found to increase by 65% and 3-fold respectively on day 8 compared to NC cells (figure 7.1). The improved phenotype reflects previous data and signifies an important role for miR-7 in CHO cells which could be leveraged for biotherapeutic production. To fully understand the knock-on effect and mechanism of miR-7 we set out to identify and quantify potential targets whose levels are significantly altered following miR-7 depletion. Quantitative label-free LC-MS/MS analysis on day 3 and 8 of culture was performed on an Orbitrap Fusion Tribrid mass spectrometer resulting in over 3000 proteins identified per sample analysed (Figure 7.2). The technological advancements offered by this mass spectrometer, notably the increased acquisition rate, achieved by a

novel design and parallel analysis in the linear trap and Orbitrap have resulted in substantial improvements in proteome characterisation (Senko et al., 2013). This analysis thus represents a novel, in-depth proteomic analysis of the impact of miR-7 on CHO DP12 cells.

Early-stage, exponential phase of batch culture was assessed by collecting samples from NC-spg and miR-7-spg cells on day 3 and performing subcellular enrichment to yield two distinct proteomic fractions; cytosolic and membrane enriched. Of the 117 proteins differentially expressed interesting candidates include striatin-3 (SG2NA), ezrin, and various ribosomal associated proteins. SG2NA expression was over 2-fold increased in miR-7-spg cells. In silico analysis of miR-7 target genes identified SG2NA as a possible direct target of miR-7 based on seed region sequence complementarity to the 3' UTR of the gene. This protein has been shown to have increased expression during S and G2 phase and minimal expression during G1 and M phases thus having cell cycle-specific expression (Landberg and Tan, 1994). The upregulation of SG2NA following miR-7 depletion suggests endogenously miR-7 could be decreasing DNA replication, a key event during the S phase of the cell cycle (Laskey et al., 1989). SG2NA has been reported to activate Akt through stabilizing DJ-1 causing enhanced proliferation in various cancer cells (Tanti et al., 2015). A marginal increase in cell density on day 3 in miR-7-spg cells is shown in this study potentially as a direct result of SG2NA upregulation in-turn enhancing the Akt pathway. Specific genetic engineering of SG2NA could expedite cell growth by targeting the Akt pathway in early stage culture of CHO cells; this hypothesis warrants future analysis. Another interesting potential target of miR-7 identified is Ezrin, identified here as 1.5-fold higher in expression in miR-7-spg cells. Ezrin is a plasma membrane protein which has also been found to interact with the Akt pathway and determine cell survival in epithelial cells (Gautreau et al., 1999). Again, this suggests miR-7 suppression of the Akt pathway could be occurring in CHO DP12 as depletion of miR-7 leads to upregulation of Akt pathway associated proteins such as SG2NA and Ezrin.

Proteomic profiling of late-stage culture CHO DP12 cells following miR-7 depletion was carried out using sub-cellular enrichment and total proteome analysis of day 8 cells.

A total of 160 unique differentially expressed proteins with an increased abundance in miR-7-spg cells compared with NC were identified. Proteins of interest identified as potential targets of miR-7 include TBC1D20 and BMS1. TBC1D20 was identified as 8-fold higher in miR-7-spg cell on day 8 of cell culture. Interestingly this protein was previously found decreased in an IgG-producing CHO cell line by a mitochondrial genome-encoded small RNA (mitosRNA) which function in a miRNA-like manner. Stable knockdown of TBC1D20 and an accompanying ER-localised protein also found to be mitosRNA target, dramatically increased productivity and cell growth through secretory pathway optimisation (Pieper et al., 2017). This protein is 8-fold upregulated upon miR-7 depletion in this study; however based on the existing data surrounding it, more specific knockdown may drive productivity and cell growth further in our CHO cell line. This finding requires future analysis to confirm this. Finally of most interest, BMS1 which functions in ribosome biogenesis was identified with a 15-fold increase in expression compared to NC in the membrane fraction of day 8 miR-7-spg cells. Unfortunately, due to a lack of commercially available antibodies against this protein validation of expression was not possible. Alteration of the BMS1 gene copy number has been shown to change the ribosomal subunit ratio in producing *S.cerevisiae* in turn maximising recombinant product yield by a factor of 78 compared to wild-type yeast (Bonander et al., 2009). BMS1 has not been reported in CHO cells previously; however this study presents novel proteins uncovered by state-of-the-art mass spectrometry which could represent new angles for improving biopharmaceutical production. Ribosomal engineering was first reported in CHO cells in 2009 by Santoro et al. whereby ribosomal RNA (rRNA) transcription rate was enhanced which led to increased productivity of recombinant proteins (Santoro et al., 2009). More recently, ribosome sequencing alongside RNA-Seq of an IgG-producing CHO cell line revealed recombinant mRNAs made up 20% of total mRNA content and sequestered 15% of all ribosomes engaged in translation with a similar translational efficiency to endogenous mRNA (Kallehauge et al., 2017). By eliminating the non-essential dominant mRNA species in the cell line, NeoR, cell growth and product titre increased; hypothesised to be a direct result of freeing up translational capacity in ribosomes. Ribosome profiling and engineering in CHO cells is a promising approach for productivity enhancement.

Identification of BMS1 in our study highlights the importance of ribosome biogenesis in bioproduction. BMS1 may be a target of miR-7 which when elevated could improve productivity by translational enhancement within ribosomes. This theory warrants future investigation of BMS1 specific engineering in CHO cells.

Comparative analysis of the early-stage and late-stage culture upregulated proteins following miR-7 stable depletion revealed 19 proteins to be consistently upregulated over the batch culture. Some of these proteins show a significantly high level of upregulation with protein abundance increasing by 2-fold or more. These consistently highly upregulated proteins are histone H3, dihydrofolate reductase and cullin-2. Histone H3 has previously been identified in CHO cells with unstable expression of recombinant antibody product during long-term culture accompanied by histone H3 hypoacetylation (Paredes et al., 2013). A subsequent study revealed that selecting CHO cell lines with high levels of acetylation of histone H3 can improve the stability of production cell lines (Moritz et al., 2016). Although acetylation modifications were not analysed in this proteomic study, the high abundance of histone H3 upregulation in this mAb producing CHO cell line correlates with other studies and suggests miR-7 depletion enables a more stable expression of the mAb product evidenced by the 3-fold increased product yield on day 8 of culture. Dihydrofolate reductase (DHFR) is also highly upregulated over 2-fold throughout DP12 cell culture and has been previously studied in CHO cells. The dihydrofolate reductase locus in CHO cells has been found to contain specific replication origin sites which CHO cells can select in order to initiate replication (Sasaki et al., 2006). Transcription through the DHFR gene is required to activate this replication origin site in early S phase CHO cells (Saha et al., 2004). The identification of upregulated levels of DHFR in our analysis thus could be correlated with the 65% increased cellular growth seen in CHO DP12 cells with miR-7 depletion on day 8 of culture. DHFR may serve as an indicator of CHO cell replication. Finally, cullin-2 which was also identified as being consistently upregulated has not been previously studied in CHO cells. However, in eukaryotes it functions as a scaffold protein to form E3 ubiquitin ligases which recognize a number of substrates and regulate their protein stability and function through poly-ubiquitination (Cai and Yang, 2016). The identification of histone H3 and DHFR as highly and consistently upregulated

proteins following miR-7 stable depletion suggests that they may serve as indicators but not drivers of stable mAb productivity and increased replication respectively, based on the previous work of other groups in CHO research described above. KEGG pathways analysis of the commonly dysregulated set of 19 proteins did not identify any significant pathways associated with this cohort. Individual KEGG pathways analysis did, however, identify 3 pathways, splicesome, RNA transport and glutathione metabolism, as significantly enriched for both early and late-stage culture potential targets of miR-7. The splicesome and RNA transport pathways are two key events for gene expression. Splicing of pre-mRNA to remove introns and join protein-coding exons is a fundamental process to produce mature mRNA which can be translated into proteins (van den Hoogenhof Maarten M.G. et al., 2016). Similarly, RNA transport from the nucleus to the cytoplasm of species such as tRNAs and rRNAs is critical for gene expression (Köhler and Hurt, 2007). These two significantly enriched KEGG pathways signifies that miR-7 depletion in the mAb producing CHO DP12 cell line results in an increase in protein synthesis machinery and processes. This finding emphasises the potential of miR-7 intervention to improve the productivity of CHO cell lines through protein synthesis upregulation.

In conclusion, we have identified a large cohort of proteins from subcellular enrichment and total proteome analysis to be differentially expressed following stable depletion of miR-7 using sponge decoy technology in CHO DP12 cells. Using different methods of sample preparation increased the proteomic coverage in this study and employing novel high-end mass spectrometry provided in-depth and accurate measurement of thousands of proteins for this label-free relative quantitation study. In-depth proteomic analysis of miR-7 depleted DP12 cells identified proteins not previously studied as phenotypic drivers in CHO; some of these proteins are involved in important cellular mechanisms such as survival, growth and ribosome biogenesis in other eukaryotes. Furthermore, proteomics has identified numerous potential targets of miR-7 which were not identified by in silico prediction software thus highlighting the importance of profiling miR-induced phenotypes. Predictive software has recently been scrutinised for biasing target prediction based solely on seed region binding, for containing a high degree of false positives and generally lacking specificity and sensitivity (Liu et al., 2010;

Loganantharaj and Randall, 2017; Pinzón et al., 2017). Novel methods of miR binding have been uncovered which are not yet taken into account in most of current prediction software. Conversely, mass spectrometry based proteomic studies identifies proteins using high-resolution technology to provide highly confident, accurately quantified identifications which provide an unbiased insight into the biology of CHO cell phenotypes (Angel et al., 2012).

Early stage of culture identified 117 unique proteins which are potential direct targets of miR-7. Upregulated proteins identified here play a role in cell cycle progression and some targets are specifically involved in the Akt pathway which promotes cell survival and growth. This finding suggests endogenous miR-7 may be suppressing CHO cell growth in the early stage of culture by binding to these targets of interest. Depletion of miR-7 shown here causes a marginal increase in cell growth compared to NC on day 3, however more specific target engineering may enhance this phenotype and push CHO cell cultures to higher densities faster by leveraging the Akt pathway. Late stage culture identified 160 unique proteins which are potentially direct targets of miR-7. From this group of proteins functional annotation revealed RNA binding and mRNA processes as enrichment functions connecting this cohort. Of most interest is the identification of proteins involved in CHO cell product secretion and ribosome biogenesis, both promising approaches for the improvement of the biopharmaceutical phenotype. Comparing these sets of early and late-stage miR-7 candidate targets, 19 proteins were consistently upregulated over culture time. Most interestingly, two of these proteins, histone H3 and DHFR, were previously studied and found to indicate stable mAb production and replication in CHO cells. Significantly enriched pathways include the splicesome and RNA transport pathways, two key events enabling protein synthesis. Identification of these phenotypic marker proteins and enriched pathways validates miR-7 depletion as a promising engineering approach to improve CHO cell growth and productivity. We have previously found that miR-7 depletion improves cell growth and productivity in CHO cells; however, this study presents the potential key regulators and underlying mechanisms of this phenotype. Targeted engineering of these proteins and pathways may enhance these phenotypes even more and such proteins represent the basis of future studies going forward.

7.5 References

Agarwal, V., Bell, G.W., Nam, J.-W., Bartel, D.P. (2015) Predicting effective microRNA target sites in mammalian mRNAs. elife 4, e05005.

Altamirano, C., Paredes, C., Illanes, A., Cairo, J.J., Godia, F., 2004. Strategies for fed-batch cultivation of t-PA producing CHO cells: substitution of glucose and glutamine and rational design of culture medium. *J. Biotechnol.* 110, 171–179.

Angel, T.E., Aryal, U.K., Hengel, S.M., Baker, E.S., Kelly, R.T., Robinson, E.W., Smith, R.D. (2012) Mass spectrometry based proteomics: existing capabilities and future directions. *Chem. Soc. Rev*. 41, 3912–3928. https://doi.org/10.1039/c2cs15331a

Barron, N., Kumar, N., Sanchez, N., Doolan, P., Clarke, C., Meleady, P., O'Sullivan, F., Clynes, M. (2011) Engineering CHO cell growth and recombinant protein productivity by overexpression of miR-7. *J. Biotechnol*. 151, 204–211.

Bollati-Fogolín, M., Forno, G., Nimtz, M., Conradt, H.S., Etcheverrigaray, M., Kratje, R. (2005) Temperature reduction in cultures of hGM-CSF-expressing CHO cells: effect on productivity and product quality. *Biotechnol. Prog.* 21, 17–21.

Bonander, N., Darby, R.A., Grgic, L., Bora, N., Wen, J., Brogna, S., Poyner, D.R., O'Neill, M.A., Bill, R.M. (2009) Altering the ribosomal subunit ratio in yeast maximizes recombinant protein yield. *Microb. Cell Factories* 8, 10.

Cai, W., Yang, H. (2016) The structure and regulation of Cullin 2 based E3 ubiquitin ligases and their biological functions. *Cell Div*. 11, 7.

Clarke, C., Henry, M., Doolan, P., Kelly, S., Aherne, S., Sanchez, N., Kelly, P., Kinsella, P., Breen, L., Madden, S.F. (2012) Integrated miRNA, mRNA and protein expression analysis reveals the role of post-transcriptional regulation in controlling CHO cell growth rate. *BMC Genomics* 13, 656.

Coleman, O., Henry, M., Clynes, M., Meleady, P. (2017) Filter-Aided Sample Preparation (FASP) for Improved Proteome Analysis of Recombinant Chinese Hamster Ovary Cells. *Methods Mol. Biol.* 1603, 187–194.

Estes, S., Melville, M. (2014) Mammalian Cell Line Developments in Speed and Efficiency, in: Zhou, W., Kantardjieff, A. (Eds.), Mammalian Cell Cultures for Biologics Manufacturing, Advances in Biochemical Engineering/Biotechnology. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 11–33.

Fischer, S., Buck, T., Wagner, A., Ehrhart, C., Giancaterino, J., Mang, S., Schad, M., Mathias, S., Aschrafi, A., Handrick, R., Otte, K. (2014) A functional high-content miRNA screen identifies miR-30 family to boost recombinant protein production in CHO cells. *Biotechnol. J.* 9, 1279–1292.

Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., Mering, C.V. (2012) STRING v9. 1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res*. 41, D808–D815.

Friedman, R.C., Farh, K.K.-H., Burge, C.B., Bartel, D.P. (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* 19, 92–105.

Gammell, P., Barron, N., Kumar, N., Clynes, M. (2007) Initial identification of low temperature and culture stage induction of miRNA expression in suspension CHO-K1 cells. *J. Biotechnol.* 130, 213–218.

Gautreau, A., Poullet, P., Louvard, D., Arpin, M. (1999) Ezrin, a plasma membrane-microfilament linker, signals cell survival through the phosphatidylinositol 3-kinase/Akt pathway. *Proc. Natl. Acad. Sci. U. S. A.* 96, 7300–7305.

Grimson, A., Farh, K.K.-H., Johnston, W.K., Garrett-Engele, P., Lim, L.P., Bartel, D.P., (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell* 27, 91–105.

Hackl, M., Borth, N., Grillari, J. (2012) miRNAs – pathway engineering of CHO cell factories that avoids translational burdening. *Trends Biotechnol*. 30, 405–406.

Hackl, M., Jakobi, T., Blom, J., Doppmeier, D., Brinkrolf, K., Szczepanowski, R., Bernhart, S.H., Siederdissen, C.H. zu, Bort, J.A.H., Wieser, M. (2011) Next-generation sequencing of the Chinese hamster ovary microRNA transcriptome: Identification, annotation and profiling of microRNAs as targets for cellular engineering. *J. Biotechnol*. 153, 62–75.

Hammond, S.M. (2015) An overview of microRNAs. *Adv. Drug Deliv. Rev*. 87, 3–14.

Henry, M., Power, M., Kaushik, P., Coleman, O., Clynes, M., Meleady, P. (2017) Differential Phosphoproteomic Analysis of Recombinant Chinese Hamster Ovary Cells Following Temperature Shift. *J. Proteome Res*. 16, 2339–2358.

Huang, D.W., Sherman, B.T., Tan, Q., Collins, J.R., Alvord, W.G., Roayaei, J., Stephens, R., Baseler, M.W., Lane, H.C., Lempicki, R.A. (2007) The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol*. 8, R183.

Huang, E.P., Marquis, C.P., Gray, P.P. (2007) Development of Super-CHO protein-free medium based on a statistical design. *J. Chem. Technol. Biotechnol. Int. Res. Process Environ. Clean Technol*. 82, 431–441.

Huang, E.P., Marquis, C.P., Gray, P.P. (2004) Process development for a recombinant Chinese hamster ovary (CHO) cell line utilizing a metal induced and amplified metallothionein expression system. *Biotechnol. Bioeng*. 88, 437–450.

Huang, T.-C., Pinto, S.M., Pandey, A. (2013) Proteomics for understanding miRNA biology. *Proteomics* 13, 558–567.

Kallehauge, T.B., Li, S., Pedersen, L.E., Ha, T.K., Ley, D., Andersen, M.R., Kildegaard, H.F., Lee, G.M., Lewis, N.E. (2017) Ribosome profiling-guided depletion of an mRNA increases cell growth rate and protein secretion. *Sci. Rep*. 7, 40388.

Kaushik, P., Henry, M., Clynes, M., Meleady, P. (2018) The Expression Pattern of the Phosphoproteome Is Significantly Changed During the Growth Phases of Recombinant CHO Cell Culture. *Biotechnol. J*. 13, e1700221.

Kiezun, A., Artzi, S., Modai, S., Volk, N., Isakov, O., Shomron, N. (2012) miRviewer: a multispecies microRNA homologous viewer. *BMC Res. Notes* 5, 92.

Kishishita, S., Katayama, S., Kodaira, K., Takagi, Y., Matsuda, H., Okamoto, H., Takuma, S., Hirashima, C., Aoyagi, H. (2015) Optimization of chemically defined feed media for monoclonal antibody production in Chinese hamster ovary cells. *J. Biosci. Bioeng.* 120, 78–84.

Köhler, A., Hurt, E. (2007) Exporting RNA from the nucleus to the cytoplasm. Nat. Rev. Mol. Cell Biol. 8, 761–773.

Landberg, G., Tan, E.M. (1994) Characterization of a DNA-binding nuclear autoantigen mainly associated with S phase and G2 cells. *Exp. Cell Res*. 212, 255–261.

Laskey, R.A., Fairman, M.P., Blow, J.J. (1989) S phase of the cell cycle. *Science* 246, 609–614.

Lee, C.-T., Risom, T., Strauss, W.M. (2007) Evolutionary conservation of microRNA regulatory circuits: an examination of microRNA gene complexity and conserved microRNA-target interactions through metazoan phylogeny. *DNA Cell Biol*. 26, 209–218.

Lewis, B.P., Burge, C.B., Bartel, D.P. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120, 15–20.

Li, C., Xiong, Q., Zhang, J., Ge, F., Bi, L.-J. (2012) Quantitative proteomic strategies for the identification of microRNA targets. *Expert Rev. Proteomics* 9, 549–559.

Lin, N., Mascarenhas, J., Sealover, N.R., George, H.J., Brooks, J., Kayser, K.J., Gau, B., Yasa, I., Azadi, P., Archer-Hartmann, S. (2015) Chinese hamster ovary (CHO) host cell engineering to increase sialylation of recombinant therapeutic proteins by modulating sialyltransferase expression. *Biotechnol. Prog*. 31, 334–346.

Liu, H., Yue, D., Chen, Y., Gao, S.-J., Huang, Y. (2010) Improving performance of mammalian microRNA target prediction. *BMC Bioinformatics* 11, 476.

Loganantharaj, R., Randall, T.A. (2017) The Limitations of Existing Approaches in Improving MicroRNA Target Prediction Accuracy. *Methods Mol. Biol*. 1617, 133–158.

Ma, C., Qi, Y., Shao, L., Liu, M., Li, X., Tang, H. (2013) Downregulation of miR-7 upregulates Cullin 5 (CUL5) to facilitate G1/S transition in human hepatocellular carcinoma cells. *IUBMB Life* 65, 1026–1034.

Meleady, P., Gallagher, M., Clarke, C., Henry, M., Sanchez, N., Barron, N., Clynes, M. (2012) Impact of miR-7 over-expression on the proteome of Chinese hamster ovary cells. *J. Biotechnol*. 160, 251–262.

Mering, C. von, Huynen, M., Jaeggi, D., Schmidt, S., Bork, P., Snel, B. (2003) STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res*. 31, 258–261.

Moritz, B., Woltering, L., Becker, P.B., Göpfert, U. (2016) High levels of histone H3 acetylation at the CMV promoter are predictive of stable expression in Chinese hamster ovary cells. *Biotechnol. Prog*. 32, 776–786. https://doi.org/10.1002/btpr.2271

Paredes, V., Park, J.S., Jeong, Y., Yoon, J., Baek, K. (2013) Unstable expression of recombinant antibody during long-term culture of CHO cells is accompanied by histone H3 hypoacetylation. *Biotechnol. Letters*. 35, 987–993. https://doi.org/10.1007/s10529-013-1168-8

Pieper, L.A., Strotbek, M., Wenger, T., Gamer, M., Olayioye, M.A., Hausser, A. (2017) Secretory pathway optimization of CHO producer cells by co-engineering of the mitosRNA-1978 target genes CerS2 and Tbc1D20. *Metab. Eng*. 40, 69–79.

Pinzón, N., Li, B., Martinez, L., Sergeeva, A., Presumey, J., Apparailly, F., Seitz, H. (2017) microRNA target prediction programs predict many false positives. *Genome Res*. 27, 234–245.

Saha, S., Shan, Y., Mesner, L.D., Hamlin, J.L. (2004) The promoter of the Chinese hamster ovary dihydrofolate reductase gene regulates the activity of the local origin and helps define its boundaries. *Genes Dev.* 18, 397–410.

Sanchez, N., Kelly, P., Gallagher, C., Lao, N.T., Clarke, C., Clynes, M., Barron, N. (2014) CHO cell culture longevity and recombinant protein yield are enhanced by depletion of miR-7 activity via sponge decoy vectors. *Biotechnol. J.* 9, 396–404.

Santoro, R., Lienemann, P., Fussenegger, M. (2009) Epigenetic engineering of ribosomal RNA genes enhances protein production. *PLoS One* 4, e6653.

Sasaki, T., Ramanathan, S., Okuno, Y., Kumagai, C., Shaikh, S.S., Gilbert, D.M. (2006) The Chinese Hamster Dihydrofolate Reductase Replication Origin Decision Point Follows Activation of Transcription and Suppresses Initiation of Replication within Transcription Units. *Mol. Cell. Biol.* 26, 1051–1062.

Schoellhorn, M., Fischer, S., Wagner, A., Handrick, R., Otte, K. (2017) miR-143 targets MAPK7 in CHO cells and induces a hyperproductive phenotype to enhance production of difficult-to-express proteins. *Biotechnol. Prog.* 33, 1046–1058.

Senko, M.W., Remes, P.M., Canterbury, J.D., Mathur, R., Song, Q., Eliuk, S.M., Mullen, C., Earley, L., Hardman, M., Blethrow, J.D. (2013) Novel parallelized quadrupole/linear ion trap/Orbitrap tribrid mass spectrometer improving proteome coverage and peptide identification rates. *Anal. Chem*. 85, 11710–11714.

Shin, C., Nam, J.-W., Farh, K.K.-H., Chiang, H.R., Shkumatava, A., Bartel, D.P. (2010) Expanding the microRNA targeting code: functional sites with centered pairing. *Mol. Cell* 38, 789–802.

Stolfa, G., Smonskey, M.T., Boniface, R., Hachmann, A.-B., Gulde, P., Joshi, A.D., Pierce, A.P., Jacobia, S.J., Campbell, A. (2018) CHO-Omics Review: The Impact of Current and Emerging Technologies on Chinese Hamster Ovary Based Bioproduction. *Biotechnol. J*. 13, 1700227.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P. (2014) STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 43, D447–D452.

Tanti, G.K., Pandey, S., Goswami, S.K. (2015) SG2NA enhances cancer cell survival by stabilizing DJ-1 and thus activating Akt. *Biochem. Biophys. Res. Commun*. 463, 524–531.

Thomas, M., Lieberman, J., Lal, A. (2010) Desperately seeking microRNA targets. *Nat. Struct. Mol. Biol.* 17, 1169.

Trummer, E., Fauland, K., Seidinger, S., Schriebl, K., Lattenmayer, C., Kunert, R., Vorauer-Uhl, K., Weik, R., Borth, N., Katinger, H. (2006) Process parameter shifting: Part I. Effect of DOT, pH, and temperature on the performance of Epo-Fc expressing CHO cells cultivated in controlled batch bioreactors. *Biotechnol. Bioeng*. 94, 1033–1044.

van den Hoogenhof Maarten M.G., Pinto Yigal M., Creemers Esther E. (2016) RNA Splicing. *Circ. Res*. 118, 454–468.

Vizcaíno, J.A., Csordas, A., del-Toro, N., Dianes, J.A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., Xu, Q.-W., Wang, R., Hermjakob, H. (2016) 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res*. 44, D447-456.

Wiśniewski, J.R., Zougman, A., Nagaraj, N., Mann, M. (2009) Universal sample preparation method for proteome analysis. *Nat. Methods* 6, 359.

7.6 Appendix D

Supplementary tables S1 – S4 are provided on the disk attached to the back cover of the thesis.

8 Future work

Prospective studies from published research

During my PhD project I have published 4 research articles, 1 methodology book chapter and 1 data article all stemming from my personal PhD work. These publications provide in-depth proteomic profiling of both pancreatic cancer and CHO cells; possible future work important for both these studies, is outlined below.

Pancreatic cancer studies

Chapter 3: A comparative quantitative LC-MS/MS profiling analysis of human pancreatic adenocarcinoma, adjacent-normal tissue, and patient-derived tumour xenografts is the first comparative proteomic analysis of PDAC which employs PDX models to identify patient tumour cell-associated proteins, in an effort to find robust targets for therapeutic treatment of PDAC. This study provides a wealth of proteomic data – both qualitative lists of all identified proteins for each sample type, and quantitative information on differentially expressed proteins between various sample types.

- Prioritisation and selection of target proteins for validation. The 129 proteins significantly overexpressed in PDAC compared to adjacent-normal tissues, and 32 human-specific proteins overexpressed in PDX F1 compared to primary PDAC are too numerous for validation. Prioritisation of the most promising candidate proteins will largely be based on fold-change, consistent expression across the current patient cohort and novelty. The 32 human-specific proteins overexpressed in PDX F1 compared to primary tumours are the highest priority for future studies. This analysis showd that these proteins are derived from the tumour cells and thus may be key drivers of PDAC tumourigenesis. Overlapping targets that are overexpressed in primary tumours compared to adjacent-normal AND overexpressed in PDX F1 compared to primary tumours may prove to be of particular interest.

- Validation of target proteins using immunohistochemical (IHC) staining. Western blot validation of CD55 expression was performed for this study; however, this form of validation for patient samples and PDX tissues proved very difficult due to the heterogeneity of the samples. Furthermore it does not distinguish between expression in different cell types. IHC validation of potential targets will give a more accurate measurement of expression levels and provide cellular location information for the targets of interest. IHC should be performed on a larger, separate cohort of PDAC patients and non-diseased pancreas tissues to validate the specificity of the target proteins. Similarly, IHC of the targets across other tissue samples from other organs will demonstrate the usefulness as a target for novel targeted therapeutics of PDAC.
- In conjunction with validating the expression of candidate proteins, *in vitro* analysis of the role of targets proteins in PDAC through cell line studies can be performed. Functional studies can be performed following knock-down or overexpression of a target to determine the biological role of the target in PDAC progression. Examples of appropriate functional analyses include assays for proliferation, migration and invasion.
- The main aim of this study is to identify novel targets for targeted therapeutic treatments of PDAC. However using the data analysis method described in this study to identify human-specific proteins a prospective study to identify human-specific proteins in the blood of PDX mice could identify potential biomarkers of PDAC. The use of human-specific protein identifications in PDX proteomic analyses signifies that these proteins are derived from the primary PDAC tumour and selectively de-contaminates the protein data resulting in a brief list of candidate targets.

CHO studies

Chapter 5: Depletion of endogenous miRNA-378-3p increases peak cell density of CHO DP12 cells and is correlated with elevated levels of ubiquitin carboxyl-terminal hydrolase 14, and the associated data article chapter 6, provide an in-depth proteomic profile of the effects of miR-378 depletion on CHO DP12 cells.

- Depletion of miR-378 significantly improved peak cell density of CHO DP12 cells by 59%. Investigation of miR-378 depletion in other CHO cell lines would be of interest to see if it has the same growth impact.
- Proteomic analysis identified a number of potential targets of miR-378. Although Usp14 follow-up did identify a role for this protein in CHO cell growth the impact was not as significant as miR-378 depletion alone. Analyses of other identified potential targets of miR-378 could determine a clearer mechanism of action for the increased cellular growth.
- Knock-down of miR-378 was achieved using sponge-decoy technology to increase CHO cell growth. Complete knock-out of miR-378 using CRISPR technology may confer a more significant effect on CHO cell growth and is worth investigating.

Chapter 7: Increased growth rate and productivity following stable depletion of miR-7 in a mAb producing CHO cell line causes an increase in proteins associated with the Akt pathway and ribosome biogenesis. This study provides an in-depth proteomic profile of the effects of miR-7 depletion on CHO DP12 cells.

- Depletion of miR-7 in CHO DP12 caused a 65% increase in cell growth and >3-fold increase in yield of secreted IgG protein. Investigation of miR-7 depletion in other industrial CHO cell lines would be interesting as it has significant advantageous phenotypic effects.
- A number of proteins such as SG2NA and ezrin were identified in early-stage culture as potential targets of miR-7 and both targets have reported roles in the Akt pathway. Specific targeting of the Akt pathway through these targets or others would be of interest to drive CHO cell growth more efficiently and directly.

- TBC1D20 was identified as 8-fold increased in abundance following miR-7 depletion during late-stage culture. Our results suggest that overexpression of TBC1D20 may be correlated with improved CHO cell phenotypes; directed engineering of this target may determine its exact biological role. BMS1 was identified as 15-fold higher in abundance upon miR-7 depletion during late-stage culture proteomic analysis. This protein has not been reported in CHO cells previously, however in *S.cerevisae* alteration of BMS1 levels changed the ribosomal subunit ratio and increased productivity by 78-fold. Ribosome biogenesis and engineering to increase translational efficiency in CHO cells is an emerging field of study. Directed engineering of BMS1 in CHO DP12 cells may increase productivity to similar levels seen in *S.cerevisae.*
- Complete knock-out of miR-7 using CRISPR technology may confer a more significant effect on CHO cell growth and productivity and is worth investigating.

9 Conclusions

The advancements in proteomics, has been inspired by the realisation that the final product of a gene is inherently more complex and responsible for the phenotypes of cells rather than the gene itself. Over the last number of decades proteomics and mass spectrometry has significantly advanced resulting in huge increases in proteome coverage and understanding. Current techniques can identify over 5000 proteins per sample in a 2-hour run. This wealth of data has been used to understand diverse biological materials and most importantly human diseases. The advent of accurate proteomic profiling studies has led to major successes in for example cancer subtyping based on certain proteins expression with treatment strategies specific to sub-types. Disease profiles are now readily exploited to create targeted therapeutics such as the treatment of HER2 positive breast cancer patients with trastuzamab. Mass spectrometry however remains an incomplete technology for the identification of the complete proteome. Future developments in sample preparation, mass spectrometers and bioinformatics are still necessary to generate a complete proteomic profile of samples in a timely, cost-efficient manner.

This thesis presents a proteomic analysis of PDAC using primary tumours, adjacent-normal tissues, PDX F1 and F2 tumours and PDAC cell lines. The analyses presented here provide a proteomic insight into the development and progression of PDAC. Membrane protein enrichment was employed with the aim of identifying novel drug targets for a targeted therapeutic for PDAC, something that is gravely missing. Using the PDX models and strict data processing techniques, I published the first comparative proteomic analysis of PDAC which exploits PDX PDAC tumours to identify patient tumour cell-associated proteins. Analysis of the two PDX generations on the proteomic level also revealed the stability of the PDAC tumour proteome over time thus proving the dependability of these models for future studies of PDAC. Proteomic analysis of the standard available PDAC cell lines revealed the reliability of these models in proteomic studies with the majority of the primary PDAC tumour proteome recapitulated in the cell line models. The PDAC work carried out in this thesis provides a wealth of

proteomic data that can be manipulated by the wider research community in future studies for biomarker evaluation and novel drug target identification.

The CHO cell factory is the most commonly used mammalian cell line for biopharmaceutical production and was the basis for the second project presented in this thesis. With the advent of a deeper understanding of many human diseases using techniques such as mass spectrometry, the need to treat these diseases by targeted therapeutics is similarly growing. MAbs are among the top selling drugs in the world and the requirement of human-like structures which contain glycosylation profiles necessitates the use of CHO cells. The ever-increasing demand for mAbs pushes the biopharmaceutical industry to reach maximum production levels. Efforts for maximising product yield have focused on engineering the CHO cell of late. This thesis focuses on microRNA engineering to improve the CHO cell phenotype. Depletion of miR-378 and miR-7 in CHO DP12 cells using sponge decoy technology resulted in improved phenotypes for both studies. Proteomic analyses were performed to identify potential targets of the respective miRs and to understand the biological drivers of CHO cell phenotypes. MiR-378 analysis revealed a role for Usp14 in increasing CHO cell growth. MiR-7 proteomic analysis identified proteins during early-stage and late-stage of culture which are known to be involved in the Akt cell proliferation pathways and ribosome biogenesis, respectively. These proteomic studies helped to define the functional protein drivers of the improved CHO cell phenotypes. Directed targeting of these proteins and pathways may further augment CHO cell growth and productivity to meet the needs of the biopharmaceutical industry.

In conclusion, proteomic analysis is an important field for understanding the phenotypes of cells. This thesis has applied proteomics and mass spectrometry to understand two diverse cell phenotypes; pancreatic cancer and biotherapeutic producing CHO cells. In-depth proteomic analyses were achieved by subcellular enrichment as a fractionation method during sample preparation prior to MS and by robust data processing. Understanding both these cell types on a proteomic level has massive potential for the targeted treatment of PDAC and maximisation of CHO cell productivity.