Landscaping the cell surface proteome of breast cancer

Following pathways through organelles to the plasma membrane

Kurbasic, Emila

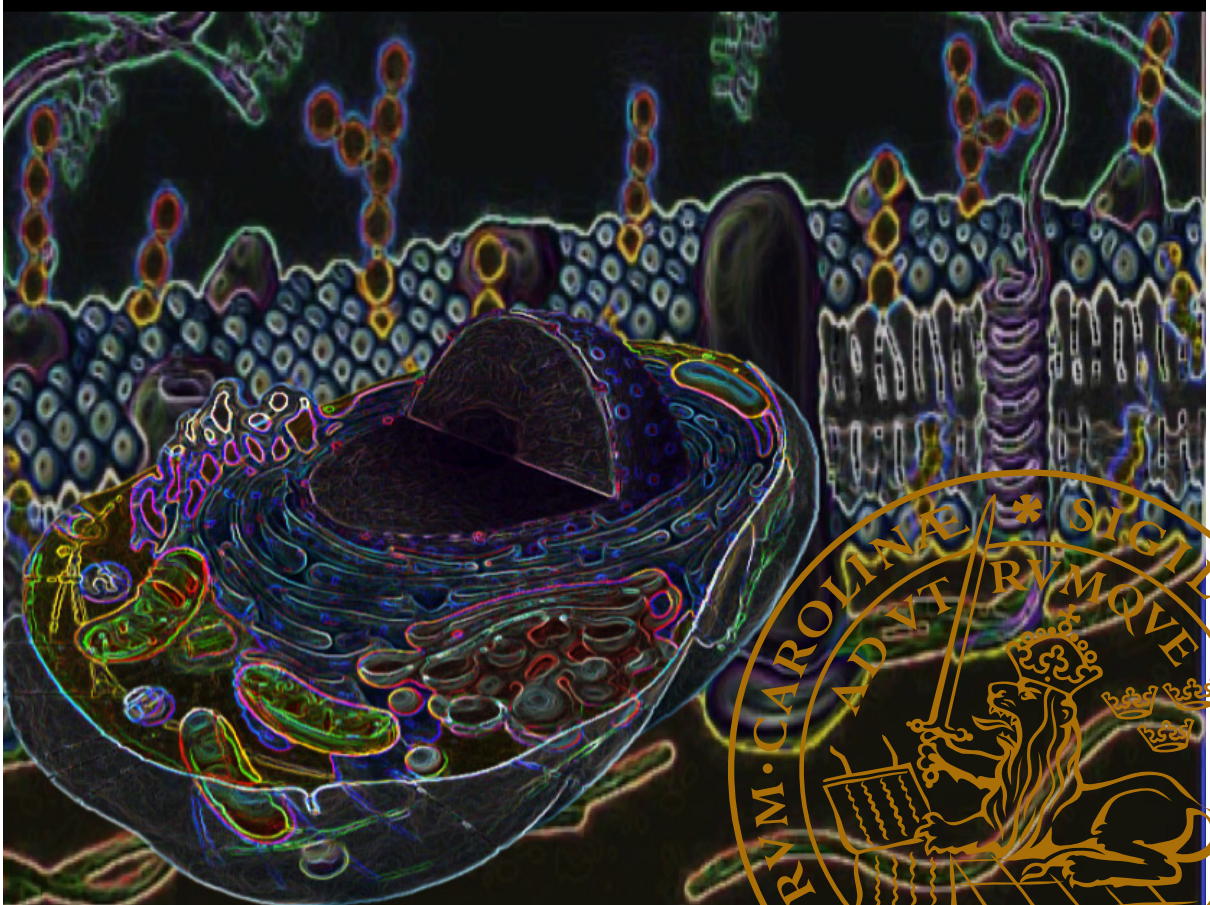
2017

Document Version:
Publisher's PDF, also known as Version of record

Landscaping the cell surface proteome of breast cancer

Following pathways through organelles to the plasma membrane

EMILA KURBASIC
FACULTY OF ENGINEERING | LUND UNIVERSITY



LANDSCAPING THE CELL SURFACE PROTEOME OF BREAST CANCER

Following pathways through organelles to the plasma membrane

Emila Kurbasic

LUND
UNIVERSITY

DOCTORAL DISSERTATION
by due permission of the Faculty of Engineering, Lund University, Sweden.
To be defended at Lundmark Salen, Sölvegatan 27, Lund.

Friday, September 29th at 09.00.

Faculty opponent
Josep Villanueva, Vall d'Hebron Institute of Oncology, Barcelona, Spain

Organization
LUND UNIVERSITY
Department of Immunotechnology
Medicon Village (building 406)
SE-223 81 Lund, Sweden

Emila Kurbasic

Document name
DOCTORAL DISSERTATION

Date of issue
2017-09-29

Sponsoring organization

Title and subtitle

Landscaping the Cell Surface Proteome of Breast Cancer
Following pathways through organelles to the plasma membrane

Abstract

Breast cancer is one of the most common cancers in women. It is most commonly treated by the surgical removal of the tumour in combination with (neo)-adjuvant therapy (hormone, chemo- or radio- therapy before or after surgery). However, a large number of patients are over treated, with approximately 60% being given adjuvant therapy when already cured by surgery alone. This causes many undesirable side effects and cost hence, there is great need for new diagnostic and prognostic methods to decide if patients are in need of adjuvant therapy.

A number of prognostic and treatment predictive factors have been established such as tumour size, hormone receptor status, histological grade and age. Hereditary predisposition to developing cancer can be a factor, with several high penetrance genes identified such as BRCA1 and BRCA2 genes as well as many as a dozen lower risk genes that are however additive in effect. Molecular subtyping has significant prognostic value allowing the differentiation of several subtypes having unique survival outcomes, particularly for tumours highly responsive or nonresponsive to hormonal or targeted drug therapies. Currently the prognostic and treatment predictive factors are mainly based on the primary tumour status, even though the distant metastases are the main reason for breast cancer related deaths. Thus, there is need for novel approaches with higher specificity and sensitivity in newly developed targeted therapies.

The aim of this thesis was to understand the changes in breast cancer tumour cells and tissues, by comparing protein expression levels in different conditions, using mass spectrometry. Attention was specifically on the analysis of patient samples, with pairs of primary tumours and metastases, in order to try to understand what happens to allow tumour cells to be able to metastasise and to identify novel molecular markers. Hence we also investigated the biological functions of proteins by integrating information about the processes and pathways in which these proteins take part in order to be able to better understand the contribution of different proteins to breast cancer development. Further we have explored molecular classification markers for breast cancer tumours into major intrinsic subtypes. Our results demonstrated a great overlap of subtypes, using gene expression and protein expression profiling. In conclusion, all our findings together show the great need for improved and early cancer detection and treatment as well as need for development and promises of personalized medicine.

Key words
Breast cancer, proteomics, membrane proteins, glycoproteins, biomarkers, pathway analysis, organelles, miRNA

Classification system and/or index terms (if any)

Supplementary bibliographical information

Language
English

ISSN and key title

ISBN
978-91-7753-393-1

Recipient's notes

Number of pages 67

Price

Security classification

Signature  Date 2017-08-22

LANDSCAPING THE CELL SURFACE PROTEOME OF BREAST CANCER

Following pathways through organelles to the plasma membrane

Emila Kurbasic

LUND
UNIVERSITY

Cover photo by Emila Kurbasic

Faculty of Engineering
Department of Immunotechnology

ISBN 978-91-7753-393-1 (print)
ISBN 978-91-7753-394-8 (pdf)

Printed in Sweden by Media-Tryck, Lund University
Lund 2017

"The greatest enemy of knowledge is not ignorance,
it is the illusion of knowledge"
Daniel J. Boorstin (Stephen Hawking)

Content

Original Papers

PAPER I

Changes in glycoprotein expression between primary breast Tumour and synchronous lymph node metastases or asynchronous distant metastases
Emila Kurbasic*,* Martin Sjöström, Morten Krogh, Elin Folkesson, Dorthe Grabau, Karin Hansson, Lisa Rydén, Sofia Waldemarson, Peter James and Emma Niméus
Clinical Proteomics **2015**, 12(1):13
doi:10.1186/s12014-015-9084-7

PAPER II

Changes in Protein Glycosylation Pathways during Breast Cancer Tumour Progression
Emila Kurbasic*,* Martin Sjöström, Sofia K. Gruvberger-Saal, Karin Hansson, Liselotte Andersson, Emma Nimeus, Peter James. (Manuscript)

PAPER III

Proteomic analysis of breast tumours confirms the mRNA intrinsic molecular subtypes using different classifiers: a large-scale analysis of fresh frozen tissue samples
Sofia Waldemarson*,* ***Emila Kurbasic****,* Morten Krogh, Paolo Cifani, Tord Berggård, Åke Borg and Peter James
Breast Cancer Research **2016**, 18(1):69
doi:10.1186/s13058-016-0732-2

PAPER IV

Proteome level changes induced by microRNAs
Emila Kurbasic*,* Bruno C. Gomes, António Sebastião Rodrigues Jose Rueff and Peter James (Manuscript)

Authors Contribution to the Papers

PAPER I

Contributed to experiment design. Performed the protein extraction, digestion, glycocapture. Ran LC-MS/MS. Performed MS alignment and database searches, helped with statistical analysis. I drafted manuscript and participated in writing the manuscript.

PAPER II

Contributed to experiment design. I performed MS alignment and database searching. Performed statistical, functional and pathway analysis. I drafted manuscript and helped writing the manuscript.

PAPER III

Designed the SRM assay, prepared samples for and performed SRM, helped with data analysis and manuscript writing.

PAPER IV

Prepared samples for all MS runs, ran shotgun LC-MS/MS. Designed the SRM assay and performed SRM analyses. Performed MS data processing and analysis.and database searching. Participated in the manuscript writing.

Abbreviations

LC-MS/MS	Liquid chromatography tandem mass spectrometry
ER	Oestrogen receptor
DM	Distant metastasis
LNM	Lymph node metastasis
PCA	Principle component analysis
CAN	Acetonitrile
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
ANOVA	Analysis of variance
PgR	Progesterone receptor
SRM	Single reaction Monitoring
PTM	Post-translational modification
TOF	Time of flight
IHC	Immunohistochemistry
ESI	Electrospray ionisation
HPLC	High-performance liquid chromatography
miRNA	micro ribonucleic acid
DNA	Deoxyribonucleic acid
EGFR	Epidermal growth factor receptor
HER2	Human epidermal growth factor receptor 2
mRNA	Messenger RNA
BRCA	Breast cancer gene
MALDI	matrix-assisted laser desorption/ionization
IDC	invasive ductal carcinoma
ILC	invasive lobular carcinoma
ECM	extracellular matrix
SERMs	oestrogen receptor modifiers

Introduction

Breast cancer is one of the most common cancers in women and its incidence is growing constantly due to changes in lifestyle and an increase in longevity. Even though prevalence is increasing, the survival rate has improved in patients over the last few decades, due to early detection with mammography but also due to new, improved treatment methods.

The most common breast cancer treatment is surgical removal of part or the whole breast with addition of radiation, chemo-, hormone-, and antibody therapy separately or in combination, depending on the spread and grade of disease. Adjuvant therapy is used as prevention to the spread of cancer cells, that can be revealed as recurrence in the breast (due to residual primary tumour) or recurrence in other parts of the body than original, often the lymph nodes (local recurrence) or other parts of the body (distant recurrence/metastases). Adjuvant therapy is given in 80% of cases, nevertheless 60% of patients are considered cured after surgery alone. Consequently, there is a large number of overtreated patients with unwanted side effects in connection to unnecessary costs for society. Thus, there is great need and demand for new diagnostic methods to confirm that patient is in the risk zone for recurrence and if this is the case, to introduce necessary adjuvant therapy.

Tumour formation requires several changes in cell behaviour that take place at different time points during tumorigenesis. Altered cells display self-sufficiency in growth-promoting signals, limitless replication, insensitivity to growth-promoting signals, avoidance of apoptosis, but also in the abilities to induce angiogenesis, invade and metastasize. Once the lesions breach the basement membrane or spread to the lymph nodes, they are considered to be one of the two types invasive breast cancers: invasive ductal carcinoma (IDC) or invasive lobular carcinoma (ILC). Most of the breast lesions are represented by IDC (80%), while ILC is poorly represented (10% -15%), while both of the types show distinguishable molecular and genetic features.

Tumours can further be grouped into histopathological types and tumour grades according to the pathological examination. Complexity is increased further by heterogeneity in tumours in different patients with genomic and transcriptomics variations and different disease outcomes. In addition, there is heterogeneity in the each of tumour's microenvironment, consequently leading to the formation of multiple molecular subtypes of the disease. Increase in the disease knowledge has

led to establishment of a number of prognostic and treatment predictive factors such as tumour size, oestrogen (ER) and progesterone (PgR) receptor status, human epidermal growth factor receptor type 2 (HER2) status, histological grade, Ki67 and age. Another factor is the hereditary predisposition to developing cancer with several low penetrance genes identified such as ATM, TP53, CHK2 and the high penetrance BRCA1 and BRCA2 genes that occur in 1 out of 250 women and have a 50% chance of developing cancer with mutations in either of the two genes (Campeau et al. 2008). Four different clinical molecular subtypes of breast cancer have been identified; luminal A (ER/PR positive, HER2 negative), luminal B (ER positive and/or PR positive, HER2 positive), HER2-enriched (ER/PR negative, HER2 positive) and basal-like (ER/PR/HER2 negative/triple-negative (TNBC)) as well as less well characterised normal-like and claudin. Further subtypes have been demonstrated such as the gene expression analysis identifying 6 distinct TNBC subtypes, each of which displayed a unique biology, additionally demonstrating the complexity of the breast cancer disease (Yao et al. 2015; Abramson et al. 2015).

Molecular subtyping has great prognostic value with different subtypes having unique survival outcomes, consequently influencing therapeutic decisions for the treatment response, particularly for tumours highly responsive or nonresponsive to hormonal or targeted drug therapies (Paper III). The prognostic and treatment predictive factors evaluating the nature of the tumour are methodically based on the primary tumour status, even though the predictive value has exhibited flaws when it comes to specificity and sensitivity, particularly in newly developed targeted therapies. This together with the fact that metastases are primarily responsible for cancer-caused deaths indicates that understanding the systemic effects of tumours and factors that regulate the ability of cancer cells to grow at distant sites are of greatest importance.

Distant metastases are the main reason in breast cancer related deaths, with a five-year survival rate of 26%. Most common metastases occur in the skeleton, lungs, liver and the brain but are found even in adrenal glands and ovaries. It is still unclear in detail, what kind of changes tumour cells are going through to be able to metastasize. Normal cells, must acquire a succession of capabilities that enables them to become tumorigenic and ultimately metastasize. Metastatic process includes loss of adhesion, breakdown of ECM and basement membrane and migration to the distant site. Distant metastases are formed in cases where cancer cells are able to survive, proliferate and stimulate angiogenesis in the distant location. The aim of this study was to try to understand the changes in breast cancer tumour cells and tissues, comparing protein expression levels in primary cancers and recurrences/metastases. Changes were measured using mass spectrometry between paired breast tumour tissues from primary tumours and metastases in a large patient group (Paper I, II and III)

In this thesis I have focused on the analysis of membrane proteins since these are the proteins that interaction with the other cells and environment is made through. Accordingly glycoproteins were studied for the most part, due to the facts that 60% of all proteins in the body are glycosylated and 50% of all drug targets are glycoproteins. Glycoproteins are most often membrane proteins with sugar molecules attached to one or more parts of the protein structure. Membrane proteins are low in abundance, compared to the entire cell content, very hard to separate from the rest of the cell and also difficult to analyse. Therefore, we have used enrichment methods (Paper I and IV) before analysis, to get a higher concentration of proteins. In other projects, we have analysed the whole proteome of the cells and tissues to get a clearer image of protein expression in general. Moreover, we have investigated biological functions of proteins by integrating information about processes and pathways in which these proteins take part to be able to better understand the contribution of different proteins to breast cancer development (Paper I-IV).

Breast Cancer

Cancer development

The term cancer is defined as uncontrolled cell growth and is caused by cells proliferating in a manner faster than normal, invading the surrounding tissues and, in some cases, even moving to distant parts of the body creating metastases. Malignant tumours are associated with the loss of tissue functions and eventually death (McAllister, Weinberg, 2014). Cancer is the result of the gradual accumulation of driver gene mutations that successively increase cell proliferation and can be caused either by environmental factors, heredity or random errors made during normal DNA replication (Tomasetti et al. 2017).

Tumours are comprised of two components; the proliferating cells and the stroma, which comprises blood vessels, white blood cells and adipose and connective tissue. For the tumour to progress to metastasis, enzymes such as those of the matrix metalloproteinase class must break down the local host tissue. Once capillaries are damaged or leaking, the broken clusters of tumour cells can travel to distant regions throughout the body. Once malignant, the tumour can induce angiogenesis allowing the tumour to grow its own blood supply to provide itself with the essential nutrients and especially oxygen to allow further growth and escape to secondary sites (Freese et al. 2015; McAllister, Weinberg, 2014). The distribution of cells within tumours is less organized than the microarchitecture of the normal tissues from which the tumour arose those it still resembles some features of the original. Thus, the use of histopathological analyses of tumour sections makes it possible to classify cancers depending on their presumed tissues of origin. Also, it is possible to understand the relationship between the microscopic features of a tumour and the effects that the tumour has on the patient and most importantly, segregate tumours into categories depending on the degree of aggressive growth. Tumours growing without invading adjacent tissues are classified as benign, while those that invade adjacent tissues and develop metastases are termed malignant (Rakha et al. 2010).

Carcinomas or abnormal, invasive cell growths are the most common cancers in humans that account for about 80% of the cancer-related deaths in the Western world. They have their origin in epithelial tissues, where most of human tumours

originate. Epithelial tissues are built up from the layer of cells that form the linings of cavities and channels in the body, beneath which lies basement membrane (basal lamina), separating epithelial cells from supporting connective tissue, stroma. The basement membranes serve the as structural scaffolding of the tissue and are a specialized type of extracellular matrix (ECM). They are mainly constructed from proteins secreted by epithelial cells to which a range of biologically active signalling molecules attaches. Carcinomas can be divided in two main categories, depending on the function they arise from; squamous cell carcinomas, arising from the epithelial cells forming protective seals to cavity or channels and adenocarcinomas, built from epithelial cells that secrete substances into the ducts or cavities that they line, in order to protect the same from their contents (Weinberg et al. 2007). Remaining 20% of cancers are sarcomas, malignant tumours arising from non-epithelial, mesenchymal tissues with origin in the mesoderm of the embryo, hematopoietic cancers of blood lineage cells and neuro-ectodermal tumours affecting the nervous system (Weinberg et al. 2007).

Breast and breast cancer

Breast cancer is one of the most common cancer types in the world and is the leading cause of cancer-related deaths among females worldwide. Although the overall mortality rate amongst breast cancer patients has decreased last few decades, due to the early detection and better adjuvant therapy, there is still no curable treatment for patients with spread to distant sites (Berry, Cronin et al. 2005). Increased mortality rates could be attributed to the increased life expectancy, but also to the changes in the life styles around the world, associated with modern way of living such as increased food and alcohol intake, smoking and physical inactivity (Jemal, Brey et al. 2011; Farley et al. 2013).

The mammary gland is a complex tissue composed of a series of milk-producing lobes connected to the lactiferous ducts. Every mammary gland contains 15–20 lobes and each respective lobe contains a succession of branched ducts ending in the nipple. Ducts are lined with epithelial cells surrounded by an outer layer of myoepithelial cells with contractile properties. Myoepithelial cells affect differentiation, polarity, proliferation and migration of the luminal epithelial cells (Figure 1a). Together, different cell types and components build a branching ductal network embedded in a fibroblast stromal mass composed of fibrous connective tissues, adipose tissues and extracellular matrix (ECM) (Polyak et al. 2010; Ali et al. 2002). Breast tumours that are still confined within the ducts or lobules are known as ductal or lobular carcinoma in situ. As soon as the lesions breach the basement membrane or spread to the lymph nodes, they are considered

as invasive breast cancer. The majority of invasive breast cancers are of epithelial origin known as carcinomas. There are two dominant types of invasive breast cancers, invasive ductal carcinoma (IDC) and invasive lobular carcinoma (ILC). Up to 80% of breast lesions are IDC, while 10% to 15% are represented by ILC. Both of these carcinomas show distinguishable molecular and genetic features (Perou et al. 2000; Arpinoet al. 2004).

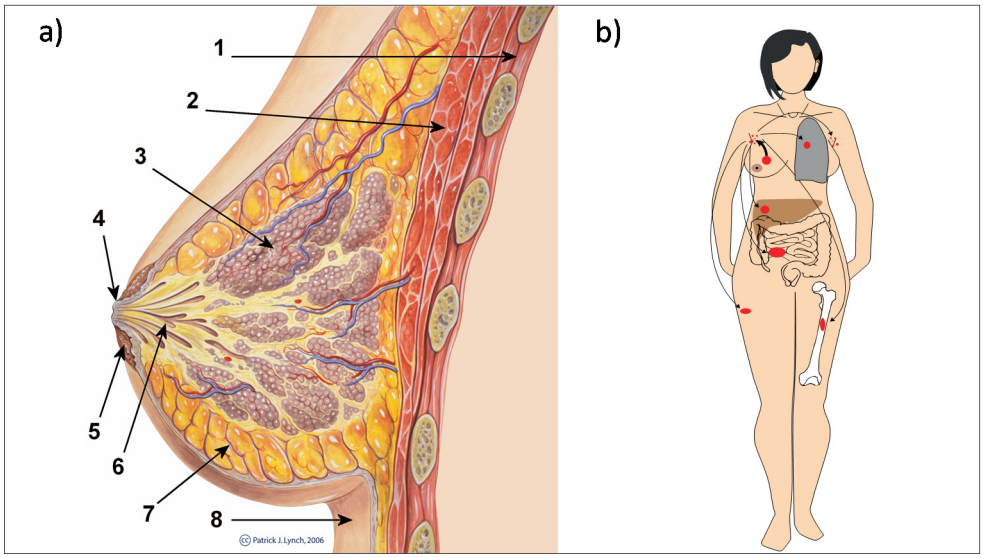


Figure 1. a) Breast anatomy scheme: 1-Chest wall; 2-Pectoralis muscles; 3-Lobules; 4-Nipple; 5-Areola; 6-Lactiferous duct; 7-fatty tissue; 8-Skin b) Schematic representation of cancer spread (Creative Commons Attribution: Patrick J. Lynch, medical illustrator; C. Carl Jaffe, MD, cardiologist).

Pathology and prognostic factors

With the incidence and mortality numbers constantly increasing, there is a need for finding better treatments and develop personalized medical care for breast cancer patients. In order to determine the prognosis of the disease and more importantly, to allow the selection of the most appropriate treatment for the individual patient, breast cancers are most often categorized according to established classification schemes, following the initial diagnosis (Viale et al. 2012). Pathological examination of breast tumours, serves as the basis for classification schemes, grouping the tumours into histopathological types, tumour grades and stages (Figure 2). Invasive carcinomas are subdivided according to their growth patterns and degree of differentiation by assessing histological type and histological grade. Histopathological grade is a well-established prognostic factor based on the degree

of differentiation of the tumour tissue. It is assessed by the Nottingham Grading System (NGS), which is based on the evaluation of three morphological features; degree of tubule or gland formation, nuclear pleomorphism, and mitotic count. According to NGS, tumours can be well-differentiated (grade I), with high homology to the normal breast terminal duct lobular unit, tubule formation, mild degree of nuclear pleomorphism, and low mitotic count; moderately differentiated (grade II) and poorly differentiated (grade III), with a marked degree of cellular pleomorphism and frequent mitoses and no tubule formation Unfortunately, most of the breast cancers have no only one of specific characteristics, thus the role of histological typing is currently limited (Engstrøm et al. 2013).

In addition, breast tumours are evaluated by immunohistochemistry (IHC) for the expression of hormone oestrogen receptor (ER) and progesterone receptor (PR) while human epidermal growth receptor 2 (HER2) is confirmed by fluorescent in situ hybridization (Hannah & Weinberg 2011). The most widely used classification scheme is World Health Organisation (WHO) TNM (tumour, node, metastasis) system. It evaluates size of the primary tumours (T: 1-4 grade), presence of the lymph node metastases (M: 0-3) and the manifestation of distant metastases (M: 0 or 1) (Sobin et al. 2002). Another classification has been developed, based on histological grade of the primary breast tumour through the microscopic analysis, investigating mitotic index, nuclear polymorphism and tubule formation. According to this evaluation of tumours as being well-, moderately- or poorly differentiated, patients with well differentiated tumours have the best prognosis. The Nottingham prognostic Index (NPI) evaluates tumour size, lymph node status and histological grade altogether to determine the prognosis and suggest the best suited therapy for a certain patient (Elston et al. 1991).

There is a high heterogeneity between tumours in different patients with genomic and transcriptomics variations and different disease outcomes, but there is also heterogeneity within each individual tumour's microenvironment (Russnes, Navin et al. 2011). The molecular heterogeneity existing within breast cancer gives rise to the formation of multiple molecular subtypes of the disease. Thus, great efforts are being made to address the challenges associated with establishing reliable markers and improve understanding of the molecular mechanisms and pathway deregulation in the breast cancer pathology. In order to prevent the development of distant spread, a combination of local and systemic treatment is used, and the choice is mainly based on the characteristics of the primary tumour. Surgical removal of the tumour is used as the standard effective treatment, but in some patients in which tumour has moved from the primary growth site, this treatment have been proved to be ineffective. When it comes to adjuvant treatments like the use of endocrine agents, they have reduced breast cancer caused deaths significantly during the last decade. Complications arise when certain patients

become resistant to endocrine therapies or do not respond to this kind of therapy at all (Ali, Coombes 2002). A number of prognostic and treatment predictive factors have been identified such as tumour size, oestrogen (ER) and progesterone (PgR) receptor status, human epidermal growth factor receptor type 2 (HER2) status, histological grade, Ki67 and age. Lymph node involvement is also assessed during surgery to determine if the tumour has spread, which requires dissection of the axilla and adjuvant treatment. The prognostic and treatment predictive factors assessing the nature of the tumour are all routinely based on the status of the primary tumour (Ignatiadis et al. 2007). Another, and maybe the most important factor is the hereditary predisposition for developing cancer and several low penetrance genes have been identified such as ATM, TP53, CHK2 as well as the two main high penetrance genes, BRCA1 and BRCA2 (Paper III) which occur in 1 out of 250 women and cause a 50% chance of developing cancer with mutations in any of the two genes (Narod, Foulkes et al. 2004; Antoniu, Phraoah et al. 2003).

Molecular subtyping

There has been an increased emphasis in using molecular approaches to improve breast cancer diagnostics due to the limited efficacy of the conventional classification methods. The receptor status of ER, PgR and HER2 has prognostic value in predicting effectiveness of targeted hormone and cytotoxic drug treatment against these receptors. However, the predictive value has insufficient specificity and sensitivity, and is ineffective for newly developed targeted therapies. Moreover, these traditional classification schemes alone are unable to capture the genetic diversity that is present within the largest IDCs group. Studies revealed that distinctive molecular features associated with IDC-classified breast tumours, such as differential expression of the three receptors and proliferative genes, such as Ki-67, could be used to segregate tumours into various intrinsic subtypes (Perou et al. 2000; Sørlie et al. 2001). There were four subtypes identified; luminal A (ER/PgR positive, HER2 negative), luminal B (ER positive and/or PgR positive, HER2 positive), HER2-enriched (ER/PgR negative, HER2 positive) and basal-like (ER/PgR/HER2 negative/triple-negative) (Sørlie et al. 2003).

Molecular subtyping has great prognostic value since each subtype has a distinctive survival outcome. The division of luminal subtypes into luminal A and B is of clinical interest because, despite both groups being positive for ER, they have different prognoses with luminal A having a better prognosis than luminal B. Similarly, significantly worse prognosis is observed in HER2-enriched and basal-like subtypes, even though lately prognosis for HER2 positive cancer patients have

been improved due to new treatments based on antibody therapies (Subbiah et al. 2014).

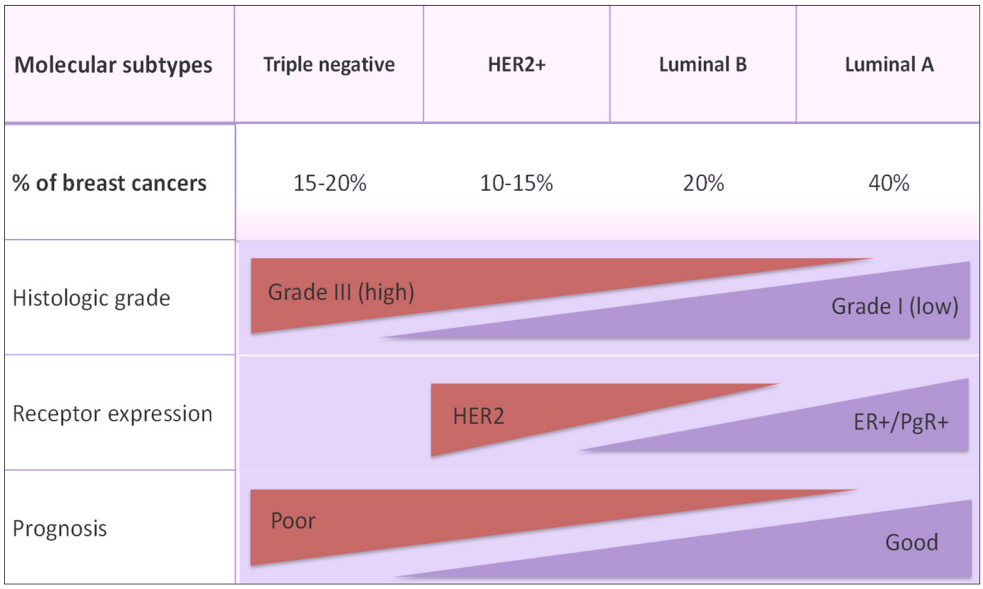


Figure 2. Breast cancer pathogenesis.

Molecular subtyping also influences therapeutic decisions serving as a predictor for the treatment response, specifically for tumours that are highly responsive or non-responsive to hormonal or targeted drug therapies. Patients with ER+ breast tumours are treated with drugs such as Tamoxifen that block ER activity while those, which are HER2+, benefit from anti-HER2 drugs such as Trastuzumab. Tamoxifen is the first choice for adjuvant therapy, especially for premenopausal women as it reduces breast cancer recurrence by 50% and mortality rate by 31%. It belongs to the selective oestrogen receptor modifiers (SERMs) and acts mainly via ER, by binding to it and blocking its transcriptional activity (Rivera-Guevara et al. 2011; Ali et al. 2016). HER2-positive breast cancer, in which HER2 is regularly overexpressed, accounts for 20–30% of all breast cancers and has the second poorest prognosis among breast cancer subtypes. Trastuzumab is a recombinant humanized monoclonal antibody that binds to the extracellular domain IV of HER2 and can have three different mechanism of action: Trastuzumab has been proposed to trigger HER2 internalization and degradation; it can attract immune cells to tumour site that overexpress HER2 and cause antibody-dependent cellular cytotoxicity (ADCC) and also act as an inhibitor of MAPK and PI3K/Akt pathways increasing cell cycle arrest and the suppression of cell growth and proliferation (Barok et al 2014; Vu et al 2012).

In contrast, patients with breast tumours that lack the three receptors (triple negative patients), i.e. ER, PgR and HER2, are not expected to respond to these targeted treatments and are only suitable for surgery and systemic chemotherapy. Although the gene expression-based stratification of breast cancer has led to better insights into the biological diversity of breast cancers, the overall underlying molecular mechanism(s) in breast tumorigenesis, is still poorly understood. Gene expression does not necessary correlate with the expression of proteins, which are known to be the key mediators of cellular processes. Single gene can have various spliced protein variants and proteins themselves undergo a wide range of post-translational modifications (PTMs) such as glycosylation and phosphorylation, which can affect their biological functions to the great extent. Proteomics as a study of protein expression and the associated PTMs is therefore a crucial method in understanding function and regulation of key biological events during malignant breast transformation (Parker et al. 2009; Falck et al. 2013).

Metastasis development

Because metastases are responsible for most cancer-related deaths, understanding the systemic effects of tumours and factors that regulate the ability of disseminated cancer cells to grow at distant sites are of primary importance (Fidler 2002). In order to evolve progressively to a neoplastic state, normal cells, must acquire a succession of the hallmark capabilities, thus cancer cells need to acquire the traits that enable them to become tumorigenic and ultimately malignant (Hannah et al. 2011). The metastatic process is initiated through the loss of adhesion, the breakdown of the ECM and the basement membrane. Cancer cells then migrate to the blood or lymph vessels and are transported to the site of anchoring, where they adhere to the vessel wall and undergo extravasation. In the case where cancer cells are able to survive, proliferate and stimulate angiogenesis, a new tumour develops at the distant location (figure 1b). Otherwise, if they cannot perform any of the process in metastatic development, they either form micro metastases or the metastatic process fails (Fidler 2002).

Tumours are complex tissues composed of multiple distinctive cell types that participate in interactions with one another and even recruiting normal cells, to form tumour-associated stroma and participate in tumorigenesis by contributing to the development and expression of hallmark capabilities (Hannah et al. 2011). Studies have indicated that less than 0.01% of the circulating tumour cells actually leads to formation of metastases (Liotta et al. 1974). Organ distribution of metastases doesn't seem to be random but is rather determine by expression of adhesion receptors on the tumour cells and the vascular endothelial cells but also

the microenvironment in the target organs, such as expression of the growth factors and chemokines (Ben-Baruch et al. 2007).

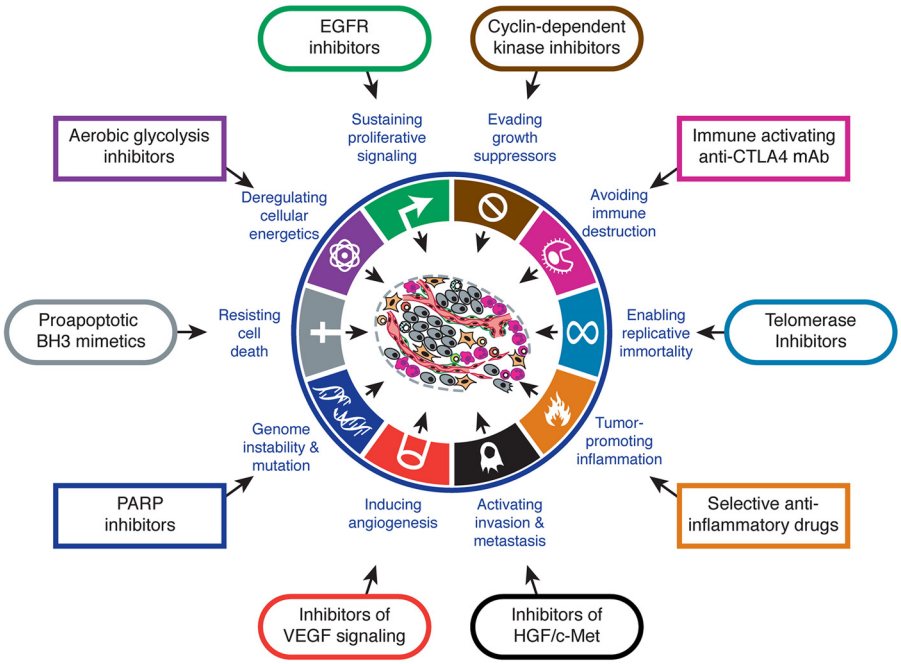


Figure 3. This illustration encompasses the six hallmark capabilities. Adopted from Hallmarks of Cancer, the Next Generation 2011.

There are six originally proposed biological capabilities acquired during the development of human tumours, so called hallmarks of cancer (figure 3) and they comprise an organizing principle for rationalizing the complexities of neoplastic disease. They include sustaining proliferative signalling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis, and activating invasion and metastasis. Underlying these, there are more hallmarks that have been recently added, such as are genome instability, which generates the genetic diversity that expedites their acquisition, inflammation reprogramming of energy metabolism and evading immune destruction. In addition to cancer cells, tumours contain a selection of recruited, seemingly normal cells that contribute to acquisition of hallmark traits by creating the tumour microenvironment (Hannah et al. 2011).

Currently, another probable hallmark of cancer is emerging in the form of somatic mutations of the disease, including base substitutions, small insertions/deletions,

rearrangements and copy number changes. The mutational theory of cancer suggests that changes in DNA sequence, driver mutations, bestow proliferative advantage of the cell and lead to outgrowth of a neoplastic clone. While some of these mutations are inherited in the germ line, most of them arise in somatic cells during the lifetime of the cancer patient, due to multiple mutational processes, such as endogenous and exogenous mutagen exposures, aberrant DNA editing, replication errors and defective DNA maintenance. A number of base substitution mutational signatures and rearrangement signatures have been found, however, dominantly acting activated fusion genes and non-coding driver mutations appear rare. Additional rarely mutated cancer genes probably exist but are not yet known and a role and occurrence of single driving genes are still to be identified (Nik-Zainal et al. 2016).

Proteome Analysis – from Cells to Tissues

Proteomics is the study of the proteome, defined as the entire set of proteins, in a particular cell or tissue, at a specified time, space and condition (Wilkins et al. 1996). It is the natural complement to genomics, and major advances in human proteomics were made possible by sequencing and publishing of the human genome (Wheelock et al. 2006). Proteomics comprises the structural analysis of proteins, protein identification and protein abundance measurements but also qualitative and quantitative characterizations of proteins and any post-translational modifications (PTMs) associated with it (Wilkins et al. 1996; Venter et al. 2004). Gel-based techniques such as two-dimensional gel electrophoresis (2DE) (Klose 1975) and differential in-gel electrophoresis (DIGE) (Unlu et al. 1997) have been long used methods for protein separation, until the development of the novel non-gel techniques and methods in proteomics. The rapid progression of proteomics has been enabled thanks to development of liquid chromatography (LC) in combination with advanced mass spectrometry (MS), allowing high throughput protein identification and quantification (Doerr et al. 2010). By combining appropriate upstream methodologies such as sample preparation and fractionation with downstream computational tools, proteomics is a powerful analytical approach to allow parallel global proteome profiling and identification of distinct protein expression patterns in tumorigenesis, in order to discover cancer biomarkers for possible treatments (Menon et al. 2010). Despite the technological advances, resulting in the rapid, accurate and highly sensitive analysis of proteins at a relatively low cost, the MS-based proteome analysis remains challenging due to the complexity of samples investigated as well as the extensive dynamic range of proteins present within the proteome. In order to overcome these issues, new methodologies have emerged to fractionate, separate or enrich samples in order to increase dynamic range and depth of analysis (Qian et al. 2006). In plasma or serum samples with a very complex content and high concentration of proteins, depletion strategies are employed in order to detect proteins of lower abundance (Bjorhall et al. 2005). In cases where the whole proteome needs to be analysed or sample complexity reduced by analysing only a portion of it, samples can be fractionated either by gel-based or gel-free techniques. Gel-based methods are capable of separating proteins, whereas gel-free methods in proteomics are

typically used to separate peptides after digestion (Apweiler et al. 1999; Scherp et al. 2011). Mainly, in this thesis gel‐based separation of proteins was performed by one-dimensional electrophoresis (1DE) on sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) followed by proteolytic digestion of proteins into peptides and LC-MS (Paper II, III, IV and V).

Mass spectrometry-based proteomics

The enzymatic release of peptides from proteins for MS characterization is referred to as "bottom-up" protein identification, commonly used in shotgun proteomics today (Yates et al. 2004). Undertaken as a global approach, with the whole proteome being investigated it can be executed as the top-down or bottom-up methodology. In the "top-down" strategy intact proteins or large protein fragments are directly analysed, to preserve the biological organization, including position of PTMs, within the protein. Due to the analytical challenges like limited throughput (different protein properties making separation difficult) and sensitivity top down strategy is best suited for studying single proteins or simple protein mixtures, while the "bottom up" method, remains the preferred approach for most current proteomics research (Compton et al. 2012). Either gel-based and gel-free separation proteomics can be coupled with advanced LC-MS/MS. The LC serves to chemically separate peptides on the basis of differential peptide hydrophobicity using reversed phase (RP) chromatography column packed with octadecylsilyl (C18) stationary phases (Lai et al. 2010).

In cases where the whole proteome needs to be analysed or sample complexity reduced by analysing only a portion of it, samples can be fractionated either by gel-based or gel-free techniques. Gel-based methods are capable of separating proteins, whereas gel-free methods in proteomics are typically used to separate peptides after digestion (Apweiler et al. 1999). In this thesis gel‐based separation of proteins was performed by one-dimensional electrophoresis (1DE) on sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) followed by proteolytic digestion of proteins into peptides.

Protein mass spectrometry

A mass spectrometer separates analytes based on their mass over charge ratio (m/z), resulting in an array of m/z values, called a mass spectrum. The analytes are, in this case, peptides from digested proteins (usually trypsin). Peptides are built from a series of amino acids (aa) into a chain. This chain can be fragmented or dissociated by adding kinetic energy to the peptide in the gas-phase by collision

with neutral gas molecules that result in fragmentation in a random fashion. Consequently, fragments of different lengths are generated, forming a-, b - and c-fragments from the N-terminal and x-, y- and z- fragments from the C-terminal end of aa (figure 4). These fragment types are the most common fragments observed with ion trap, triple quadrupole, and q-TOF mass spectrometers. This nomenclature widely used and was modified by Biemann, from an earlier proposed nomenclature by Roepstorff and Fohlman (Roepstorff and Fohlman al. 1984; Biemann 1998).

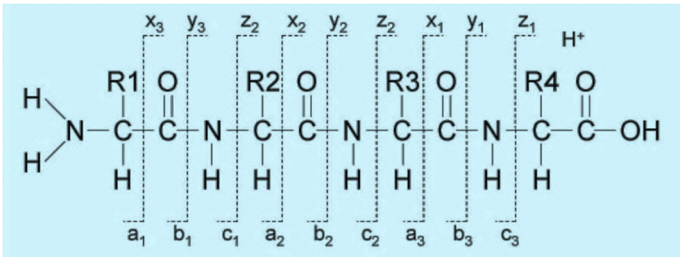


Figure 4. Peptide fragmentation and ions generated through mass spectrometry procedure.

The chemistry of fragmentation is still not fully understood and there is no certain way to predict how a peptide will fragment. Fragmentation is both length- and composition-dependent (Simpson 2003). Several different fragmentation methods are in use as present, most common one being collisional induced dissociation (CID) (Cramer et al. 2001) that works by colliding analytes into innate gas such as argon, producing mainly y- an b- fragments. Higher-energy collision dissociation (HCD) (Phanstiel et al. 2008) produces mostly y- fragments. Electron-capture dissociation (ECD) and electron transfer dissociation (ETD) are accomplishing fragmentation by low-energy electron donation and produce predominantly c- and z- fragments (Palmblad et al. 2002; Mikesh et al. 206). The spectrum of an intact analyte before fragmentation is called an MS1, and the spectrum of its fragments is called MS2.

After fragmentation gas phase ions are separated in the mass analyser, which is normally a combination of electric and magnetic fields. Most widely used and analysers are quadrupole, quadrupole ion-trap, Fourier transform-ion cyclotron resonance (FT-ICR) and time-of-flight (TOF) analysers. The relative abundances of ions generated, are read as magnitude of current produced and measured in detector as function of time giving result in m/z values or intensities of the ion (Simpson 2003). Increasing gradient of hydrophobic organic solvent, usually acetonitrile (ACN) elutes bound peptides progressively to subject them to ionization in the MS. The two most popular ionization sources for MS are matrix-assisted laser desorption/ionization (MALDI) (Karas and Hillenkamp 1988) and

electrospray ionization (ESI) (Fenn et al. 1989). Although both methods of ionization are useful in proteomics, ESI typically produces a range of multiply

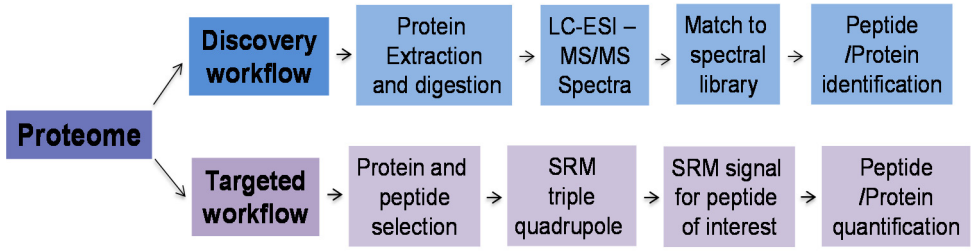


Figure 5. Proteomic workflow, discovery vs. targeted workflow

charged ions that can be detected in ideal m/z ranges of all common types of mass analysers for bio molecular analysis. Thus, efficient identification and characterization of large biologically important macromolecules such as peptides, proteins, nucleic acids and carbohydrates is made possible (Benerjee et al.2012). ESI utilizes electrical voltage and heat in the solution, followed with a narrow capillary forming an electrical spray, composed of charged small drops. These micro-drops are evaporated, increasing number of the charges on their surface. Thus, ions are transformed from liquid to the gas phase generating peptide precursor ions (Ho et al. 2003; Macher and Yen 2007). MALDI involves embedding the analyte molecules (i.e. peptides) in an organic matrix, which ablates and ionizes when the matrix absorbs energy from a laser (Liyanage et al. 2006). The matrix should be a highly UV absorbing substance, normally a small organic acid, producing, predominantly single charged ions. It should also be able to adsorb energy from the laser and protect analyte molecules at the same time (Canas et al 2007).

Protein Quantification

The ability to accurately quantify protein expression in comparative studies represents an important but challenging task enabling determination of proteins that may play key biological roles in disease development and potential biomarkers or drug targets. Currently, MS-based relative quantitation of individual proteins from different samples can be undertaken either via label-assisted or label-free techniques. Label-free quantitation strategy involves unlabelled peptide mixtures being analysed in separate LC-MS/MS runs. The relative protein

abundance is determined by comparing the precursor peak ion intensities or MS/MS spectral counts of the corresponding peptides across the samples (Liu et al. 2004; Sandin et al. 2014). Precursor ion intensity is quantified by measuring the area under curve in the LC-MS, nevertheless, peak area can vary from run to run, and between two injections thus extensive normalization is required to account for variation amongst samples. Alignment of runs, also called match-between-runs, is conducted before normalization, and is accomplished by different algorithms to automatically align the peptide peaks or extracted ion chromatograms prior to normalization and further comparisons (Lai et al. 2013).

A major issue in label-free is the potential bias arising from sample preparation and data acquisition (Callister et al. 2006). However, such bias can be minimized with careful sample handling and implementation of good practises for LC-MS/MS data acquisition, for example, the inclusion of quality control runs and performing all analyses in a single batch in randomized order. Furthermore: Intensity normalization as part of the data processing workflow can help in reducing technical variation (Callister 2006, Chawade 2014). In the work in this thesis, signals were normalized using the algorithms in MaxQuant, Progenesis and R. In addition, to increase the statistical significance of the measured protein fold-change, it is necessary to perform sample analysis in either multiple biological or technical replicates. The main advantage of label-free approach is the ability to analyse and compare many samples in a single experiment with a relatively low cost. Label-free quantitation has been shown to have superior and more profound proteome coverage than iTRAQ (Patel et al. 2009). It is in particular suitable for large-scale studies of differentially expressed proteins and mapping of unique molecular signatures associated with specific conditions. Therefore, the label-free quantitation method was chosen as the method of choice for the quantitative proteomics analyses in most of our studies (Paper I, II, IV and V).

Due to the complexity of biological samples there is increase in a requirement for detection and quantification of large numbers of proteins in proteomic studies. Therefore, a frequently used methodology is the “shotgun” approach, in which the sample proteins are enzymatically digested to peptides, and then peptides detected in the sample are automatically selected by the mass spectrometer in a process named Data-dependent acquisition (DDA) (Lange et al. 2008).

Another, targeted approach, selected reaction monitoring (SRM) has been used for protein quantification as well as for targeted MS measurements of preselected sets of proteins, such as candidate biomarkers or proteins constituting cellular networks. Specifically, SRM can be a useful tool when a number of proteins need to be measured across multiple samples in a reproducible manner (Picotti et al. 2012). SRM exploits the unique advantage of triple quadrupole (QQQ) MS (Figure 5) to use first and the third quadrupoles as filters to select predefined m/z

values of peptide ions while the second quadrupole serves as collision cell in which peptides are fragmented. Several transitions per peptide are monitored over time, giving a set of chromatograms with intensities as a function of retention time (Papers III and V). The SRM method does not measure full mass spectra, which leads to increased sensitivity compared with shotgun technique. In addition, it enables the detection of low abundant proteins in complex mixtures, which is of importance in any systematic quantitative proteomic study (Lange et al. 2008).

Bioinformatics tools for protein identification

Protein identification starts with matching the obtained MS/MS spectra to a theoretical database generated from in silico digestion of protein sequences, with the suitable parameters such as cleavage procedures, modifications, species, precursor and fragment ion mass tolerance and charge states. The most widely used identification tool in proteomics is curated non-redundant and publicly available Uni-Prot/Swiss-Prot database, including the entire set of known human proteins and their predicted fragment spectra (Bairoch et al. 2000). In addition, high-quality spectral libraries with experimental spectra have been created for more accurate spectral matching and scoring (Lam et al. 2007).

The calculation of a false discovery rate (FDR) for validation of the results of database searches is necessary to minimize false-positive identifications. A decoy databases, obtain by shuffling or by the protein sequences been reversed from the true target protein database, is used to estimate the number of random matches in the target database. The acceptable FDR is pre-specified before the database search and the value is calculated afterwards and reported as the ratio of the number of matches in the decoy database to the number of matches in the target database (Elias et al. 2005).

One of the main factors for obtaining confident peptide identifications from complex mixtures is the resolving power and mass accuracy of the mass spectrometer, which, in turn, affects the stringency and the FDR of the database search. Predominant fragments ions, produced from the cleavage of the polypeptide backbone, in MS/MS spectra are usually the b and y ions. Database searches are based on the computer algorithms that match experimentally derived fragment ions against pre-defined database and return a list of peptide sequences, ranked by their probability score and FDR. (Nesvizhskii and Aebersold 2005). A number of complex algorithms for the protein identification from LC-MS/MS data have been developed, the most commonly used are Sequest (Eng et al. 1994), Mascot (Perkins et al. 1999) and X!Tandem (Craig and Beavis 2004). In recent years, a number of proteomics software has been developed, containing pipeline

features that allow comprehensive analysis of high quality MS data including protein and PTM identification as well as peptide/protein validation and quantitation. These include open source proteomics software, such as Trans Proteomics Pipeline, OpenMS and Proteios Skyline and freely available MaxQuant but also commercial ones as Proteome Discover and ProteinPilot (Craig and Beavis 2004). Mascot and X!Tandem were the main software tools used for protein identification in this thesis, together with Skyline (Deutsch et al 2010) and MaxQuant (Cox et al 2009).

Functional analysis - Pathways and interaction networks

The final goal of our MS analysis of the biological samples was to identify potential biomarkers for breast cancer diagnosis, prognosis and targets for therapy. To achieve that, an improved understanding of the global and integrated view of molecular mechanisms underlying breast cancer biology is essential. A number of studies using high throughput proteomics approaches have generated a vast amount of high quality data enabling comprehensive system-wide investigation of differential protein regulation in breast cancer. Cancer is a systems biology-wide disease with multiple oncogenic proteins involved simultaneously in different cellular processes (Hornberg et al. 2006). Thus, an emerging objective is a functional interpretation of the proteome-wide changes. Currently, with the overwhelming amount of information collected from the large data sets, we rely heavily on the effective use of bioinformatics tools for the functional proteome pathway and protein-protein network analysis.

The most basic approach in functional proteome analysis is to categorize the identified proteins using Gene Ontology (GO) terms consisting of defined descriptors that relate proteins with their biological processes, molecular functions or cellular components (Ashburner et al. 2000). This is often accompanied by looking for overrepresentation of GO terms in a dataset, using for example Gene-Set enrichment analysis (GSEA) (Subramaniana et al. 2005). More in-depth analysis can be performed from the perspective of biological pathways or protein-protein networks, which seek to understand the key processes underlying the functional roles of differentially expressed proteins by statistically evaluating their relationships and interactions with one another in a given condition. Bioinformatics tools proved to be indispensable for this type of data mining as they are able to organize and reduce the complexity of large volumes of data to present a visual view of significantly important biological patterns and relationships. A number of commercial tools are available such as Ingenuity pathway analysis IPA (Ingenuity System) and MetaCore (Nikolsky et al. 2005). In

this thesis MetaCore, Panther web tools Version 11.1 (released 2016-10-24) and Functional Evaluation of Expression Regulation (FEvER) (Kirik et al. 2012), an in-house developed tool, were mostly used for functional and pathway analysis. UniProt was also used in order to further explore functions and GO terms of important proteins.

In an effort to gain better mechanistic insights, proteome-based studies that explored the functional aspects of the resulting proteomes identified important oncogenic processes and protein interaction networks that are critical for cancer progression (Hu et al. 2011) indicating the usefulness of these approaches to interpret large data sets. Some recurring themes that emerged from these analyses include; perturbations of cellular structural integrity, changes to the extracellular matrix (ECM) composition, abnormal intracellular signalling, increased cell locomotion and an activated immune system in the cancer pathophysiology. These altered processes seem to be coordinated by changes in the expression of groups of functionally similar proteins such as cytoskeletal proteins, extracellular matrix proteins, cell surface integrins, tyrosine kinases, adhesion proteins and peptide-presenting proteins. Individual groups of proteins can interact within their own network as well as work synergistically with other protein groups to promote cancer invasion and metastasis (Wu et al. 2010).

Dissecting the Proteome

One of the main challenges in functional proteomics today is the separation of protein subpopulations from complex protein mixtures to be able to perform detection of low abundance proteins and allow for reliable quantitative and qualitative analysis. Because of the complexity of a proteome and differences in protein properties, subdividing the proteome into partial proteomes, thus maximising the coverage of low abundance proteins possible has become a standard procedure. Different methods, depending on what kind of sub-proteome one wants to explore have been developed to create standardized and reproducible procedures for sample preparation and protein separation (Michelsen and Hagen 2009).

Membrane Proteins

Cells, but also the organelles inside them are coated and protected by lipid and protein containing membranes (Bränden et al. 1999). The lipid bilayer encompasses proteins whose chains cross the lipid bilayer (integral membrane proteins), proteins that anchored to the membrane (anchored membrane proteins) and proteins strongly bounded to the integral membrane (associated membrane proteins) (Elofsson et al.). Membrane proteins are responsible for connecting cells to each other, for cell–cell communication but also large number of them has important functions, as receptors and transporters for traffic regulation across the membranes and between organelles. The plasma membrane contains a number of anchor proteins responsible for attachment of cytoskeletal proteins and the extracellular matrix both inside and between cells (Wu et al. 2003). Due to their important functions and accessibility at the surface of the cell, it makes sense that 70% of currently available drugs target membrane proteins. In order to understand any important functions of membrane proteins, they must be first isolated and purified. Membrane proteins are difficult to purify, for several different reasons, including low abundance, low solubility of the lipid bilayer, they are embedded in and difficulties with their digestion. Therefore, there is a need for modifications to the purification protocols for membrane protein isolation, most of all when it comes to detergents required in maintaining the solubility and function of these

proteins (Lin et al. 2009). Enzymatic digestion adjustments are harder to implement, since the trypsin, one of the most used enzymes in proteomic studies, lack of ability to efficiently digest hydrophobic trans-membrane domain peptides (Coughenour et al. 2004). The reasons for this is the structural accessibility problem as well as the low number of Lysine and Arginine amino acids in the trans-membrane protein regions, making it almost impossible for enzyme to digest protein into proteins of the correct length for MS analysis (Schluesener et al. 2005). Membrane protein and functional analysis in proteomics is even more complicated by the fact that membrane proteins are present in addition to the cell membrane, in the array of subcellular compartments (organelles), resulting in the number of sub-proteomes (Pasquali et al 1999).

An additional difficulty in 'subproteome' analysis is the fact that numerous subcellular membranes are interconnected, such as connection between the outer nuclear membrane and endoplasmatic reticulum (ER), resulting in overlap in the distribution of membrane proteins between organelles (Macher et al. 2007). Many of membrane proteins are discharged from the membranes and can be then transported to the different parts of the cells, resulting in the multiple localisations (co-localisation) for approximately 39% of all organelle proteins (Foster et al. 2006). Due to all the above difficulties, the wide use approach in membrane protein analysis is enrichment prior to the analysis, either by the rate-sedimentation centrifugation or density gradient centrifugation, for organelle membrane separation and enrichment (Macher et al 2007). Different PTMs are used in order to both fractionate samples, for easier analysis, but also because of their biological significance in the cells and diseases. Common PTMs used for sample fractionation are phosphorylation and glycosylation, latter being investigated into more depth in this thesis (Paper I) (Temporini et al. 2008).

Glycoprotein Enrichment

Early detection of cancer is perhaps the most promising approach to ensure the long-term survival of patients and in recent years, the focus has been to identify reliable biomarkers and efficient methods for screening the same in biological fluids and tissues. Thus, the efforts have been made for identifications of specific alternations, at the cellular level, during the oncogenesis. One such alternation is glycosylation, which is affected by malignant transformation through the altered expression of glycosylated enzymes (Silva 2015).

Structural diversity created by the N-glycans on membrane-bound and secretory proteins is fundamental to establishment of a number of biological functions, including cell proliferation, differentiation, migration, cell-cell integrity and

recognition, immune modulation and signal transduction (Christiansen et al. 2013). Aberrant protein N-glycosylation can therefore disrupt normal cellular functions leading to lack of cellular homeostasis and pathophysiological conditions (Sakamoto et al. 1986). It has become increasingly evident that aberrant protein glycosylation is associated with a number of pathological conditions including many human cancers (Christiansen et al. 2013). It has been shown that glycan-associated antigens on breast tumour tissues positively correlate with the increased metastatic potential and a poor prognosis for breast cancer patients (Madjd et al. 2005).

Protein glycosylation, is a very extensive (more than 50% of all proteins are glycosylated) and biologically significant post-translational modification that occurs in all forms of life starting with prokaryotes, to complex eukaryotic systems and organisms in which an array of glycoproteins is generated (Spiro et al. 2002). A vast array of different glycosylation forms and structural diversity are presented in the form of glycoproteins, in spite the limited number of monosaccharide building blocks utilized for mammalian glycan synthesis. The diversity of the glycosylated proteome, and hence the diverse functions of glycoproteins, is dramatically increased by protein macro-heterogeneity (such as partial glycan occupancy of various glycosylation sites on the polypeptide backbones) in combination with the extensive micro-heterogeneity, caused by variation in the glycan length (number of building blocks), the monosaccharide compositions (the type of building blocks), the topology or branching of the chains and the linkage types (Moremen et al. 2012).

Most common types of glycosylation in humans encompass glycans enzymatically attached to the protein via either N- or O-glycosidic linkages. In protein N-glycosylation, an N-glycan precursor is added via reducing-end N-acetylglucosamine (GlcNAc) residue to asparagine (Asn) residues on polypeptides found within consensus peptide sequences or “Asn-sequons” displaying Asn-X-Serine/Threonine (Ser/Thr), where X can be any amino acid residue except for proline (Bause et al. 1983), whereas O-glycosylation occurs at serine or threonine residues without any strong consensus sequence (Tylor et al. 2006).

Altered expression leads to changes in glycan biosynthesis and diversity, affecting the post-translational modifications of proteins. The end results are aberrant, cancer specific and antigenically distinct glycoproteins, which can, together with their corresponding antibodies, be targeted for cancer biomarker research (Silva 2015). Together with the fact that more than 50 % of proteins are glycosylated and that approximately 60% of drug targets are glycoproteins, the need for glycomic research is undeniable and it is expanding rapidly, covering many aspects of scientific research from basic science, over therapeutic areas with fusion to neighbouring analytical disciplines such as proteomics and glycoproteomics. In

cancer research, where the attention is focused on identifying unique expression patterns of N-glycans, special attention is placed on detecting N-glycan changes associated with cancer and possible candidate N-glycan biomarkers of sufficiently high sensitivity and specificity for early diagnosis and monitoring of cancer progression. The significance of N-glycosylation in cellular functions during tumorigenesis might also become understood using a comprehensive approach based on functional glycomic studies (Turnbull et al. 2007).

N-glycome enrichment (figure 6) is an approach to capture the N-glycosylation status of cancer cells and biological samples (Paper I). It is performed to allow analysis of glycoproteins, which are first separated from protein mixtures by binding of N-glycans to the hydrazide resins on beads. The detachment of N-glycopeptides is accomplished enzymatically by N-glycosidase F (PNGase F) treatment, which specifically hydrolyses the amide bond between N-glycan and the Asn residue, converting the Asn to an aspartic acid residue in the process (Takashi et al. 1978). However, this method is rather inefficient from the point of global glycol analysis, since the cleavage entails the loss of information on the protein isoform of origin of the released glycopeptides but also information on the varieties of glycans attached to these peptides.

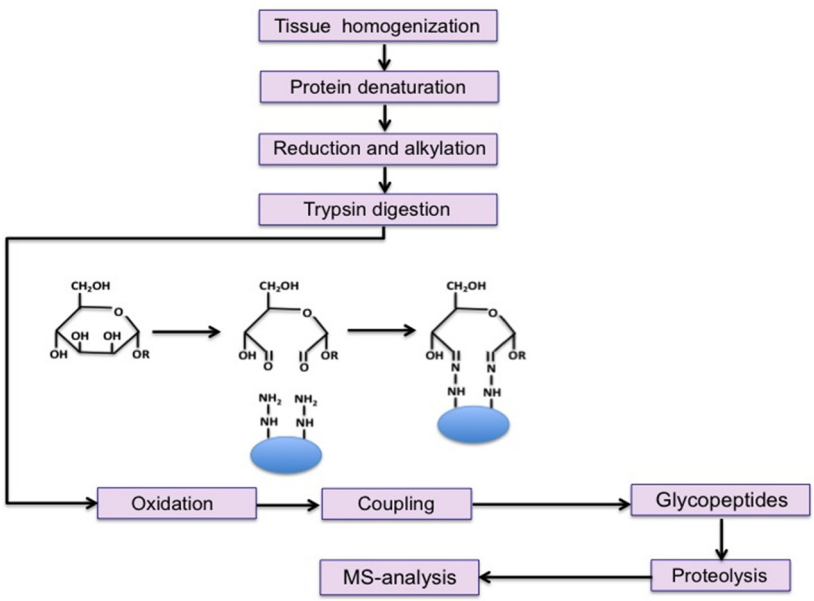


Figure 6. A schematic representation of the Glycocapture method: The method selectively isolates and quantifies N-linked glycopeptides from mixtures of glycoproteins. Oxidation: periodate oxidation converts the cis-diol groups of carbohydrates to aldehydes; Coupling: the aldehydes react with hydrazide groups immobilized on a solid support to form covalent hydrazone bonds. The non-glycosylated peptides are removed by washing (and can be collected, for further analysis), and the N-linked glycopeptides are subsequently released from the solid-phase by Peptide N-glycosidase F treatment for MS analysis.

A vast array of different glycosylation forms exists giving rise to a rich structural diversity of glycosylated proteins despite the fact that the number of monosaccharide building blocks is rather limited. The first source of variability comes about from partial glycan occupancy of various glycosylation sites on the polypeptide backbones and the second one is caused by variation in number of building blocks which in turn affects the glycan length, monosaccharide compositions, branching and linkage types. This diversity of glyco-proteome and consequently the diverse functions of glycoproteins, in combination with protein macro-heterogeneity and micro-heterogeneity (Moremen et al. 2012), indicates a great need to combine the knowledge gained from glyco-proteomics with that of glycomics and further investigate the conformation/linkages of sugar molecules connected to the proteins as well as their position in the protein. In global glycan profiling, relative abundances between glycans within glycome populations are studied and the glycoprofiles are then compared between samples. As a part of the thesis, I tried to investigate changes in glycomes between breast cancer cells, from cell culture with xenografts grown in mice during a placement in one of the leading glycomics laboratories in the world under the guidance of Prof. Nicki Packer. Unfortunately, this study was not successful due to the low amounts of material available and the relative insensitivity of glycomic vs. proteomic analysis.

Organelle Proteome Enrichment

Beyond the role in protecting living cells from their external environment, biological membranes in eukaryotes serve to compartmentalize intracellular organelles, such as the envelope of the nucleus, the outer and inner membranes of the mitochondria, membrane cisternae complex of the ER, Golgi apparatus, as well as lysosomes and secretory vesicles (Tan et al. 2008). Organelles are the entities within eukaryotic cells, each with their own chemical characteristics and different highly specialized functions, depending on their organ and tissue-specific localizations in the organism (Sadowski et al 2008).

Through the knowledge of the localisation of proteins we are able to evaluate and study their function. Thus, there is a need for reliable protein subcellular localisation determination. Only through the thorough investigation of proteins sub-cellular localization, the whole picture of the proteome, protein functions and the cellular processes they enrol in can be resolved (Gatto et al. 2010). In order to be able to interact, proteins are often co-localized in the same sub-cellular structure (Huh et al 2003). Thus, compartmental location of a protein is essential, when it comes to the knowledge of potential interaction partners, which are limited by its location. The function of a protein is therefore, often defined by the

organelle in which it is located and the proteins that it can interact with (Gatto et al. 2010). A certain stimulus, like stress or a disease can also drive translocation of proteins between organelles, changing their function, either in turning on/off function of the proteins, or by a complete change in a protein function. (Boisvert et al. 2010; Mulvey et al. 2013). Even though proteins in organelles can be located in the lumen, one third of the proteome is believed to be associated with the membrane (Stevens et al. 2000). As membranes contain many low-abundant proteins, information on its proteome is not easily obtained from total proteome analysis.

The main challenges in proteomic work with analysis and characterisation of membrane proteins are the poor aqueous solubility and their various, mostly low concentrations. Contamination during enrichment is an additional issue, due to poor organelle separation or inconsistency in the lab work and technical replicates are necessary to ensure more confident measurements. Other challenges include the wide dynamic range of expression but also the fact that protein composition is constantly changing, with proteins moving between different organelles and even interacting with proteins from other compartments (Wu et al. 2003).

Characterisation of the organelle proteins is performed using two main types of techniques. *In situ* direct microscopic visualisation of the location of individual proteins is done using imaging approaches, either expression linkage with the green fluorescent protein (GFP) or tagging with a fluorescently labelled antibody specific to the protein of interest. MS-based proteomic techniques involve the isolation of organelle fractions of a sample to identify and quantify the protein content of the cell. MS-based approaches involve numerous methods for purifying organelle fractions, among many others, subtractive proteomics and gradient approaches (Simpson et al. 2006). Subtractive proteomics involves partial purification of subcellular organelles biochemically and partially in silico. Two different fractions, the biochemical fraction of interest and a fraction enriched in proteins known to contaminate it, are isolated and separately analysed by MS. Thereafter proteins found in both fractions are subtracted from the dataset, thus identifying proteins that are restricted to the organelle of interest (Kavanagh et al. 2007; Michelsen and Hagen 2009). In the gradient organelle separation, a number of fractions are sampled along the continuous gradient, to determine their subcellular localization (Graham 2015). Multidimensional protein identification technology (MudPIT) is often used to reduce proteome complexity on readily enriched organelles. This technique is based on a gel-free two-dimensional high-performance liquid chromatography (HPLC) approach in which complex protein mixture is denatured and digested with proteases and separated in a capillary packed with a strong cation exchange (SCX) resin following with a reversed phase (RP) material and interfaced with a quaternary HPLC system coupled to a tandem mass spectrometer (Michelsen and Hagen 2009).

Gradient approaches

A widely used technique for separation based on the density and size differences in a mixture of organelles and proteins is centrifugation, that benefits of the fact that, bigger and heavier molecules settle at the bottom of a centrifuge tube in a shorter time. Standard mixture of cell homogenate contains organelles of different sizes, densities and shapes, which can therefore can be separated by centrifugation. Each fraction obtained in this way contains a number of different organelles with similar sedimentation velocities. Sedimentation velocity depends on both size and the density of the molecule, thus the fraction can be further separated based on density, independent on their sizes, by a process known as density gradient centrifugation (Michelsen and Hagen 2009).

Iodixanol density gradient centrifugation is used for the analysis of subcellular membranes. It relies on the creation of a density gradient using sucrose or iodixanol, layering of a cell extract on the top of the gradient and then centrifugation until the organelles equilibrate out at the isopynic point and can then be isolated by fractionation of the gradient (figure 7)(Graham 2015).

In order to decide a proteins subcellular position, proteins are matched to the distribution of known organelle markers, employing statistical pattern recognition techniques. Thus, using proteins with identified organelle localization as the basis, other proteins co-localizing in the same organelle can be identified by matching their distribution to the distribution of the biomarker (Gatto et al. 2010).

Protein correlation profiling (PCP) is a label-free quantification gradient-based approach in which abundance of every protein in each fraction is calculated from the area of the obtained ion currents. Abundances are then used to calculate peptide-specific distributions, which are then matched against a specific profile to map sub-cellular proteomes (Foster et al. 2006).

The analysis of the proteome of an organelle depends not only on protein identification but also on quantification, evaluation of contamination, reliability of organelle markers and normalisation. The challenge of clarifying the distribution of the multiply located proteins is difficult to solve and complicates the interpretation, due to either poor migration during the fractionation or due the to the problem of the fractions that contain no established markers to enable annotation. Another limitation is that some organelles are too low in abundance to be identified by current technologies. In addition there is a need to increase resolution by applying multiple (as orthogonal as possible) gradients, since one density gradient may resolve some of the organelles while different gradients are needed to separate the other organelles (Gatto et al. 2010).

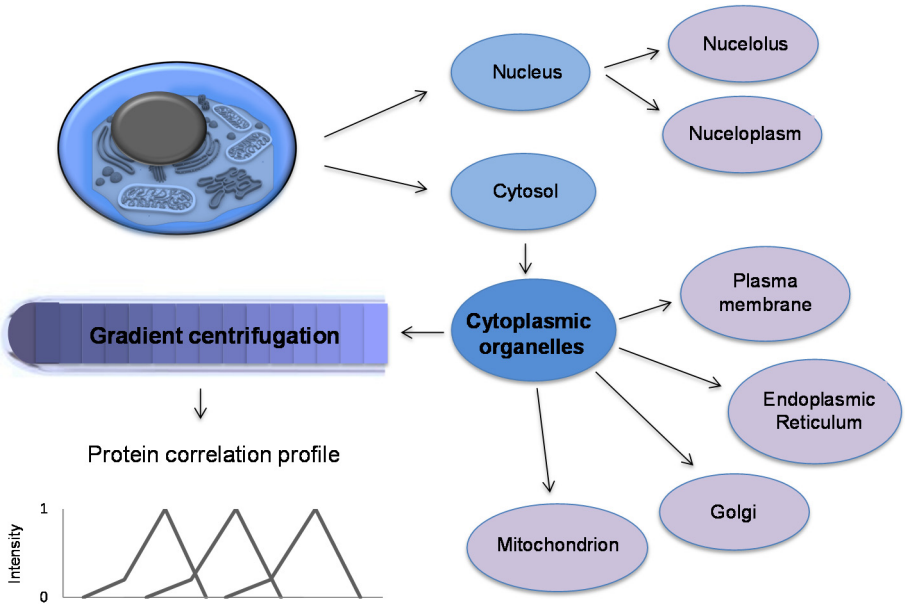


Figure 7. Gradient Centrifugation Profiling of organelles. Cell lysates with intact organelles are separated by iodixanol density gradient centrifugation and ion intensities for peptides from marker proteins are measured by mass spectrometric analysis of each gradient fraction and called Protein Correlation Profiles (PCPs)

Chemotherapy, DNA damage and repair

Chemotherapy presents the most efficient treatment for the majority of advanced tumours. However, each patient responds to the treatment individually and thus high variability in response is quite common. Resistance of tumour cells to various chemotherapeutic agents often leads to limited therapeutic options and unfavourable outcome of patients.

Doxorubicin (Adriamycin), is an anthracycline glycoside antibiotic which has been widely used either as a single agent or in combination with other chemotherapeutic regimens for adjuvant treatment in cancer patients for over 30 years (Ji et al. 2015; Tacar et al. 2013). Doxorubicin acts by entering the nucleus, and interfering with enzymes involved in DNA replication and through intercalation causing DNA double-strand breaks in dividing cells (Andersen et al. 1999). By this interference with DNAs uncoiling, doxorubicin induces a range of cytotoxic effects in process of anti-proliferation, resulting in DNA damage. In response to DNA damage, cells attempt to repair the DNA-breaks and if this fails, the apoptosis pathway is triggered leading to programmed cell death. Intercalation into the DNA, leads to inhibition of both DNA and RNA polymerase, eventually

terminating both DNA replication and RNA transcription (Tacar et al. 2013, Forrest et al. 2012). Doxorubicin is also known to target multiple molecules to produce a range of cytotoxic effects such as activation of various molecular signals from AMP-activated protein kinase inducing apoptosis (AMPK) to alter the Bcl-2/Bax ratio in the Bcl-2/Bax apoptosis pathway, leading to downstream activation of different caspases and apoptosis. While providing a cure, doxorubicin causes toxicity to most major organs as well, especially life-threatening cardiotoxicity, which forces the treatment to become dose limiting. Doxorubicin also induces apoptosis and necrosis in healthy tissue, causing toxicity, there have been many efforts to develop a drug delivery system that would diminish these affects including liposomes, hydrogel and nanoparticle systems (Tacar et al. 2013).

DNA is constantly exposed to the both endogenous and exogenous treats such as chemicals and radiation, resulting in the range of damages like covalently bound bases, base adducts and breaks in one or both DNA backbones. Damaged cells have developed an advanced repair system, due to the fact that these kinds of damages can prevent DNA from being properly replicated and transcribed (De Bont 2004). DNA damage stimulates cellular response called the DNA Damage Response (DDR), inducing molecular pathways to detect and repair, or in cases of or irreparable damages, apoptosis or cellular senescence (Pearl et al. 2016). There are eight DNA damage signalling and repair pathways up to the date, including: Direct damage signalling (DDS), Direct reversal repair (DDR), Trans-lesion synthesis (TLS), nucleotide excision repair (NER), base excision repair (BER), mismatch repair (MMR), homologous recombination repair (HRR) and non-homologous end joining (NHEJ) (Hoeijmakers 2009, Scott et al. 2014).

Due to the a lack of specificity and the high variability in response between patients after chemotherapy, a number of studies have demonstrated that exposure of cells to a single drug may result in cross-resistance to that or many other structurally and pharmacologically unrelated drugs. This phenomenon is known as the multidrug resistance (MDR) (Kunicka et al. 2014). In order to improve patient treatment and survival, it would be imperative to understand the mechanisms behind MDR and identify the proteins, changes in their modifications and thus functions responsible for resistance. It is well known that proteins activity and function can be changed depending on where it is in the cell. Investigation of proteins sub-cellular localization is a crucial step in getting the full picture of the proteome of the cell and any possible functions and cellular processes that the proteins are enrolling in. Often proteins are co-localized in the same sub-cellular structure in order to be able to interact, thus co-localization could be an effective tool in determining the yet unknown function of a protein (Yang et al. 2013).

Conclusion on organelle proteome enrichment

The aim of this study was to try to understand cellular responses to DNA damage as important determinants of both cancer development and cancer outcome following radiation therapy and chemotherapy. Identification of molecular pathways governing DNA damage signalling and DNA repair in response to different types of DNA lesions allows for a better understanding of the effects of radiation and chemotherapy on normal and tumour cells. Although deregulation of the DNA damage response (DDR) is associated with predisposition to cancer development, it can also result in hypersensitivity or resistance of tumours to therapy and can be exploited for improvement of cancer treatment (Goldstein et al. 2015).

Unfortunately, after looking at our data in PRoloc package for analysis of protein localization in organelles in R, we realized our data was too weak to make proper conclusions. After looking through our data, conclusion was that our replicates were not overlapping properly and on top of that; we had to many missing values in each replicate. This resulted in a very low number of eligible proteins for analysis (see small square in figure 0) making the data set unworkable.

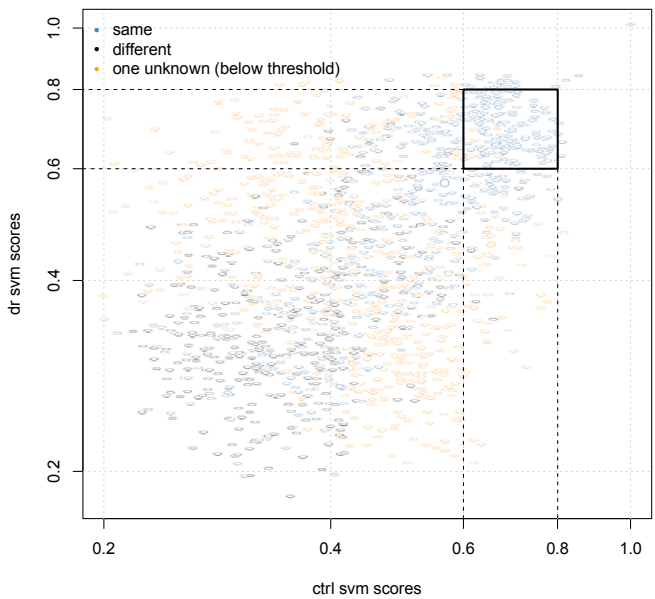


Figure 0. Classification results for ctrl The distributions of classification scores per organelle. Those with scores equal to 1 are only composed of markers and have not actual classification result. Each point in the figure represents a protein and axes are the classification scores for ctrl (x axis) and dr (y axis). The colors agreeing (blue) or disagreeing (black) are results after thresholding. When one classification result is unknown, the dot is yellow.

microRNA Involvement in Breast Cancer

It has been long known that only approximately 2% of the human genome codes for proteins, while remaining genes are referred to as “junk DNA”. Nevertheless, recent studies have shown that non-coding DNA has more importance than previously thought in cell homeostasis and that RNA is not only a molecule of information transmission, through messenger RNA (mRNA) (Alexander et al., 2010; Esteller, 2011). New technologies have enabled identification of numerous non-coding RNA transcripts (ncRNAs), categorized into two main, distinctive RNA classes: short or small ncRNAs and long ncRNAs (lncRNAs). Aberrant expression of ncRNAs has been linked to several diseases, including cancer (Silva et al., 2015). Among the all small ncRNAs, such as endogenous small interfering RNAs (endo siRNAs), transcription start site associated RNAs (TSSa-RNAs), transcription initiation RNAs (tiRNAs), promoter-associated small RNAs (PASRs) and promoter upstream transcripts (PROMPTs), microRNAs (miRNAs) has been the most studied one (Esteller, 2011). The biogenesis of miRNAs is a complex process that takes place in the nucleus and in the cytoplasm. (Westholm and Lai, 2011). Between small size of miRNAs and no need for the perfect pairing between a miRNA and a target site there is wide selection of genes that can be target to regulation, thus one miRNA can regulate multiple mRNAs. On the other hand, these properties make it much harder to predict the putative targets. miRNAs can act through mRNA destabilization, translational repression and can even activate gene expression, with destabilization of the target mRNA being common and the latter two much more rare events. These activities make miRNAs extremely relevant to many cell mechanisms and are most probably involved in many diseases such as infections, diabetes and cancer (Balatti et al., 2015). Recent work has shown that even supposedly ncRNAs actually encode small proteins (Akilesh Pandey Natuer 2014).

miRNAs in cancer

Shortly after their discovery, miRNAs were associated with cancer (Johnson et al. 2005) and soon it became clear that miRNA expression levels differ between normal and tumour cells. Furthermore, miRNAs displayed tissue specific expression signatures and evidence of promotion or suppression of tumour development and progression, influencing hallmarks of cancer (Hanahan and Weinberg, 2011). miRNAs can also function as tumour suppressors in the number of cancers, acting by repressing oncogenes. Oncogenic miRNAs, are much less frequent and normally are either up-regulating oncogenes or down-regulating tumour suppressor genes (Bonci et al. 2016).

There are several groups of tumour suppressor miRNAs that have been studied and one of the most examined is probably the miR-200 family due to their important role in tumour cell adhesion, migration, invasion, and metastasis (Korpal et al. 2008). Out of thirteen miRNAs believed to be involved in tumour metastases, four are members of the miR-200 family: miR-200a, miR-200b, miR-200c, and miR-141 (Gravgaard et al. 2012). Some studies have revealed the dual effects of the miR-200 family as the tumour suppressor and yet with overexpress during metastasis. Their functions in the epithelial-to-mesenchymal transition (EMT) and mesenchymal–epithelial transition (MET) are critical for the late stages of metastasis, enabling the tumour cells to grow at distant sites (Banyard et al. 2013). Thus, the miR-200 family is down-regulated in the primary tumours to enable cells to enter in EMT and then in distant recurrences the same cells up-regulate miR-200 family to facilitate establishment of MET (Zhang et al. 2012, Jabbari et al. 2014). miR-203 is labelled as a tumour suppressor in an array of cancer types, including breast cancer (Zhang et al. 2011) and it also influences metastasis by direct targeting of SNAI2 (Shi et al., 2015; Zhao et al., 2015).

It is not still fully understood how miRNA deregulation occurs, possibly during miRNAs biogenesis. The control of miRNAs expression begins by transcription inhibition of the miRNAs genes, such as genetic alterations (mutations, deletions), epigenetic regulation like methylation of promoter sequences and histone modifications or repression of transcription factors by oncogenes (Kim et al. 2009). Another possible way of deregulation happens at the pre-miRNA processing, by genetic mutations and transcriptional regulation control (Lin and Gregory, 2015).

Soon after their discovery, miRNAs were associated with phenomena of developed drug resistance (DDR). miRNAs have been linked with drug resistance pathways such as drug metabolism, drug transport and DNA repair. High number of miRNAs are transcriptionally induced after the DNA damage (Templin et al.

2011), inversely, many miRNAs target DDR genes, controlling feed-back and feed-forward loops of the response (Bottai et al. 2014).

One of the first indications that miRNAs are involved in the regulation of the DDR was the study in which knockdown of the miRNA biogenesis pathway resulted in increased sensitivity to ultraviolet light (UV) and altered cell cycle after UV damage (Pothof et al. 2009). Many reports after that showed that different DNA damaging agents induce changes in patterns of miRNAs expression (Wouters et al. 2011) indicating that alterations in miRNAs are implicated in tumour response to anti-cancer agents.

BRCA1 is a major component of the DDR pathway involved in HR but also in other DNA repairs pathways. It encodes for a nuclear phosphoprotein whose function it is to maintain genomic stability and has critical roles in DNA repair, cell cycle checkpoint control, transcriptional regulation, apoptosis and mRNA splicing (Savage and Harkin 2015). Mutations in BRCA1 are associated with an increased risk of developing breast and ovarian cancer. In breast cancers, levels of the miRNAs are inversely correlated with BRCA1 protein levels, miRNAs are over-expressed in TNBC, the most common type of breast cancer in women with BRCA1 mutations (Garcia et al. 2011).

miRNAs in breast cancer

In recent years, the use of genome-wide approaches has enabled the establishment of miRNAs profiles in tumours and in its normal tissue counterpart, resulting in miRNA expression signatures (miRNAome) making it possible to discriminate different types of cancers and the tissue of origin with high accuracy (Volinia et al. 2006). Several of the miRNAs have been associated with different subtypes of breast cancer; a number of miRNAs are differently expressed in luminal, HER2, basal-like and normal-like subtypes of breast cancer (Serpico et al. 2014). Some miRNAs are differentially expressed between normal and breast cancer tissues, having both tumour suppressive and oncogenic functions. Their possible clinical value, as novel biomarkers for cancer is being currently investigated. The main difficulty with identifying biomarkers is the identification of the target mRNAs as well as proteins with biological significance in breast cancer and the pathways they are involved in. Rather high rates of both false-positives and false-negatives are present in the current miRNA target prediction platforms, thus it is important to experimentally identify significant miRNA targets and the pathways involved in carcinogenesis (Garzon et al. 2010). Studies of the mechanisms behind regulation of miRNAs expression levels and the possible effect of miRNAs in breast cancer therapy resistance showed that miR-200c seems to have opposing effects in MCF-

7 and MDA-MB-231 cells regarding treatment with chemotherapeutic agents and with miR-203 in combination (Gomes et al. 2016). Also, the expression levels of miR-200c and miR-203 in breast tumour tissues on their putative targets, SIX1 and SOX2 showed little effect of miR-200c, expression between normal breast and tumour tissues, while miR-203 effects were significant, showing that miR-203 is involved in, mainly, the early stages of breast cancer development. Accordingly, miR-203 might be a potential marker to determine different stages in invasive lobular carcinoma (Gomes et al. 2016).

Most studies rely on the effects of one miRNA on one or two targets but due to the complex mechanism of its action, it does not show the real impact of a selected miRNA on the cell processes. Therefore, transiently expression or inhibition of miRNAs, the approach that relies on the observation of multiple miRNAs targets could provide much more information about what proteins and pathways are influenced by a miRNA in question. Quantitative proteomics is widely used to detect difference in protein expression between conditions or to resolve molecular portraits of different cancer subtypes. In a similar way, this approach can be a powerful tool to experimentally identify new miRNAs targets and confirm putative targets discovered by bioinformatics tools (Geng et al. 2016, Gamez-Pozo et al. 2015).

In Paper IV, we address the validation of putative targets of the miR-200c and miR-203, by conducting a comparative proteomic study between wild-type MDA-MB-231 and MCF-10A breast cancer cell lines and the same cell lines with transiently expressed or inhibited miR-200c and miR-203. TargetScan (Lewis et al. 2005) that predicts biological targets of miRNAs by searching for the presence of conserved 8mer, 7mer, and 6mer sites that match the seed region of each miRNA was, used in this study. DIANA Tools (Reczko et al. 2012) was also used for miRNA-recognition elements (MREs).

Concluding Remarks

Scientific advances are becoming more and more data driven and existing trend that data can be produced and stored more massively and cheaply, is likely to accelerate in the future. This has immense impact on science, giving us new opportunities as well as challenges. The biggest challenge might be the demand for valid statistical analysis for large data sets, with goals to develop effective methods that can accurately predict the future disease outcomes and at the same time increase understanding of the relationship between the disease topographies. Large sample size, and consequently large data sets give rise to opportunity of understanding heterogeneity across different sub-populations but also extracting important common features across many sub-populations with large individual variations (Fan et al. 2014).

In the past few decades, substantial amount of proteomics data of breast cancer samples and cell lines has been accumulated in online platforms. These data include data from mice models, human tumour xenografts models in mice, and clinical biological specimen such as serum and tumour tissues. Underlying these efforts is the anticipation that comparing proteins expressions between disease states will possibly provide molecular signatures or yield insight but also comprehension of signalling pathways and biological processes leading to initiation and progression of breast cancer. In turn, this knowledge could be used to identify novel biomarkers, which in future could be used as drug targets (Ponomerako et al. 2016).

Metastasis is a terrible event leading to death in the most cancers. The growth of cancer cells at distant organs from a different tissue requires complex processes of detaching from the original tissue; invasion through the basement membrane; movement in the bloodstream or lymphatic system; and anchorage in other organs (Zielinska et al. 2015).

The trend in Sweden is toward an increasing number of breast cancer cases, due in part to the improved detection of tumours. The prognosis has improved with new treatment modalities and earlier detection. However there is currently no curative approach for patients with distant metastases. The subsequent adjuvant therapy is usually based on the analysis of the primary tumour even though as we have shown in our studies, hormone receptor status can change after migration. In order to be able to make a diagnosis and eventually develop the treatment for metastasis,

the analysis has to be based not only on the primary tumours but also local recurrences and metastases.

The motivation underlying efforts of our studies was the assumption that the protein expression between different stages of disease (and normal state), possibly providing molecular signatures or produce an insight into the intracellular signalling pathways leading to initiation and progression of breast tumours. Consequently, such knowledge may help us finding novel biomarkers and drug targets for cancer detection and treatment.

Proteomic studies are high-throughput and one of the most dynamic tools for detection of differential expression of molecules, posttranslational modifications, and involvement in complex structures, so when it comes to the greater context of interpretations of protein interactions and functional networks, proteomics strength is indisputable. In our study on miRNA influence on expression of the large number of proteins, mass spectrometric measurements were clear choice, enabling us to generate a great amount of data, with lists of hundreds of proteins that are differentially expressed. To achieve the goals of precision medicine, we have focussed on the new markers that can be used in metastatic breast cancer to discover new drug targets, and also prognostic markers that could enhance the early detection of metastatic disease. Our aim was to identify possible markers of metastatic breast cancer using two cell line models that represent non-aggressive tumour cells and metastatic breast cells. The results were in line with the idea that breast cancer is a very heterogeneous disease and that changes between a primary tumour and the metastases is specific to every individual patient. We also demonstrated that the tumours group when considering functional changes that occur. The results indicate that there are cell-wide changes in protein expression preparing the cell for escape and invasion and adaptation to the new tissue environments that the cells find themselves in. The new niches are not random but are prepared (pre-metastatic niches) by proteins and miRNAs that are secreted by the primary tumour that prepare distant sites to accept incoming tumour cells.

Further, our findings support and expand the molecular classification markers for breast cancer tumours into major intrinsic subtypes, demonstrating the great overlap of subtypes formed using gene expression and protein expression profiling. These findings have important implications for the use of genomics and expression analysis for prediction of protein expression, such as receptor status and drug target expression. Our findings give us a possible platform for the development of protein expression profiling for identifying novel molecular markers, providing the promise of the personalized medicine, in which cancer detection and treatment with cheap characterization of tumour tissue and suitable choice of treatment is made possible.

Our studies revealed also that biologically significant proteins are differentially expressed in the secretions and in the membranes of recurrent breast tumours relative to their counterparts. Further, pathways significantly involved in tumorigenesis have been identified. A set of membrane and glycoproteins as possible biomarkers has been discovered due to the analytical depth and the diverse molecular focus of this thesis. However, whether these proteins and processes are causing or resulting from breast tumorigenesis needs further assessment by more targeted studies.

The studies of the glycans, connected to the proteins, would be of great value, if they were successful. Still there are some efforts in the finding the ways to study proteins and/or peptides together with the attached modifications, which could improve the knowledge of their functions, interactions with other proteins and consequently involvement in the diseases.

After using several different approaches of analysing proteins and proteome (and even sugars) I can see the need in science of combining different approaches in order to really find the answers to the questions and solve puzzles of cancers but even other diseases. There is no one single best method in doing research but the collective efforts are needed in finding the solutions. Also there is a great need in improving the data analysis, especially when it comes to the big data sets, that are difficult to handle, partly due to the huge amount information generated but even more so complexity of the biological samples analysed.

It is evident, from the studies presented in this thesis, that deep molecular characterisation and understanding could generate solid knowledge platforms from which potential biomarkers and therapeutics can be developed, is needed to drive precision medicine and avoiding extensive overtreatment, and allowing for precise patient stratification.

Populärvetenskaplig Sammanfattning

Bröstcancer är en av de vanligaste tumörformerna hos kvinnor och ökar stadigt på grund av ökad livslängd och förändrade levnadssätt. I Sverige drabbas var tionde kvinna. Trots ökad sjukdomförekomst, så har överlevnadsfrekvensen bland cancerpatienter ökat de senaste årtiondena, tack vare tidig upptäckt med mammografiundersökningar men även nya, förbättrade behandlingsmetoder.

Den vanligaste behandling av bröstcancer idag är bortoperation av tumören med tillägg av strålbehandling, cellgifter, hormonterapi eller antikroppsbaserad behandling, enskilt eller i kombination, beroende på sjukdoms spridning och svårighetsgrad. Utökad terapi används i syfte att minska risken för spridning av cancerceller, som kan påvisas i form av återfall i bröstet (primära tumörer), spridning till andra delar av kroppen än den ursprungliga, ofta till närliggande lymfnoder (lokalrecidiv) eller i en annan del av kroppen (fjärrmetastaser). Tilläggsbehandling ges i 80% av fallen, även då ca 60% av patienter antas vara botade efter operationen. Detta medför att många patienter överbehandlas med obefogade biverkningar som följd med onödiga kostnader för samhället., Nya metoder for diagnostisering behövs för att säkerställa om patienter är i riskzonen för återfall. Om så är fallet, förändras prognosen drastiskt och tilläggsbehandling blir nödvändig.

Alla individers kroppar fungerar olika och därför reagerar alla inte likadant på olika cancerbehandlingar. Det kan bero på skillnader mellan patienter, som ålder eller tumör relaterade olikheter, som storleken och histologisk gradering. Samtidigt finns det ett antal skillnader på molekylär nivå, som upptagningsmekanismer av läkemedel och substanser men även på olika egenskaper hos cancerceller, som vävnads ursprung och vilka proteiner (genprodukter) som uttrycks i cellen. Därför behövs bättre molekylär karaktärisering av bröstcancer för att kunna behandla varje individ på rätt sätt. I den här avhandlingen har jag följaktligen valt att analysera skillnader i uttrycket (mängder) av proteiner, mellan olika tillstånd i bröst cancer celler och vävnader.

Skillnader i proteinuttryck i metastaser antas vara större, jämfört med lokala återfall (lokalrecidiv). För att bilda tumörer i andra ställen i kroppen måste

cancercellerna ta sig in i blodkärlen och transporteras vidare, till en miljö lämplig för deras delning. För att uppnå det, måste de anpassa sig till kroppens naturliga försvarsmekanismer som den bildat för att skydda sig själv, bland annat ett välutvecklat filtreringssystem i lymfnoderna som cancercellerna måste passera för att ta sig vidare till kroppen. Det är fortfarande oklart vilka förändringar cancerceller undergår för att klara av metastasen. Vårt mål var att försöka förstå vilka förändringar som händer i cancerceller på proteinnivå, och därför har vi jämfört parade tumörvävnadsprover mellan primära tumörer och metastaser i en större grupp av patienter (Paper I och III). Glycoproteiner är oftast membranproteiner som har sockermolekyler fästa vid en eller flera positioner i proteinmolekylen. Eftersom 60% av alla proteiner i kroppen är glycoproteiner och 50 % av alla läkemedels måltavlor är glycoproteiner, valde vi i den här avhandlingen att lägga vår koncentration på analys av membranproteiner. Membranproteiner är väldigt svåra att separera och analysera, därför har vi i vissa projekt (Paper I), använt oss av anrikningsmetoder innan analysen. I andra projekt har vi analyserat hela proteinuppsättningen i cellen (proteomet) för att få en helhetsbild av vad som försiggår med proteinuttrycket. Utöver analysen av protein uttryck, har vi också studerat deras biologiska funktioner och vilka processer i kroppen de deltar i, för att bättre förstå deras roll i bröst cancer utvecklingen.

För bestämning av proteinmängd i de olika tillstånden har vi använt oss av global analys av proteiner, vilket kallas proteomik. De olika proverna har analyserats med masspektrometri för att upptäcka peptider, som proteiner är uppbyggda av. Därefter identifierades proteiner och deras mängder jämfördes mellan olika prover. Eftersom cellerna har komplexa uppsättningar av proteiner, behövs statistiska analyser och bioinformatiska metoder för att hitta signifikanta signaler i jämförelserna.

Vi kom fram att det finns skillnader i uttryck av vissa proteiner mellan primära tumörer och lokala återfall i samma patient. Ännu fler skillnader hittades mellan primära tumörer och metastaser (Paper I och III) men störst skillnad i mängder av proteiner var mellan olika vävnader (Paper III). Våra studier visar ett antal proteiner som kan, i framtiden användas som biomarkörer (proteiner vilkas uttryck är specifika för en viss kategori cancerpatienter) och kan hjälpa att upptäcka sjukdomen vid ett tidigt skede, i en större patient antal innan cancer sprider sig. Dessutom kan man använda sig av dessa markörer för att förutsäga om patienter kommer att utveckla metastaser och underlätta beslut om den bästa behandlingen för enskilda patienten.

Acknowledgements

The last 5 years have been a journey and they might have been the most exciting time of my life, if my life hadn't been so eventful before that. It was the most exciting time of my life, in a good way at least. Many people helped me along this journey, the last part being little bit bumpy making the help and support of lovely people around me even more appreciated.

Peter James, my supervisor and tutor, both in science and life, thank you, first of all, for making my work possible. You are the wisest, most helpful and kind supervisor a person can wish for, always putting your students first. Even when you're not obligated and now during your hardships, you are trying to do your best and guide me to the finish. Thank you for all science and life lessons, and inspiring me to become the scientist I am today. And last but not least, thanks for all the good times you organized for our group either at the conferences or fun afterworks, we had a blast! It was a privilege working with you.

I owe a special thanks to my co-supervisor, **Fredrik** Levander, for all the extra help during the last year, taking over some of the supervisors' responsibilities on top of his own. I admire his huge knowledge and his motivation to always help and teach us, oblivious biologists some bioinformatics stuff.

Prof. **Nicki** Packer, thank you for having me in your lab and your group at Macquarie University. You provided an amazing opportunity for me to learn all the sweet things about sugars and an insight in how science works outside of Sweden. At the same time, it was a priceless opportunity to explore and learn lots about the beautiful, amazing Australia and it's amazing people. I learned heaps from you and I respect your wisdom and experience in science.

The entire glyco-group made me welcome to Macquarie, still few people made my days so much more enjoyable with great chatters in office, during the lunches, afterworks and dinners, so thank you **Chris A**, **Chris K, Mafruha** and **Zhiping** for good times. **Shaz** and **Daniel** my friends in and outside the lab made my time in Sydney so much more pleasurable. Thank you for the talks, lunches, dinners and great parties. Daniel, extra thanks to you and Erick for nice dinners and sharing your house with me.

Anna and Shezri my 'Village' buddies, without you living in the student village would have been a nightmare. **Shezri** always happy, understanding and wise, thank you for the parties and travels and the most, to both you and your parents, for having me as your guest in the beautiful Sri Lanka. **Anna** my only sane housemate, positive and glad, I am grateful you were there for mutual confiding, sharing food and cooking together but also doing fun stuff like tasting different foods, partying and not to forget climbing the Sydney Bridge.

Tim R. thank you for your kindness, for making me laugh at the toughest period in my life and making me realize what really matters in life. I really enjoyed yours and your family's hospitality and company in Hobart.

Ling, my lab supervisor at Macquarie but also life supervisor in Sydney, your help with work matters, as well as the every day matters connected to my new life in Sydney are extremely appreciated. Not to forget all our travels together, I had the great time. You are kind person and good friend and I cannot ever thank you enough for your support.

For all the help, support and good times at work, afterworks and conferences members of my group, former and current: **Johan M**., **Paolo**, **Ufuk**, **Michele**, **Linn**, **Johan**, **Sofia**, **Ola**, **Kristoffer**, **Christoffer**, **Aakash**, **Leena**, **Dani**, **Valentina**, **Elena**, **Marianne**, I thank you.

Special thanks to **Karin**, **Mats** and **Liselotte,** for all the help in the lab and with the cranky instruments during the years.

Thanks to all the supervisors for support and students from the dark side (haha) for fikas, sport activities, afterworks and parties.

Thanks to all people at 404 for making my PhD time more fun: **Minerva**, **Jeanette**, **Anthony**, **Alan**, **Alissa**, **Eliana**, **Eugenia**, **Mischa**, **Matteo**, **Tina** and **Margareta.**

Tommie, **Marianne** and **Cornelia**, thank you for fixing all the administrative work during the years, you made things so much easier for all of us. **Tommie** thanks for always being happy (even when grumpy) and making me laugh. Extra thanks for fixing cakes and candy ☺

Special thanks to **Jonas** and Media Tryck for doing great job and simplifying process of submission.

Anson, **Olof** and **Vignesh** we had some great times cooking, eating and having interesting discussions together Anson you are kind and good listener but always late (got you there!). Olof we will never agree on many things but you are good guy and I enjoyed our discussions and sharing life anecdotes. **Sebastian** I will miss our lunches and discussions on religion and atheism and you know what else.

My friends of many years: **Mahsa** thank you for chats, lunches, dinners and our travels together. I miss you now when you moved to Jönköping. **Shub,** my best Indian friend, for always being there to listen, sharing food and drinks. Not to mention a number of parties we got crazy on. It is extra fun finishing our thesis at the same time. **Aysel,** my Turkish girl that made my days so much more fun during her visit, I miss you. Hope to see you soon in Copenhagen.

Monika and **Pawel** polish party crew, I am grateful to have you as my friends. You helped me transit to the new chapter of my life and were always there for me with tons of laughs, good food and good times. Your presence in Malmö is missed but I'm happy for your new life in Warsaw.

Lotta my best work friend during my time at BMC, still caching up for lunches, dinners and drinks and talking about everything and anything. It is great to have a friend like you.

My university study friends, **Suado**, **Emina**, **Eleonor**, **Frida**, **Kamille**, **Caroline** and **Jennifer** thank you for still being by my side and having good times together.

Sofia N you have been in my life for a short while but stayed by my side when I needed you the most as a true friend. Thank you for always being there for me.

I am grateful to my family, my sisters **Elmina**, **Ajka**, **Ehlimana**, nieces **Danira** and **Amila** and nephews **Rino** and **Roni** for always being by my side. Thanks to the rest of my family in Bosnia for always believing in me.

My sister **Amra,** I am so lucky to have you in my life, it is a treasure to have a sister you can talk about anything with and on top of that she always make you laugh.

Thank you Mum for always being understanding, supportive and cooking my favourite foods.

Hvala Mama na razumijevanju, podrsci i najboljoj hrani.

Huge thanks to my 'adoptive' family Sorensen-Kurbasic. Thank you **Azra** and **Ole** for taking care of me during all these years. My sister **Azra** you are one of the kindest people I know, trying to take care of everybody. Having a freedom to come over for dinners or whatever else I need, makes my life so much easier. **Elvira** and **Nora** you were and are the lights of my life, making me smile and laugh every single time I meet you.

Kind, gentle **Robert** I cannot thank you enough for taking care of me during these last months. You thought me the value of patience and I appreciate your presence and friendship more than you know. I enjoyed (almost) every minute of time I spent with you either at home or away traveling.

References

1. A. OverbergA. HassenbürgerF. Hillenkamp. Laser Desorption Mass Spectrometry. Part II Performance and Applications of Matrix-Assisted Laser Desorption/Ionization of Large Biomolecules. Mass Spectrometry in the Biological Sciences: A Tutorial. 1989; 353:181-197.
2. Abramson VG1, Lehmann BD, Ballinger TJ, Pietenpol JA. Subtyping of triple-negative breast cancer: implications for therapy. Cancer. 2015;121(1):8-16.
3. Ahmedin Jemal DVM, PhD, Freddie Bray PhD, Melissa M. Center MPH, Jacques Ferlay ME, Elizabeth Ward PhD, David Forman PhD. Global cancer statistics. CA: Cancer journal for clinicians. 2011; 61(2):69-90.
4. Alexander, R. P., Fang, G., Rozowsky, J., Snyder, M., and Gerstein, M. B. Annotating non-coding regions of the genome. Nature reviews Genetics 2010; 11, 559-571.
5. Ali S, Coombes RC. Endocrine-responsive breast cancer and strategies for combating resistance. Nat Rev Cancer. 2002; 2(2):101-12.
6. Andersen A, Holte H, Slørdal L. Pharmacokinetics and metabolism of doxorubicin after short-term infusions in lymphoma patients. Cancer Chemother Pharmacol. 1999; 44(5):422-6.
7. Andreyev, A. Y., Shen, Z., Guan, Z., Ryan, A. et al. Application of proteomic marker ensembles to subcellular organelle identification. Mol. Cell. Proteomics 2010; (9) 388–402.
8. Antoniou A, Pharoah PD, Narod S, Risch HA, Eyfjord JE, Hopper JL, Loman N, Olsson H, Johannsson O, Borg A, Pasini B, Radice P, Manoukian S, Eccles DM, Tang N, Olah E, Anton-Culver H, Warner E, Lubinski J, Gronwald J, Gorski B, Tulinius H, Thorlacius S, Eerola H, Nevanlinna H, Syrjäkoski K, Kallioniemi OP, Thompson D, Evans C, Peto J, Lalloo F, Evans DG, Easton DF. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. Am J Hum Genet. 2003; 72(5):1117-30
9. Apweiler, R., H. Hermjakob, and N. Sharon. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. Biochimica et Biophysica Acta. 1999; 1473(1):4-8.
10. Aravind Subramaniana,b, Pablo Tamayoa,b, Vamsi K. Moothaa,c, Sayan Mukherjeed, Benjamin L. Eberta,e, Michael A. Gillettea,f, Amanda Paulovichg, Scott L. Pomeroyh, Todd R. Goluba,e, Eric S. Landera,c,i,j,k, and Jill P. Mesirova, K. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. PNAS. 2005; 102(43): 15545–15550.

11. Ashburner, M., et al. Gene Ontology: tool for the unification of biology. Nat Genet. 2000; 25(1):25-29.
12. Bairoch, A. and R. Apweiler, The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res. 2000; 28(1):45-8.
13. Balatti, V., Pekarky, Y., and Croce, C. M. Role of microRNA in chronic lymphocytic leukemia onset and progression. Journal of hematology & oncology. 2015; 8, 12.
14. Banerjee, S. and S. Mazumdar. Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte. International Journal of Analytical Chemistry. 2012; 2012:40.
15. Banyard J, Chung I, Wilson AM, Vetter G, Le Béchec A, Bielenberg DR, Zetter BR. Regulation of epithelial plasticity by miR-424 and miR-200 in a new prostate cancer metastasis model. Sci Rep. 2013; 3:3151.
16. Bause, E., Structural requirements of N-glycosylation of proteins. Studies with proline peptides as conformational probes. Biochem. J. 1983; 209(2):331-336.
17. Ben-Baruch, A. Organ selectivity in metastasis: regulation by chemokines and their receptors. Clin Exp Metastasis 2007.
18. BenitoCañasa, CarmenPiñeirob, EnriqueCalvoc, DanielLópez-Ferrerd, Jose ManuelGallardob. Trends in sample preparation for classical and second-generation proteomics. Journal of Chromatography A. 2007; 1153(1–2):235-258.
19. Biemann K. Contributions of mass spectrometry to peptide and protein structure. Biomed Environ Mass Spectrom. 1988;16(1-12):99-111.
20. Bjorhall, K., T. Miliotis, and P. Davidsson. Comparison of different depletion strategies for improved resolution in proteomic analysis of human serum samples. Proteomics. 2005; 5(1):307-17.
21. Boisvert FM, Lamond AI. p53-Dependent subcellular proteome localization following DNA damage. Proteomics. 2010;10(22):4087-97.
22. Bonci, D., Coppola, V., Patrizii, M., Addario, A., Cannistraci, A., Francescangeli, F., Pecci, R., Muto, G., Collura, D., Bedini, R., et al. A microRNA code for prostate cancer metastasis. Oncogene 2016; 35, 1180-1192.
23. Bottai G, Pasculli B, Calin GA, Santarpia L. Targeting the microRNA-regulating DNA damage/repair pathways in cancer. Expert Opin Biol Ther. 2014; 14(11):1667-83
24. Bränden, C.I. and Tooze, J. Introduction to protein structure. 1999; New York, Garland Pub.
25. Callister, S.J., et al. Normalization Approaches for Removing Systematic Biases Associated with Mass Spectrometry and Label-Free Proteomics. Journal of Proteome Research. 2006; 5(2):277-286.
26. Campeau PM, Foulkes WD, Tischkowitz MD. Hereditary breast cancer: New genetic developments, new therapeutic avenues. Human Genetics 2008; 124(1):31–42.
27. Caroline F. Thorn, Connie Oshiro, Sharon Marsh, Tina Hernandez-Boussard, Howard McLeod, Teri E. Klein,aand Russ B. Altman. Doxorubicin pathways: pharmacodynamics and adverse effects. Pharmacogenet Genomics. 2011; 21(7): 440–446.

28. Chanchal Kumar, Matthias Mann. Bioinformatics analysis of mass spectrometry-based proteomics data sets. FEBS Letters. 2009; 583:1703–1712.
29. Chelius, D. and P.V. Bondarenko. Quantitative Profiling of Proteins in Complex Mixtures Using Liquid Chromatography and Mass Spectrometry. Journal of Proteome Research. 2002; 1(4):317-323.
30. Christiansen, M.N., et al., Cell surface protein glycosylation in cancer. Proteomics. 2013.
31. Claudia Rivera-Guevara, Javier Camacho. Tamoxifen and its New Derivatives in Cancer Research. Recent Patents on Anti-Cancer Drug Discovery. 2011; 6 (2):237 – 245.
32. Compton. P.D. and N.L. Kelleher.Spinning up mass spectrometry for whole protein complexes. Nat Meth. 2012; 9(11):1065-1066.
33. Coughenour, HD. Spaulding, RS. Thompson, CM. The synaptic vesicle proteome: a comparative study in membrane protein identification. Proteomics. 2004; 4(10):3141-55.
34. Cox, J., et al., A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. Nat Protoc. 2009; 4(5):698-705.
35. Craig, R. and R.C. Beavis, TANDEM: matching proteins with tandem mass spectra.
36. Cramer R, Corless S. The nature of collision-induced dissociation processes of doubly protonated peptides: comparative study for the future use of matrix-assisted laser desorption/ionization on a hybrid quadrupole time-of-flight mass spectrometer in proteomics. Rapid Commun Mass Spectrom. 2001;1 5(22):2058-66.
37. Cristian Tomasetti, Lu Li, Bert Vogelstein. Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. Science 2017;(355) 6331:1330-1334
38. De Bont R, van Larebeke N. Endogenous DNA damage in humans: a review of quantitative data. Mutagenesis. 2004; 19(3):169-85.
39. Deutsch, E.W., et al., A guided tour of the Trans-Proteomic Pipeline. Proteomics. 2010; 10(6):1150-9.
40. Doerr, A. Targeted proteomics. Nat Meth. 2010; 7(1):34-34.
41. Donald A. Berry, Ph.D., Kathleen A. Cronin, Ph.D., Sylvia K. Plevritis, Ph.D., Dennis G. Fryback, Ph.D., Lauren Clarke, M.S., Marvin Zelen, Ph.D., Jeanne S. Mandelblatt, Ph.D., Andrei Y. Yakovlev, Ph.D., J. Dik F. Habbema, Ph.D., and Eric J. Feuer, Ph.D. Effect of Screening and Adjuvant Therapy on Mortality from Breast Cancer for the Cancer Intervention and Surveillance Modeling Network (CISNET) Collaborators. N Engl J Med 2005; 353:1784-1792.
42. Elias, J.E., et al., Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. Nat Meth. 2005; 2(9):667-675.
43. Elofsson A, von Heijne G. Membrane protein structure: prediction versus reality. Annu Rev Biochem. 2007; 76:125-40.
44. Elston, C. W., and Ellis I. O. Pathological prognostic factors in the breast cancer. I. The value of histological grade in breast cancer: Experience from large study with long term follow-up. Histopathology 1991;19:403-410.

45. Emad A Rakha, Jorge S Reis-Filho, Frederick Baehner, David J Dabbs, Thomas Decker, Vincenzo Eusebi, Stephen B Fox, Shu Ichihara, Jocelyne Jacquemier, Sunil R Lakhani, José Palacios, Andrea L Richardson, Stuart J Schnitt, Fernando C Schmitt, Puay-Hoon Tan, Gary M Tse, Sunil Badve, and Ian O Ellis. Breast cancer prognostic classification in the molecular era: the role of histological grade. Breast Cancer Res. 2010; 12(4): 207.
46. Eng, J., A. McCormack, and J. Yates, An approach to correlate tandem mass spectral data ofpeptides with amino acid sequences in a protein database. Journal of the American Society for Mass Spectrometry.1994; 5(11):976-989.
47. Engstrøm, M.J., Opdahl, S., Hagen, A.I. et al. Molecular subtypes, histopathological grade and survival in a historic cohort of breast cancer patients Breast Cancer Res Treat. 2013; 140:463.
48. Esplen MJ1, Hunter J, Leszcz M, Warner E, Narod S, Metcalfe K, Glendon G, Butler K, Liede A, Young MA, Kieffer S, DiProspero L, Irwin E, Wong J.A multicenter study of supportive-expressive group therapy for women with BRCA1/BRCA2 mutations. Cancer. 2004; 101(10):2327-40.
49. Esteller, M. Non-coding RNAs in human disease. Nature reviews Genetics. 2010; 12, 861-874.
50. Falck, A.K., et al., St Gallen molecular subtypes in primary breast cancer and matched lymph node metastases--aspects on distribution and prognosis for patients with luminal A tumours: results from a prospective randomised trial. BMC Cancer, 2013; 13: 558.
51. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. Science. 1989 Oct 6;246(4926):64-71. Review.
52. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, Rosso S, Coebergh JWW, Comber H, Forman D, Bray F. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. Eur J Cancer. 2013; 49(6):1374-403.
53. Fidler, I. J. The organ microenvironment and cancer metastasis. Differentiation 2002; 70:498-505
54. Filipowicz, W., Bhattacharyya, S. N., and Sonenberg, N. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nature reviews Genetics. 2008; 9, 102-114.
55. Forrest RA, Swift LP, Rephaeli A, Nudelman A, Kimura K, Phillips DR, Cutts SM. Activation of DNA damage response pathways as a consequence of anthracycline-DNA adduct formation. Biochem Pharmacol. 2012; 83(12):1602-12.
56. Foster LJ, de Hoog CL, Zhang Y, Zhang Y, Xie X, Mootha VK, Mann M. A mammalian organelle map by protein correlation profiling. Cell. 2006; 125(1):187-99.
57. Foster, L. J., de Hoog, C. L., Zhang, Y., Xie, X. et al., A mammalian organelle map by protein correlation profiling. Cell 2006; 125:187–199.
58. Garcia AI, Buisson M, Bertrand P, Rimokh R, Rouleau E, Lopez BS, Lidereau R, Mikaélian I, Mazoyer S. Down-regulation of BRCA1 expression by miR-146a and miR-146b-5p in triple negative sporadic breast cancers. EMBO Mol Med. 2011; 3(5):279-90.

59. Gatto L., Vizcaı´no JA., Hermjakob H., Huber W. and Lilley KS. Organelle proteomics experimental designs and analysis. Proteomics. 2010; 10:3957–3969.
60. Goldstein M, Kastan MB. The DNA damage response: implications for tumor responses to radiation and chemotherapy. Annu Rev Med. 2015; 66:129-43.
61. Graham, JM. Fractionation of Subcellular Organelles Current Protocols in Cell Biology. 2015.
62. Gravgaard KH, Lyng MB, Laenkholm AV, Søkilde R, Nielsen BS, Litman T, Ditzel HJ. The miRNA-200 family and miRNA-9 exhibit differential expression in primary versus corresponding metastatic tissue in breast cancer. Breast Cancer Res Treat. 2012; 134(1):207-17.
63. Hall, S. L., Hester, S., Griffin, J. L., Lilley, K. S., Jackson, A. P. The organelle proteome of the DT40 lymphocyte cell line. Mol. Cell. Proteomics 2009; (8) 1295–1305.
64. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011; 144(5):646-74.
65. Hanahan, D., and Weinberg, R. A. Hallmarks of cancer: the next generation. 2011; Cell 144, 646-674.
66. He, L., Thomson, J. M., Hemann, M. T., Hernando-Monge, E., Mu, D., Goodson, S., Powers, S., Cordon-Cardo, C., Lowe, S. W., Hannon, G. J., and Hammond, S. M. A microRNA polycistron as a potential human oncogene. Nature 2005; 435, 828-833.
67. Ho, C.S., et al. Electrospray ionisation mass spectrometry: principles and clinical applications. Clin Biochem Rev, 2003. 24(1):3-12.
68. Hoeijmakers JH. DNA damage, aging, and cancer. N Engl J Med. 2009;361(15):1475-85.
69. Hondermarck, H., Breast cancer: when proteomics challenges biological complexity. Mol Cell Proteomics. 2003; 2(5):281-91.
70. Hu, Z.-Z., et al. Proteomic Analysis of Pathways Involved in Estrogen-Induced Growth and Apoptosis of Breast Cancer Cells. PLoS One. 2011; 6(6):204-10.
71. Huh, W. K., Falvo, J. V., Gerke, L. C., Carroll, A. S. et al., Global analysis of protein localization in budding yeast.Nature 2003; 425, 686–691.
72. Ignatiadis, M., Xenidis, N., Perraki, M., Apostolaki, S., Politaki, E., Kafousi, M., et al. Different Prognostic Value of Cytokeratin-19 mRNA Positive Circulating Tumor Cells According to Estrogen Receptor and HER2 Status in Early-Stage Breast Cancer. Journal of Clinical Oncology. 2007; 25, 5194-5202.
73. Ingenuity Pathway Analysis (IPA) (Ingenuity System) http://www.ingenuity.com
74. Jabbari N, Reavis AN, McDonald JF. Sequence variation among members of the miR-200 microRNA family is correlated with variation in the ability to induce hallmarks of mesenchymal-epithelial transition in ovarian cancer cells. J Ovarian Res. 2014; 7:12.
75. Jeggo PA, Pearl LH, Carr AM. DNA repair, genome stability and cancer: a historical perspective. Nat Rev Cancer. 2016;16(1):35-42.
76. Ji Na Kong, Qian He, Guanghu Wang, Somsankar Dasgupta, Michael B. Dinkins, Gu Zhu, Austin Kim, Stefka Spassieva, and Erhard Bieberich. Guggulsterone and bexarotene induce secretion of exosome-associated breast cancer resistance protein

and reduce doxorubicin resistance in MDA-MB-231 cells. Int J Cancer. 2015; 137(7): 1610–1620.

77. Jianqing Fan, Fang Han, and Han Liu. Challenges of Big Data Analysis. Natl Sci Rev. 2014; 1(2): 293–314.
78. Johnson, S. M., Grosshans, H., Shingara, J., Byrom, M., Jarvis, R., Cheng, A., Labourier, E., Reinert, K. L., Brown, D., and Slack, F. J. RAS is regulated by the let-7 microRNA family. Cell 2005; 120, 635-647.
79. Kavanagh DM, Powell WE, Malik P, Lazou V, Schirmer EC. Organelle proteome variation among different cell types: lessons from nuclear membrane proteins. Subcell Biochem. 2007; 43:51-76.
80. Kevin Lawlor, Arpi Nazarian, Lynne Lacomis, Paul Tempst, and Josep Villanueva. Pathway-Based Biomarker Search by High-Throughput Proteomics Profiling of Secretomes. Journal of Proteom Research. 2009; 8, 1489–1503.
81. Khoury, G. A., Baliban, R.C., and Floudas, C. A. Proteome-wide post-translational modification statistics: frequency analysis and curation of the Swiss-prot database. Sci Rep. 2011; 1.
82. Kim MS, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, Madugundu AK, Kelkar DS, Isserlin R, Jain S, Thomas JK, Muthusamy B, Leal-Rojas P, Kumar P, Sahasrabuddhe NA, Balakrishnan L, Advani J, George B, Renuse S, Selvan LD, Patil AH, Nanjappa V, Radhakrishnan A, Prasad S, Subbannayya T, Raju R, Kumar M, Sreenivasamurthy SK, Marimuthu A, Sathe GJ, Chavan S, Datta KK, Subbannayya Y, Sahu A, Yelamanchi SD, Jayaram S, Rajagopalan P, Sharma J, Murthy KR, Syed N, Goel R, Khan AA, Ahmad S, Dey G, Mudgal K, Chatterjee A, Huang TC, Zhong J, Wu X, Shaw PG, Freed D, Zahari MS, Mukherjee KK, Shankar S, Mahadevan A, Lam H, Mitchell CJ, Shankar SK, Satishchandra P, Schroeder JT, Sirdeshmukh R, Maitra A, Leach SD, Drake CG, Halushka MK, Prasad TS, Hruban RH, Kerr CL, Bader GD, Iacobuzio-Donahue CA, Gowda H, Pandey A.A draft map of the human proteome. Nature. 2014;509(7502):575-81
83. Kim S, Hwang DW, Lee DS. A study of microRNAs in silico and in vivo: bioimaging of microRNA biogenesis and regulation. FEBS J. 2009; 276(8):2165-74.
84. Kirik U, Cifani P, Albrekt AS, Lindstedt M, Heyden A, Levander F. Multimodel pathway enrichment methods for functional evaluation of expression regulation. J Proteome Res. 2012; 11(5):2955-67.
85. Klose J. Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. Humangenetik. 1975; 26(3):231-43.
86. Korpal M, Kang Y. The emerging role of miR-200 family of microRNAs in epithelial-mesenchymal transition and cancer metastasis. RNA Biol. 2008; 5(3):115-9.
87. Kyle E. Freese, Lauren Kokai, Robert P. Edwards, Brian J. Philips, M. Aamir Sheikh, Joseph Kelley, John Comerci, Kacey G. Marra, J. Peter Rubin and Faina Linkov. Adipose-Derived Stems Cells and Their Role in Human Cancer Development, Growth, Progression, and Metastasis: A Systematic Review. Cancer Research 2015; 75(7).

88. Lai, T.-C., et al. Secretomic and Proteomic Analysis of Potential Breast Cancer Markers by Two-Dimensional Differential Gel Electrophoresis. Journal of Proteome Research. 2010; 9(3):1302-1322.
89. Lai, X., L. Wang, and F.A. Witzmann. Issues and Applications in Label-Free Quantitative Mass Spectrometry. International Journal of Proteomics. 2013; 2013:13.
90. Lam, H., et al., Development and validation of a spectral library searching method for peptide identification from MS/MS. Proteomics. 2007; 7(5):655-67.
91. Lange, V. Picotti, P. Domon, B. Aebersold, R. Selected reaction monitoring for quantitative proteomics: a tutorial. Molecular Systems Biology. 2008; 4:222
92. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell. 2005;120(1):15-20.
93. Li, J., et al., A Targeted Proteomics Approach for Biomarker Discovery Using Bilateral Matched Nipple Aspiration Fluids. Clinical Proteomics, 2010; 6(3):57-64.
94. Lin SH, Guidotti G. Purification of membrane proteins. Methods Enzymol. 2009; 463:619-29.
95. Liotta, L. A., Kleinerman, J. and Saidel G. M. Quantitative relationships of intravascular tumor cells, tumor vessels and pulmonary metastases following tumor implantation. Cancer Res 1974; 57:32-35.
96. Liu, H., R.G. Sadygov, and J.R. Yates, 3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal Chem, 2004; 76(14):4193-201.
97. Liyanage, R. and J.O. Lay. An Introduction to MALDI-TOF MS, in Identification of Microorganisms by Mass Spectrometry. John Wiley & Sons, Inc. 2006; p. 39-60.
98. Macher BA, Yen TY. Proteins at membrane surfaces-a review of approaches. Mol Biosyst. 2007; 3(10):705-13.
99. Madjd, Z., et al. High expression of Lewisy/b antigens is associated with decreased survival in lymph node negative breast carcinomas. Breast Cancer Research. 2005; 7(5):780-787.
100. Mark Barok, Heikki Joensuu and Jorma Isola. Trastuzumab emtansine: mechanisms of action and drug resistance. Breast Cancer Research 2014; 16:209.
101. Maureen E. taylor and Kurt Drickamer. Introduction to Glycobiology. Oxford university press, 2006. ISBN:0-19-928278-1.
102. Menon, R. and G.S. Omenn. Cancer Proteomic characterization of novel alternative splice variant proteins in human epidermal growth factor receptor 2/neu-induced breast cancers. Res. 2010; 70(9):3440-9.
103. Michelsen., U. and Hagen, J. Isolation of Subcellular Organelles and Structures. Methods in Enzymology. 2009; 463:305-328.
104. Mikesh LM, Ueberheide B, Chi A, Coon JJ, Syka JE, Shabanowitz J, Hunt DF. The utility of ETD mass spectrometry in proteomic analysis. Biochim Biophys Acta. 2006; 1764(12):1811-22.
105. Moremen, K.W., M. Tiemeyer, and A.V. Nairn. Vertebrate protein glycosylation: diversity, synthesis and function. Nat Rev Mol Cell Biol. 2012; 13(7): 448-462.

106. Mulvey CM, Tudzarova S, Crawford M, Williams GH, Stoeber K, Godovac-Zimmermann J. Bioinformatics. 2004; 20(9): 1466-7.

107. Nesvizhskii, A.I. and R. Aebersold, Interpretation of shotgun proteomic data: the protein inference problem. Mol Cell Proteomics. 2005; 4(10):1419-40.

108. Nikolsky Y1, Ekins S, Nikolskaya T, Bugrim A. A novel method for generation of signature networks as biomarkers from complex high throughput data. Toxicol Lett. 2005; 158(1):20-9.

109. Palmblad M, Tsybin YO, Ramström M, Bergquist J, Håkansson P. Liquid chromatography and electron-capture dissociation in Fourier transform ion cyclotron resonance mass spectrometry. Rapid Commun Mass Spectrom. 2002; 16(10):988-92.

110. Parker, J.S., et al., Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes. Journal of Clinical Oncology, 2009. 27(8):1160-1167.

111. Pasquali C1, Fialka I, Huber LA. Subcellular fractionation, electromigration analysis and mapping of organelles. J Chromatogr B Biomed Sci Appl. 1999; 722(1-2):89-102.

112. Patel, V.J., et al. Comparison of Labeling and Label-Free Mass Spectrometry-Based Proteomics Approaches. A Journal of Proteome Research. 2009; 8(7):752-3759.

113. Perkins, D.N., et al., Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis. 1999; 20(18):3551-67.

114. Perou, C., et al., Molecular portraits of human breast tumours. Nature, 2000. 406(6797):747-752.

115. Phanstiel D, Zhang Y, Marto JA, Coon JJ. Peptide and protein quantification using iTRAQ with electron transfer dissociation. J Am Soc Mass Spectrom. 2008; 19(9):1255-62.

116. Picotti, P. Aebersold, R. Selected reaction monitoring–based proteomics: workflows, potential, pitfalls and future directions. Nature Methods. 2012; 9(6):555–566.

117. Polyak, K. and R. Kalluri, The role of the microenvironment in mammary gland development andcancer. Cold Spring Harb Perspect Biol, 2010; 2(11).

118. Ponomarenko, E. A., et al. The size of the human proteome: The width and Depth. Int J anal Chem. 2016;

119. Pothof J, Verkaik NS, Hoeijmakers JH, van Gent DC. MicroRNA responses and stress granule formation modulate the DNA damage response. Cell Cycle. 2009; 8(21):3462-8.

120. Pullen, T. J., da Silva Xavier, G., Kelsey, G., and Rutter, G. A. miR-29a and miR-29b contribute to pancreatic beta-cell-specific silencing of monocarboxylate transporter 1 (Mct1). Molecular and cellular biology 2011; 31, 3182-3194.

121. Qian, W.J., et al. Advances and challenges in liquid chromatography-mass spectrometry-based proteomics profiling for clinical applications. Mol Cell Proteomics. 2006; 5(10):1727-44.

122. Reczko M., Maragkakis M., Alexiou P., Grosse I., Hatzigeorgiou A.G. Functional microRNA targets in protein coding sequences. Bioinformatics. 2012;28:771–776.

123. Robert A. Weinberg, Robert A Weinberg. The Biology of Cancer. 2007; ISBN: 9780815342205.

124. Roepstorff P, Fohlman J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. Biomed Mass Spectrom. 1984;11(11):601.

125. Rohan RM, Fernandez A, Udagawa T, Yuan J, D'Amato RJ. Genetic heterogeneity of angiogenesis in mice. FASEB J. 2000;14(7):871-6.

126. Russnes HG, Navin N, Hicks J, Borresen-Dale AL. Insight into the heterogeneity of breast cancer through next-generation sequencing. J Clin Invest. 2011; 121(10):3810-8.

127. Sadowski PG, Groen AJ, Dupree P, Lilley KS. Sub-cellular localization of membrane proteins. Proteomics. 2008;8(19):3991-4011.

128. Sakamoto, J., et al. Expression of Lewisa, Lewisb, X, and Y Blood Group Antigens in Human Colonic Tumours and Normal Tissue and in Human Tumour-derived Cell Lines. Cancer Research. 1986; 46(3): 1553-1561.

129. Sandin, M. Teleman, J. Malmström, J. Levander, F. Data processing methods and quality control strategies for label-free LC–MS protein quantification. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics. 2014; 1884(1): 29-41.

130. Sandra S. McAllister & Robert A. Weinberg. The tumour-induced systemic environment as a critical regulator of cancer progression and metastasis. Nature Cell Biology. 2014; (16): 717–727 .

131. Savage KI, Harkin DP. BRCA1, a 'complex' protein involved in the maintenance of genomic stability. FEBS J. 2015; 282(4):630-46.

132. Scherp, P. Ku G. Coleman, L. Kheterpa, I. Gel-based and gel-free proteomic technologies. Methods Mol Biol. 2011; 702:163-90.

133. Schluesener D, Fischer F, Kruip J, Rögner M, Poetsch A. Mapping the membrane proteome of Corynebacterium glutamicum. Proteomics. 2005; 5(5):1317-30.

134. Scott TL, Rangaswamy S, Wicker CA, Izumi T. Repair of oxidative DNA damage and cancer: recent progress in DNA base excision repair. Antioxid Redox Signal. 2014; 20(4):708-26.

135. Serena Nik-Zainal, Helen Davies, Johan Staaf, Manasa Ramakrishna, Dominik Glodzik, Xueqing Zou, Inigo Martincorena, Ludmil B. Alexandrov, Sancha Martin, David C. Wedge, Peter Van Loo, Young Seok Ju, Marcel Smid, Arie B. Brinkman, Sandro Morganella, Miriam R. Aure, Ole Christian Lingjærde, Anita Langerød, Markus Ringnér, Sung-Min Ahn, Sandrine Boyault, Jane E. Brock, Annegien Broeks, Adam Butler, Christine Desmedt et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature 2016; 534, 47–54.

136. Serpico D, Molino L, Di Cosimo S. microRNAs in breast cancer development and treatment. Cancer Treat Rev. 2014; 40(5):595-604.

137. Shazia Ali, Mahmood Rasool, Hani Chaoudhry, Peter N Pushparaj, Prakash Jha, Abdul Hafiz, Maryam Mahfooz, Ghufrana Abdus Sami, Mohammad Azhar Kamal, Sania Bashir, Ashraf Ali and Mohammad Sarwar Jamal. Molecular mechanisms and mode of tamoxifen resistance in breast cancer. Biomedical Informatics. 2016; 12(3):135-139.

138. Shuibin Lin and Richard I. Gregory. MicroRNA biogenesis pathways in cancer. Nature Reviews Cancer 2015; 15:321–333.

139. Silva ML. Cancer serum biomarkers based on aberrant post-translational modifications of glycoproteins: Clinical value and discovery strategies. Biochim Biophys Acta. 2015.

140. Simpson , R.J. Proteins and Proteomics: A Laboratory Manual. 2003 Cold Spring Harbor Laboratory Press.

141. Simpson, J. C., Pepperkok, R., The subcellular localization of the mammalian proteome comes a fraction closer. Genome Biol. 2006; (7) 222.

142. Sobin, L. H. and Wittekind, C. TNM classification of malignant tumours (UICC). New York: Willey-Liss. 2002.

143. Sørlie, T., et al., Gene expression patterns of breast carcinomas distinguishes tumor subclasses with clinical implications. Proceedings of the National Academy of Sciences, 2001; 98(19):10869-10874.

144. Sørlie, T., et al., Repeated observation of breast tumor subtypes in independent gene expression data sets. Proceedings of the National Academy of Sciences, 2003. 100(14):8418-8423.

145. Spiro, R.G., Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. Glycobiology. 2002; 12(4):43R-56R.

146. Steven A. Narod & William D. Foulkes. BRCA1 and BRCA2: 1994 and beyond. Nature Reviews Cancer 2004; 4:665-676

147. Stevens, T. J., Arkin, I. T., Do more complex organisms have a greater proportion of membrane proteins in their genomes? Proteins 2000; 39:417–420.

148. Subbiah IM, Gonzalez-Angulo AM. Advances and future directions in the targeting of HER2-positive breast cancer: implications for the future. Curr Treat Options Oncol. 2014; 15(1):41-54.

149. Subcellular proteomics reveals a role for nucleo-cytoplasmic trafficking at the DNA replication origin activation checkpoint. J Proteome Res. 2013; 12(3):1436-53.

150. Tacar O1, Sriamornsak P, Dass CR. Doxorubicin: an update on anticancer molecular action, toxicity and novel drug delivery systems. J Pharm Pharmacol. 2013; 65(2):157-70.

151. Takahashi, N. and H. Nishibe, Some characteristics of a new glycopeptidase acting on aspartylglycosylamine linkages. J Biochem. 1978; 84(6): 1467-73.

152. Tan S., Tan HT. and Chung M.C.M. Membrane proteins and membrane proteomics. Proteomics. 2008.

153. Templin T, Paul S, Amundson SA, Young EF, Barker CA, Wolden SL, Smilenov LB. Radiation-induced micro-RNA expression changes in peripheral blood cells of radiotherapy patients. Int J Radiat Oncol Biol Phys. 2011; 80(2):549-57.

154. Temporini C, Calleri E, Massolini G, Caccialanza G. Integrated analytical strategies for the study of phosphorylation and glycosylation in proteins. Mass Spectrom Rev. 2008; 27(3):207-36.

155. ThuyVu and Francois X. Claret. Trastuzumab: updated mechanisms of action and resistance in breast cancer. Frontiers in Oncology. 2012; 2 (62):1-6.

156. Turnbull, J.E. and R.A. Field, Emerging glycomics technologies. Nat Chem Biol. 2007; 3(2):74-77.

157. Unlü M, Morgan ME, Minden JS. Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. Electrophoresis. 1997; 18(11):2071-7.

158. Uwe Michelsen, Jörg von Hagen. Guide to Protein Purification, 2nd Edition. Chapter 19: Isolation of Subcellular Organelles and Structures. 2009; 463:305–328

159. Venter JC, Smith HO, Adams MD. The Sequence of the Human Genome. Clin Chem. 2015; 61(9):1207-8.

160. Viale, G.The current state of breast cancer classification. Annals of Oncology, 2012; 23(10):207-x210.

161. Volinia S, Calin GA, Liu CG, Ambs S, Cimmino A, Petrocca F, Visone R, Iorio M, Roldo C, Ferracin M, Prueitt RL, Yanaihara N, Lanza G, Scarpa A, Vecchione A, Negrini M, Harris CC, Croce CM. A microRNA expression signature of human solid tumors defines cancer gene targets. Proc Natl Acad Sci U S A. 2006; 103(7):2257-61.

162. Westholm, J. O., and Lai, E. C. Mirtrons: microRNA biogenesis via splicing. Biochimie 2011; 93, 1897-1904.

163. Wheelock AM. and Goto S. Effects of post-electrophoretic analysis on variance in gel-based proteomics. Expert Rev Proteomics. 2006; 3(1):129-42.

164. Wilkins, M.R., et al. Progress with Proteome Projects: Why all Proteins Expressed by a Genome Should be Identified and How To Do It. Biotechnology and Genetic Engineering Reviews. 1996; 13(1):19-50.

165. Wouters MD, van Gent DC, Hoeijmakers JH, Pothof J. MicroRNAs, the DNA damage response and cancer. Mutat Res. 2011; 717(1-2):54-66.

166. Wu X, Hasan MA, Chen JY. Pathway and network analysis in proteomics. J Theor Biol. 2014; 362:44-52.

167. Wu, C. C., Yates, J. R. The application of mass spectrometry to membrane proteomics. Nat. Biotechnol. 2003; 21:262–267.

168. Wu, G., X. Feng, and L. Stein. A human functional protein interaction network and its application to cancer data analysis. Genome Biology. 2010; 11(5):53.

169. Yao K, Goldschmidt R, Turk M, Wesseling J, Stork-Sloots L, de Snoo F, Cristofanilli M. Molecular subtyping improves diagnostic stratification of patients with primary breast cancer into prognostically defined risk groups. Breast Cancer Res Treat. 2015;154(1):81-8.

170. Yates, J.R Mass. Spectral Analysis In Proteomics. Annual Review of Biophysics and Biomolecular Structure, 2004; 33(1):297-316.

171. Zhang Z, Liu ZB, Ren WM, Ye XG, Zhang YY. The miR-200 family regulates the epithelial-mesenchymal transition induced by EGF/EGFR in anaplastic thyroid cancer cells. Int J Mol Med. 2012; 30(4):856-62.

Printed by Media-Tryck, Lund 2017 NORDIC ECOLABEL, 3041 0903

Lund University
Faculty of Engineering
Department of Immunotechnology
ISBN 978-91-7753-393-1