



University of Bradford eThesis

This thesis is hosted in [Bradford Scholars](#) – The University of Bradford Open Access repository. Visit the repository for full metadata or to contact the repository team



© University of Bradford. This work is licenced for reuse under a [Creative Commons Licence](#).

ONCOPROTEOMIC APPLICATIONS FOR
DETECTION OF BREAST CANCER

S. SHAHEED

PhD

UNIVERSITY OF BRADFORD

2017

ONCOPROTEOMIC APPLICATIONS FOR DETECTION OF BREAST
CANCER

Proteomic profiling of breast cancer models and biopsies

Sadr-ul SHAHEED

Submitted for the Degree of
Doctor of Philosophy

Faculty of Life Sciences
University of Bradford

2017

Abstract

Sadr-ul Shaheed

Title: ONCOPROTEOMIC APPLICATIONS FOR DETECTION OF BREAST CANCER

Sub-title: Proteomic profiling of breast cancer models and biopsies.

Keywords: Breast cancer, biomarkers, proteomics, bio-fluids, Nipple Aspirate Fluid, Luminal, Basal, biopsies.

The heterogeneity of breast cancer (disease stage and phenotype) makes it challenging to differentiate between each subtype; luminal A, luminal B, HER2, basal-like and claudin-low, on the basis of a single gene or protein. Therefore, a collection of markers is required that can serve as a signature for diagnosing different types of breast cancer. New developments in proteomics have provided the opportunity to look at phenotype-specific breast cancer cell lines and stage-specific liquid biopsies (nipple aspirate fluid [NAF], plasma samples) to identify disease and phenotype specific signature.

An 8-plex iTRAQ quantification strategy was employed to compare proteomic profiles of a range of breast cancer and 'normal-like' cell lines with primary breast epithelial cells. From this, 2467 proteins were identified on Orbitrap Fusion and Ultraflex II, of which 1430 were common. Matched pairs of NAF samples from four patients with different stages of breast cancer, were analysed by SCX-LC-MS and a total of 1990 unique gene products were identified. More than double the number of proteins previously published data, were detected in NAF, including 300 not detected in plasma. The NAF from the diseased patients have 138 potential phenotype biomarkers that were significantly changed compared to the healthy volunteer (7 for luminal A, 9 for luminal B, 11 for HER2, 14 for basal-like and 52 for claudin-low type). The average coefficient of variation for triplicate analyses by multiple reaction monitoring mass spectrometry (MRM-MS), was 9% in cell lines, 17 % in tissue biopsies, 22% in serum samples and 24% in NAF samples.

Overall, the results provide a strong paradigm to develop a clinical assay based on proteomic changes in NAF samples for the early detection of breast cancer supplementary to established mammography programmes.

Dedications

This thesis is dedicated to the memory of my wonderful father, Dr Ahmad Khan Miana. He was a constant source of support and encouragement and had made an untold number of sacrifices for the entire family. You were well-loved everywhere you served. He was a great inspiration to me. As you look down from heaven, I hope you're proud of your son.

In laughter and in sorrow,

In sunshine and through rain,

I know you are by my side

Acknowledgements

First and foremost, I would like to express my gratitude to my supervisors, Dr Chris Sutton and Prof Laurence Patterson for their support, patience, and encouragement throughout my research project. I am also extremely grateful for their rigorous proofreading of this thesis.

My thanks also go to Professor Kyriacos Kyriacou, for providing tissues biopsy and serum samples, and unlimited support and access to MS laboratories at The Cyprus Institute of Neurology & Genetics (CING). Also, I would like to thank Dr Kleitos Sokratous for his help during my work and stay at CING. I would like to thank Ethical Tissue Bank at University of Bradford, and doctors and nurses at Bradford Teaching Hospital Trust, for collection of NAF samples.

I know that I could not have completed this project without the support of my family. They receive my deepest gratitude and love for their dedication and the many years of support during my studies that provided the foundation for this work. Their support and encouragement was in the end what made this dissertation possible.

Last, but not least, I would like to thank all patients and volunteers, who participated in this study.

Publications and Contributions

Scientific journals:

2017

- Sadr-ul Shaheed, Catherine Tait, Kyriacos Kyriacou, Joanne Mullarkey, Wayne Burrill, Laurence H. Patterson, Richard Linforth, Mohamed Salhab, Chris Sutton, "*Nipple aspirate fluid - a liquid biopsy for diagnosing breast health*", *PROTEOMICS-Clinical Applications* 2017, 1862-8354. (Impact factor: 2.959).

2016

- Sellars, Jonathan D., Mark Skipsey, Sadr-ul Shaheed, Hamza Abumansour, Klaus Pors, Laurence H. Patterson, and Chris W. Sutton. "*Rational Development of Novel Activity Probes for the Analysis of Human Cytochromes P450.*" *ChemMedChem* 11, no. 11 (2016): 1122-1128. (Impact factor: 2.968).

2013

- Sadr-ul Shaheed, Nitin Rustogi, Andrew Scally, Julie Wilson, Helene Thygesen, Maria A Loizidou, Andreas Hadjisavvas, Andrew Hanby, Valerie Speirs, Paul Loadman, Chris W. Sutton, "*Identification of Stage-Specific Breast Markers Using Quantitative Proteomics*", *Journal of proteome research*, 12 (2013), 5696-708. (Impact factor: 5.001).

Scientific meetings and posters:

2016

- Mario. Ortega, Sadr-ul Shaheed, Chris W. Sutton, Steve D. Shnyder "*Development of drug-resistant colorectal cancer cell lines to investigate mechanisms of resistance and identify predictive biomarkers of drug response*, NCRI Cancer Conference Liverpool 6-9 November-2016.
- Chris W. Sutton, Sadr-ul Shaheed, Laurence H. Patterson, Catherine Tait, Rick Linforth, Mohamed Salhab, "*Proteomics of Nipple Aspirate Fluid*", BSPR meeting 2016: Proteomic Approaches to Health and Disease, Glasgow, United Kingdom, 25-27 July 2016.

2015

- Sadr-ul Shaheed, Paul. M. Loadman, Kleitos Sokratous, Catherine Tait, Rick Linforth, Mohamed Salhab, Kyriacos Kyriacou, Chris W. Sutton, "*The Case for Using Nipple Aspirate Fluid for Early Detection of Breast Cancer*", BACR Special Meeting, "Breast Cancer - Bridging gaps in our knowledge to improve patient outcome", SAGE Newcastle United Kingdom, 7-9 October 2015.
- Sadr-ul Shaheed, Andreas Hadjisavvas, Kleitos Sokratous, Paul Loadman, Kyriacos Kyriacou, Chris Sutton, "*Analysis of HSP10 as a putative biomarker of breast cancer*", AACR Annual Meeting 2015, Philadelphia, PA USA, April 18-22, 2015.

2014

- Chris W. Sutton, Sadr-ul Shaheed, Paul. M. Loadman, Kyriacos Kyriacou, *“Multiple Biomarker analysis in Clinical Biopsies using MRM MS*, 62nd ASMS Conference on Mass Spectrometry and Allied Topics 2014, Baltimore, USA, 15–19 June 2014.

2013

- Sadr-ul Shaheed, Paul M. Loadman, Kyriacos Kyriacou, Chris Sutton *“Quantitative Comparison of Matched Normal and Tumour Breast Proteomic Profiles”*, BSPR annual scientific meeting, Nottingham United Kingdom, 3rd December 2013.
- Sadr-ul Shaheed, Paul M. Loadman, Kyriacos Kyriacou, Chris Sutton *“Development of a quantitative proteomics method for detection of specific candidate biomarkers in early-stage breast cancer”*, YCR annual scientific meeting, Harrogate United Kingdom, 26th June 2013.
- Sadr-ul Shaheed, Andrew Hanby, Valerie Speirs, Kyriacos Kyriacou, Chris Sutton *“Expression profiling of cofilin-1 in breast cancer cell lines and biopsies”*, AACR Annual Meeting 2013 Washington, DC USA Tuesday, April 09, 2013.
- Sadr-ul Shaheed, Paul M. Loadman, Kyriacos Kyriacou, Chris Sutton *“Quantitative Comparison of Matched Normal and Tumour Breast Proteomic Profiles”*, BSPR annual scientific meeting, Nottingham United Kingdom, 3rd December 2013.

Speaker:

2016

- University of Bradford UK *“A New Approach to Detect Breast Cancer at Its Earliest Stages”*, Wednesday, 6 December 2016.
- University of Bradford UK *“The case for using nipple aspirate fluid for early detection of breast cancer”*, Wednesday, 1st June 2016, (best presenter award).

2015

- The Cyprus Institute of Neurology and Genetics Nicosia Cyprus *“Oncoproteomics for Early Detection of Breast Cancer”*, Thursday, 9th July 2015.
- University of Bradford UK *“Proteome profiling of breast cancer cell lines”*, Friday 20th February 2015.

Scholarship:

2015

- **Erasmus Plus Programme**, Cyprus Institute of Neurology & Genetics, Nicosia, Cyprus. April 2015 – July 2015.

Table of Contents

Abstract	I
Dedications	II
Acknowledgements.....	III
Publications and Contributions	IV
List of Figures	XII
List of Tables	XV
Abbreviations.....	XVII
CHAPTER 1. INTRODUCTION.....	1
1.1 Breast Cancer.....	1
1.2 Stages, Types and Grades of breast cancer	4
1.2.1 Histopathological types.....	4
1.2.2 Breast cancer staging and grades	6
1.2.3 Genetic phenotypes.....	8
1.2.3.1 Luminal-type breast cancer	9
1.2.3.1.1 Luminal A and Luminal B.....	9
1.2.3.1.2 HER2 enriched breast cancer.....	10
1.2.3.2 Basal-type breast cancer	11
1.2.3.2.1 Basal-like (Basal A)	11
1.2.3.2.2 Claudin-low (Basal B).....	12
1.3 Methods for detecting Breast cancer	13
1.3.1 Self-examination	13
1.3.2 Molecular imaging for breast cancer.....	13
1.3.2.1 Mammography.....	13
1.3.2.2 Other imaging techniques.....	15
1.4 Biomarkers	19
1.4.1 Biomarkers of breast cancer	19
1.5 Proteomics.....	22
1.6 Proteomics and Mass Spectrometry	25
1.6.1 Sample preparation	27
1.6.2 Preparation of peptides mixture	28
1.6.3 Fractionation of peptides mixture.....	29
1.6.3.1 Peptides IPG-IEF fractionation	29
1.6.3.2 Liquid chromatography	30

1.6.3.3	Ion-exchange chromatography	31
1.6.4	Types of Mass spectrometry	31
1.6.4.1	MALDI-MS	32
1.6.4.1.1	MALDI-TOF/TOF ion source.....	34
1.6.4.1.2	Laser	34
1.6.4.1.3	MS mode for analytes acquisition.....	35
1.6.4.1.4	MS/MS mode for analyte identification	35
1.6.4.2	ESI-MS	36
1.6.4.2.1	Electrospray Ionisation Process	36
1.6.4.2.2	Orbitrap Fusion.....	37
1.6.4.2.3	Triple quadrupole mass spectrometer	41
1.6.5	Peptide fragmentation.....	42
1.6.6	Shotgun Proteomics	44
1.6.7	Database searching.....	45
1.6.8	Target-Decoy peptide searching.....	47
1.6.9	Quantitative Proteomics.....	48
1.6.9.1	Label-free quantitation	49
1.6.9.2	Label-based quantitation	49
1.6.10	Targeted mass spectrometry (MRM/PRM)	52
1.7	Biological samples for identification of breast cancer biomarkers.....	54
1.7.1	Cell culture.....	56
1.7.2	Tissue biopsies	57
1.7.3	Plasma/Serum:	57
1.7.4	Milk and Colostrum	58
1.7.5	Breast Cyst Fluid	59
1.7.6	Ductal Lavage (DL).....	60
1.7.7	Random Peri-areolar Fine Needle Aspiration (RPFNA).....	61
1.7.8	Nipple Aspirate Fluid (NAF)	62
1.8	NAF Proteomics.....	67
1.9	Aims and objectives.....	69
CHAPTER 2. MATERIALS AND METHODS.....		72
2.1	Chemical and reagents.....	72
2.2	Cell culture.....	73
2.2.1	Breast cell lines.....	73

2.2.2	Cells washing for proteomics	74
2.3	Breast tissues procurement	74
2.4	Collection of serum samples.....	75
2.5	NAF collection.....	75
2.6	Protein extraction.....	76
2.6.1	Cell lines	76
2.6.2	Tissue Biopsies.....	76
2.7	Protein quantification	77
2.8	SDS-PAGE and Western blotting	78
2.8.1	SDS-PAGE	78
2.8.2	Coomassie blue staining.....	78
2.8.3	Western blotting.....	79
2.9	Peptide preparation for LC-MS	82
2.9.1	Trypsin digestion.....	82
2.9.2	Checking the efficiency of tryptic digestion	82
2.9.3	Isobaric tag peptide labelling	83
2.9.4	OffGel fractionation.....	83
2.9.5	SCX fractionation.....	84
2.10	LC-MS analysis	85
2.10.1	Ultraflex II	85
2.10.2	Orbitrap Fusion	86
2.10.3	MRM-MS	87
2.11	MS data processing.....	89
2.12	Post-processing of MS data	91
2.12.1	Statistical analysis	91
2.12.2	Hierarchical clustering	92
2.12.3	Limma t-statistics.....	93
2.12.4	Gene Ontology (GO) enrichment analysis	94
2.13	Comparison of Western Blotting, iTRAQ and MRM-MS quantitation methods.....	94
2.13.1	Processing of immune-blot	94
2.13.2	Processing of iTRAQ labelled peptides	95
2.13.3	Processing of MRM peptides.....	95
2.13.4	Calculating the amount of protein	95
CHAPTER 3.PROTEOME PROFILING OF BREAST CANCER CELL LINES		96

3.1	Introduction.....	96
3.2	Materials and methods	107
3.2.1	Breast cell lines.....	107
3.2.2	Protein extraction from cell lines.....	107
3.2.3	LC-MS analysis.....	108
3.2.3.1	Trypsin digestion and iTRAQ labelling.....	108
3.2.3.2	Peptide fractionation.....	108
3.2.3.3	Ultraflex II analysis	109
3.2.3.4	Fusion Orbitrap analysis	110
3.2.3.5	MS data processing.....	112
3.2.3.6	Statistical analysis	113
3.2.4	Western blot analysis.....	114
3.3	Results.....	115
3.3.1	Comparison of protein and peptide identification from two MS platforms	115
3.3.2	Quantitative comparison of breast cell lines proteome	120
3.3.3	Evaluation of Orbitrap Fusion data	122
3.3.4	Hierarchical Clustering of breast cell lines	126
3.3.5	Breast cancer stratification using cancer cell line proteome profiles	128
3.3.5.1	Luminal vs Basal	129
3.3.5.2	Luminal A (MCF-7)	131
3.3.5.3	Luminal B (ZR-75)	133
3.3.5.4	HER2 positive (MDA-MB-453).....	136
3.3.5.5	Basal-like specific (MDA-MB-468)	138
3.3.5.6Stem cell characteristics of Claudin-low cell line MDA-MB-231	142
3.3.6	Western blot analysis.....	146
3.4	Discussion	149
3.4.1	MS2 and SPS quantification	149
3.4.2	Immortalised cell lines compared to primary cells.....	152
3.4.3	Breast cancer classifiers.....	157
3.5	Conclusions	167
CHAPTER 4. PROTEOME PROFILING OF NIPPLE ASPIRATE FLUID.		168
4.1	Introduction.....	168
4.2	Materials and methods	172

4.2.1	Patients and sample collection	172
4.2.2	Sample preparation	175
4.2.3	Proteomic analysis.....	175
4.2.3.1	SCX Peptide fractionation.....	175
4.2.3.2	Fusion Orbitrap analysis	176
4.2.3.3	Data Analysis.....	176
4.2.4	Western blot analysis.....	178
4.3	RESULTS	179
4.3.1	NAF sample characterisation.....	179
4.3.2	NAF proteomic analysis.....	181
4.3.3	Comparison of identified 2D NAF proteome	183
4.3.4	Cellular localisation of NAF Proteome	185
4.3.5	Functional annotation of NAF Proteome.....	187
4.3.6	Expression of cancer related proteins.....	191
4.3.7	Pathway analysis of differently expressed proteins in the NAF proteome.	201
4.3.8	Verification of presence of selected targets by Western blot. ..	202
4.4	DISCUSSION	204
4.4.1	Comparison of identified NAF proteome with published NAF data.	204
4.4.2	Comparison of NAF proteome with Plasma proteome.....	205
4.4.3	Phenotype specific protein expression changes	206
4.4.4	Cancer specific protein expression changes	209
4.4.5	Diagnostic application.....	213
4.5	CONCLUSIONS	218
CHAPTER 5. DEVELOPMENT OF A NOVEL MULTIPLEX MRM-MS ASSAY FOR DETECTION OF BREAST CANCER.....		220
5.1	Introduction.....	220
5.2	Selection of proteins and peptides for MRM-MS assay	229
5.3	Materials and methods	235
5.3.1	Chemicals and Reagents.....	235
5.3.2	Synthetic peptides	235
5.3.2.1	Stock peptides	235
5.3.2.2	Working solution	235
5.3.3	MALDI Mass spectrometry	236
5.3.4	Sample preparation	236

5.3.5	Protein determination.....	236
5.3.6	LC-MRM-MS analysis.....	239
5.3.7	MRM-MS Data analysis	240
5.4	Results.....	241
5.4.1	MALDI MS and MS/MS data.....	241
5.4.2	Optimizing the MRM-MS transitions	242
5.4.3	Calibration curves of the synthetic peptides	246
5.4.4	Multiplex analysis of biological materials	253
5.4.4.1	Breast cell lines	253
5.4.4.2	Tissue biopsies.....	258
5.4.4.3	Serum.....	263
5.4.4.4	Nipple aspirate fluid	266
5.4.4.5	MRM-MS assay coefficient of variation in biological samples	269
5.5	Discussion	271
5.6	Conclusions	279
CHAPTER 6. FINAL DISCUSSION AND FUTURE DIRECTIONS		280
6.1	Discussion	280
6.2	Future directions	290
CHAPTER 7. REFERENCES.....		292

List of Figures

Figure 1-1: Average number of new cases per year and age-specific incidence per 100,000 population of female breast cancers in UK (CancerResearchUK, 2016b).....	1
Figure 1-2: The significance of Proteomics on human life.	24
Figure 1-3: Strategies for discovering novel breast cancer protein biomarkers.	25
Figure 1-4: Workflow illustrating typical proteomic experiment (Domon and Aebersold, 2010).	27
Figure 1-5: Schematic view of a MALDI-TOF/TOF mass spectrometer (Suckau et al., 2003).....	33
Figure 1-6: Schematic view of Orbitrap Fusion Tribrid mass spectrometer.....	38
Figure 1-7: Schematic view of a triple quadrupole mass spectrometer system (Ho et al., 2003).	42
Figure 1-8: Peptide backbone fragmentation: peptide structure, selected fragmentation techniques, and fragment ion nomenclature (Zhurov et al., 2013).....	44
Figure 1-9: Schematic diagram of MRM/PRM.	53
Figure 2-1: Representation of transfer “sandwich” assembly for Western blotting.	79
Figure 2-2: Overview of the different stages of Proteome discover and ProteinScape for LC-MS data processing.....	90
Figure 2-3: R script for Hierarchical clustering.	92
Figure 2-4: R script for Limma t-statistics.....	93
Figure 3-1: Chemical structures of iTRAQ 4-plex, TMT 6-plex and iTRAQ 8-plex reagents.	100
Figure 3-2: Model of the human mammary epithelial grading linked to breast cancer subtypes (Prat and Perou, 2009).	103
Figure 3-3: Workflow illustrating the proteomics-based approach.	111
Figure 3-4: Schematic representation of the MS data processing workflow..	113
Figure 3-5: Comparison of Ultraflex II and Orbitrap Fusion proteome.	116
Figure 3-6: Dynamic range of proteins identified with Ultraflex II vs Orbitrap Fusion.....	118
Figure 3-7: Pearson correlation coefficient between log ₂ ratios of common proteins in both MS datasets – (MS/MS; Ultraflex ratio and SPS; Orbitrap Fusion ratios).	119
Figure 3-8: Hierarchical clustering of breast cell lines (cancer and normal-like) proteome compared to primary human mammary epithelial cells.....	127
Figure 3-9: Venn diagram comparison of signature proteomes for breast cancer cell lines.	129
Figure 3-10: Volcano plot analysis of differentially expressed proteins in Luminal vs Basal breast cancer cell lines.	130

Figure 3-11: Volcano plot analysis of differentially expressed proteins in MCF-7 (Luminal A).	131
Figure 3-12: Volcano plot analysis of differentially expressed proteins in ZR-75 (Luminal B).	133
Figure 3-13: Volcano plot analysis of differentially expressed proteins in MDA-MB-453 (HER2 positive).	136
Figure 3-14: Volcano plot analysis of differentially expressed proteins in MDA-MB-468 (Basal-Like).	138
Figure 3-15: Volcano plot analysis of differentially expressed proteins in MDA-MB-231 cell line (Claudin-Low).	143
Figure 3-16: Western blot analysis of ACTB, CD44, ALDH2, CALR, PTGES3, HSPE1 in breast cell lines.....	146
Figure 3-17: Comparison of western blot analysis and MS results for HSPE1 in breast cell lines.	150
Figure 3-18 Comparison of MS ² fragmentations for a peptide of HSPE1 and dynamic range of reporter ions on Orbitrap Fusion and Ultraflex II.....	152
Figure 3-19: Up-regulated DNA replication protein-protein interactions ubiquitous to all cell lines compared to HMEC.....	154
Figure 3-20: Protein-protein interaction map of down regulated proteins for all cell lines compared to HMEC.....	156
Figure 3-21: Expression of HER2 and EGFR in breast cancer cell lines.	158
Figure 3-22: Comparison of SPS analysis and MS results for PTGES3 in breast cell lines.	160
Figure 3-23: Comparison of SPS analysis and MS results for CD44 in breast cell lines.	164
Figure 3-24: Comparison of SPS analysis and MS results for Calreticulin in breast cell lines.	166
Figure 4-1: SDS PAGE analysis of matched pairs of NAF samples.	180
Figure 4-2: Comparison of 1D and 2D NAF proteome profiling.	182
Figure 4-3: Comparison of NAF proteome acquired by 2D-LC/MS analysis.	183
Figure 4-4: Comparison of the matched 2D NAF protein profiles.	184
Figure 4-5: Pearson correlation coefficient for each NAF profile compared to each other, based on median normalised peak area of each protein.	185
Figure 4-6: Normal cellular localisation of the 567 proteins common to all the NAF samples.	186
Figure 4-7: NAF proteins detected in breast cancer signalling pathways.....	201
Figure 4-8: Western blot analysis of ACTB, CD44, CYP3A4, VIM, CALR, EGFR, CFL1 and PIP in NAF.	203
Figure 4-9: Overlapping identities with the data from the NAF study by Pavlou et al 2010.	204
Figure 4-10: overlapping identities with the Human Plasma Proteome database.	205
Figure 4-11: Significantly increased in the healthy volunteer (HV), compared to diseased samples.	207

Figure 4-12, A: Significantly increased in the cancer patients compares to healthy volunteer and patient with benign phyllodes.....	211
Figure 4-13 B: Significantly increased in the cancer patients compares to healthy volunteer and patient with benign phyllodes.....	212
Figure 5-1: Number of PubMed Publications on MRM-MS and Proteomics.	221
Figure 5-2: Workflow for the design of a MRM-MS assay.....	232
Figure 5-3: Representation of proteomics work flow for the preparation of samples.	237
Figure 5-4: Retention time and charge status of target precursors in multiplex MRM-MS assay.	242
Figure 5-5: Representative standard curves from the multiplex LC-MRM/MS analysis of 6 peptides (CFL1 and HSPE1).....	248
Figure 5-6: Representative standard curves from the multiplex LC-MRM/MS analysis of 6 peptides (SH3BGRL3, PTGES3 and ALB).....	250
Figure 5-7: Representative standard curves from the multiplex LC-MRM/MS analysis of 6 peptides (AOC3, LIPE and ACTB).	251
Figure 5-8: Sensitivity comparison of multiplex MRM-MS assay on three different platforms.	252
Figure 5-9: Expression of Cofilin-1, HSPE1, SH3BGRL3 and PTGES3 in breast cell lines.	256
Figure 5-10: Expression of AOC3, LIPE, ACTB and ALB in breast cell lines.	257
Figure 5-11: Expression of CFL-1, HSPE1, PTGES3 and SH3BGRL3 in breast tissue biopsies.	260
Figure 5-12: Expression of ACTB, ALB, LIPE and AOC3 in breast tissue biopsies.....	262
Figure 5-13: Expression of CFL1, HSPE1, PTGES3 and SH3BGRL3 in serum samples.	264
Figure 5-14: Expression of AOC3, LIPE, ACTB and ALB in serum samples.	265
Figure 5-15: Expression of CFL1, HSPE1, PTGES3 and SH3BGRL3 in nipple aspirate fluid.	267
Figure 5-16: Expression of AOC3, LIPE, ACTB and ALB in nipple aspirate fluid.	268
Figure 5-17: CV values for the multiplexed MRM-MS assays measured in breast cell lines (BCL), tissue biopsies, serum, and nipple aspirate fluid (NAF), consisting of targets (18 peptides, 8 proteins).	270
Figure 5-18: Sensitivity of multiplex MRM-MS assay in breast cell lines (BCL), tissues biopsies, serum, and nipple aspirate fluid (NAF).	275
Figure 6-1: (A) ACTB, CFL-1, HSPE1 and PTGES3 Western blot analysis of phenotypic cell lines. Comparison of three analytical methods; Western blotting, iTRAQ and MRM-MS, for expression of (B) CFL-1, (C) HSPE1 and (D) PTGES3 in breast cell lines.	288

List of Tables

Table 1-1: Histopathological types and characteristics of breast cancer (DeSantis et al., 2014).	5
Table 1-2: TNM (Tumour size, Lymph Nodes and Metastasis) system for different stages of breast cancer.	7
Table 1-3: Molecular classification of breast carcinoma (Holliday and Speirs, 2011).	8
Table 1-4: Methods for detecting breast cancer.	18
Table 1-5: Biological samples used for biomarker proteome of breast cancer.	55
Table 1-6: Potential biomarkers in nipple aspirate fluid.	63
Table 2-1: Antibodies used for Western blotting.	81
Table 2-2: Strong Cation Exchange Elution Buffer (KCl) Concentrations.	84
Table 2-3: List of instruments and columns optimised gradients for each experiment.	88
Table 3-1: Breast cell lines attributes and iTRAQ labelling.	109
Table 3-2: Proteome data comparison of instruments.	116
Table 3-3: Median normalised iTRAQ ratios of heat shock proteins in breast cell lines, acquired on Orbitrap Fusion (SPS) and Ultraflex II (MS ²).	121
Table 3-4: Significantly up-regulated proteins in cancer and premalignant cell lines compared to primary human mammary epithelial cells.	123
Table 3-5, a: Significantly down-regulated proteins in cancer and premalignant cell lines compared to primary human mammary epithelial cells.	124
Table 3-6: Signature proteins for luminal A type breast cancer.	132
Table 3-7: Signature proteins for luminal B type breast cancer.	134
Table 3-8: Signature proteins for HER2 positive breast cancer.	137
Table 3-9: Signature proteins for Basal-like breast cancer.	139
Table 3-10: Signature proteins for Claudin-Low breast cancer.	144
Table 3-11: Quantitative expression of ACTB, CD44, ALDH2, CALR, PTGES3 and HSPE1 in breast cell lines.	148
Table 4-1: Healthy volunteer and patient parameters and NAF characteristics.	174
Table 4-2: List of cell adhesion (GO:0007155; GO Biological Process) proteins identified in the NAF samples.	188
Table 4-3: Selected mitogenic factors identified in NAF.	193
Table 4-4: Selected mitogenic receptors identified in NAF.	194
Table 4-5: List of proteolysis proteins identified in the NAF samples.	197
Table 4-6: Biomarkers under investigation by the NCI Early Detection Research Network.	214
Table 5-1: Comparison of Xevo TQD, Quattro Premier XE and Quattro Ultima, used for multiplex MRM-MS assay.	225
Table 5-2: List of proteins selected for MRM-MS assay and expression in breast cancer.	231

Table 5-3: Panel of proteins and peptides for MRM-MS analysis.	234
Table 5-4: List of samples used in this study with biopsy types and characteristics.	238
Table 5-5: Verification of synthetic peptides analysed by MALDI MS against SwissProt database.	241
Table 5-6: Optimised Dwell time, Cone voltage and Collision energy (CE) conditions for multiplex MRM-MS assay, using synthetic peptides of target proteins.	244
Table 5-7: Optimised Retention time (RT) and transitions for multiplex MRM/MS assay, using synthetic peptides of target proteins.	245
Table 5-8: Summary of LoD and LoQ all standard peptides on three platforms.	249
Table 5-9: Replicate analyses of three peptides of Cofilin-1 in breast cell lines.	254
Table 6-1: Comparison of phenotype signature with clinical stages of breast cancer.	284
Table 6-2: The expression of phenotype specific proteins in healthy and cancerous NAF.	286

Abbreviations

2D	Two-dimensional
AA	Amino acid
ABC	ATP-binding cassette
ACN	Acetonitrile
ACTB	Actin, cytoplasmic 1
AGR2	Anterior gradient protein 2 homolog
AK1	Adenylate kinase 1
AKT	Protein kinase B
ALB	Serum albumin
ALDH1A3	Aldehyde dehydrogenase family 1 member A3
ALDH2	Aldehyde dehydrogenase, mitochondrial
AMBIC	Ammonium bicarbonate
AOC3	Membrane primary amine oxidase
AP-1	Activator protein 1
APF	Antiproliferative factor
BCF	Breast Cyst Fluid
BCL2	B-cell lymphoma 2 protein
BCSCs	Breast cancer stem cells
BRCA-1	Breast cancer 1 early onset
BRCA-2	Breast cancer 2 early onset
BSA	Bovine serum albumin
CADM4	Cell adhesion molecule 4
CALML5	Calmodulin-like protein 5
CALR	Calreticulin
CAM	Cell adhesion molecule
CapG	Macrophage-capping protein
CAV1	Caveolin-1
CCNB1	Cyclin B1
CD10	Neprilysin
CD34	Ki-67
CD44	Antigen CD44
CDC	Cell division cycle protein
CDH1	E-cadherin
CDK	Cyclin-dependent kinase
CDK1	Cyclin-dependent kinase 1
CE	Collision energy
CFL1	Cofilin-1
CGN	Cingulin
CHAPS	3-cholamidopropyl dimethylammonio]-1-propanesulfonate
CHCA	α -cyano-4-hydroxy cinnamic acid
CID	Collision-induced dissociation
CKAP4	Cytoskeleton-associated protein 4
CM	Conditioned medium
CMIP	C-Maf-inducing protein Cor1A Coronin 1A
CRP	C-reactive protein
CV%	Coefficient of Variations

CV	Cone voltage
CVD	Cardiovascular disease
CXCR	Chemokine (C-X-C motif) receptor
DAVID	The database for annotation, visualisation and integrated discovery
DBT	Digital breast tomosynthesis
DCIS	Ductal carcinoma in situ
DDA	Data-dependent acquisition
DIA	Data independent acquisition
DL	Ductal Lavage
DMEM	Dulbecco's modified Eagle's medium
DMSO	Dimethyl sulfoxide
DNMT	DNA methyltransferase
DOC	Sodium deoxycholate
DTT	Dithiothreitol
ECD	Electron-capture dissociation
ECM	Extracellular matrix
EDD	Electron-detachment dissociation
EDTA	Ethylene diamine tetra acetic acid
EGF	pro-epidermal growth factor
EGFR	Epidermal growth factor receptor
ELISA	Enzyme-linked immunosorbent assay
EMT	Epithelial-to-mesenchymal transition
Eno1	Alpha-enolase
ER	Estrogen receptor
ERK	Extracellular-signal-regulated kinases
ESI	Electrospray ionisation
ESI-MS	Electrospray ionisation mass spectrometry
ETD	Electron-transfer dissociation
FA	Formic acid
FACS	Fluorescents assisted cell sorting
FBS	Foetal bovine serum
FDR	FALSE discovery rate
FGF2	Fibroblast growth factor 2
FTICR	Fourier transform ion cyclotron resonance
FTMS	Fourier transform ion cyclotron mass spectrometry
G3BP2	Ras GTPase-activating protein-binding protein 2
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GO	Gene ontology
HCD	High-energy collisional dissociation
HD	Healthy donor
HMEC	Human mammary epithelial cells
HPLC	High performance liquid chromatograph
HRAM	High-resolution, accurate-mass
Hsp90	Heat shock protein 90
HSPE1	10kDa heat shock protein, mitochondrial
HSPs	Heat shock proteins
HV	Healthy volunteer
IAA	Iodoacetamide
IC	Invasive carcinoma
ICAT	Isotope-coded affinity tags

ICPL	Isotope-coded protein labels
ID	Inner diameter
IDC-NST	Invasive ductal breast cancer of no special type
IEC	Ion-exchange chromatography
IEF	Peptide isoelectric focusing
Ig	Immunoglobulin
IGF1	Insulin-like growth factors I
IGFBP3	Insulin-like growth factors-binding proteins 3
IGHV	Ig heavy-chain variable-region gene
IL16	Interleukin 16
IL5	Interleukin 5
IL5RA	Interleukin 5 receptor alpha
IPA	Ingenuity pathway analysis
IPG	Immobilized pH gradients
ITG	Integrin ITIH Inter-alpha-trypsin inhibitor
ITGA2	Integrin alpha-2
ITGAV	Integrin alpha-V
ITGB4	Integrin beta-4
iTRAQ	isobaric tags for relative and absolute quantitation
KIT	Stem cell growth factor receptor
KLK	kallikrein
KRT	Keratin
KYNU	Kynureninase
LC	Liquid chromatography
LC-MS	Liquid chromatography–mass spectrometry
LIPE	Hormone-sensitive lipase
LIT	Linear ion trap
LoD	Limits of detection
LoQ	Limit of quantification
LP	Light peptides
m/z	Mass to charge ratio
MALDI	Matrix-assisted laser desorption/ionization
MAPK	Mitogen activated protein kinases
MCM	Mini-chromosome maintenance
MEGM	Mammary Epithelial Cell Growth Medium
MgCl ₂	Magnesium chloride
MHC	Major histocompatibility complex
miR	Micro RNA
MKI67	Antigen Ki-67
MMTS	Methyl methanethiosulfonate
Moe	Moesin
MP	Mobile phase
MRI	Magnetic resonance imaging
MRM	Multiple reaction monitoring
MRM-MS	Multiple reaction monitoring mass spectrometry
MS	Mass spectrometry
MS/MS	(MS ²) Tandem mass spectrometry
MUCL-1	Mucin-like protein-1
MudPIT	Multidimensional Protein Identification Technology
Mw	Molecular weight

MWCO	Molecular weight cut-off
Myo9	Myosin-9
NAF	Nipple Aspirate Fluid
NCE	Normalized collision energy
NCI	National Cancer Institute
ND	Nipple Discharge
nETD	Negative electron-transfer dissociation
NGS	Next-Generation Sequencing
niECD	Negative-ion electron capture dissociation
niETD	Negative electron-transfer dissociation
NOTCH1	Neurogenic locus notch homolog protein 1
O/N	Over night
OD	Optical density
p53	Tumour protein 53 gene product
PAGE	Polyacrylamide gel electrophoresis
PAI-1	Plasminogen activator inhibitor-1
PARP	Poly [ADP-ribose] polymerase
PBS	Phosphate buffered saline
PBST	PBS tween
PCR	Polymerase chain reaction
PDCD4	Programmed cell death protein 4
PDGF	Platelet-derived growth factor
PDGF-C	Platelet-derived growth factors C
PE	Phycoerythrin
PEP	Posterior error probability
PET-CT	Positron Emission Tomography with computer tomography
PI3K	Phosphatidylinositol-3-kinase
pIEF	Peptide isoelectric focusing
PIGR	Polymeric Ig receptor
PIP	Prolactin-inducible protein
PLMS	Post lift metastable suppressor
PND	Pathologic Nipple Discharge
PR	Progesterone receptors
pRB	Retinoblastoma protein
PRM	Parallel reaction monitoring
PRMT1	Arginine methyltransferase 1
PSA	Prostate specific antigen
PSM	Petide-spectrum match
PTGES3	Prostaglandin E synthase 3
PTN	Patient nipple aspirate fluid
PTPN	Tyrosine-protein phosphatase non-receptor
PTT	Patient tissue
PVDF	Polyvinylidene fluoride
QM-PCR	Quantitative multiplex methylation-specific polymerase chain reaction
RF	Rradio frequency
ROR1	Receptor tyrosine kinase-like orphan receptor 1
ROS	Reactive oxygen species
RP	Reverse-phase
RPFNA	Random Peri-areolar Fine Needle Aspiration
RPMI	Roswell Park Memorial Institute medium

RRM	Ribonucleotide reductase M
RS	Regulation score
RT	Retention time
RTK	Receptor tyrosine kinase
SATB1	Special AT-rich sequence-binding protein 1
SAX	Strong anion-exchange
SCX	Strong cation exchange
SD	Standard deviation
SDS	Sodium dodecyl sulphate
SEC	Size exclusion chromatography
SF3B1	Splicing factor 3B subunit 1
SID	Stable isotope dilution
SILAC	Stable isotope labelling by amino acids in cell culture
SLC	Solute carrier
SOD-1	Superoxide Dismutases
SOD3	Extracellular superoxide dismutase
SPS	Synchronous MS3 precursor selection
STAT	Signal transducer and activator of transcription
TAE	TRIS, acetate, EDTA
TEAB	Triethylammonium bicarbonate
TEMED	Tetra methyl ethylenediamine
TF	Thomsen-Friedenreich
TFA	Trifluoroacetic acid
TFR	Ferritin transferrin protein
TGF	Transforming growth factor
TGF β R2	Transforming growth factor beta receptor type II
TGF β R3	Transforming growth factor beta receptor type III
TIS	Timed ion selector
TMA	Tissue microarrays
TMT	Tandem mass tag
TNBC	Triple-negative breast cancer
TNM	Tumour size, Lymph Nodes and Metastasis
TOF	Quadrupole time of flight
TP53	Tumour Protein p53
TRIS	Tris(hydroxymethyl)aminomethane
Tris-HCl	Tris hydroxymethyl aminomethane hydrochloride
Tween	Tween 20
uPA	Urokinase-dependent plasminogen activator
VDAC	Voltage-dependent anion-selective channel
VEGFR1	Vascular endothelial growth factor receptor
VIM	Vimentin
WB	Western blotting
WEE1	Wee1-like protein kinase
WT	Wildtype

CHAPTER 1. INTRODUCTION

1.1 Breast Cancer

Breast cancer is the leading cause of death in the world among women aged 40-59 years, with an estimated 1.67 million new cancer cases diagnosed in 2012 (25% of all cancers). It is the most common cancer; with 883,000 cases each year in developing and 794,000 cases in more advanced regions of the world. (Ferlay J et al., 2013). In UK, around 55,200 new cases of breast cancer were diagnosed in 2014, which was an average of 150 cases diagnosed every day. The breast cancer mortality rate in UK is 31 deaths every day according to Cancer Research UK (CancerResearchUK, 2016b). The breast cancer incidence is high, almost 60%, in women aged 55 and over, while women under the age of 40 have a considerably lower chance of developing the breast cancer (Figure 1-1) (CancerResearchUK, 2016b).

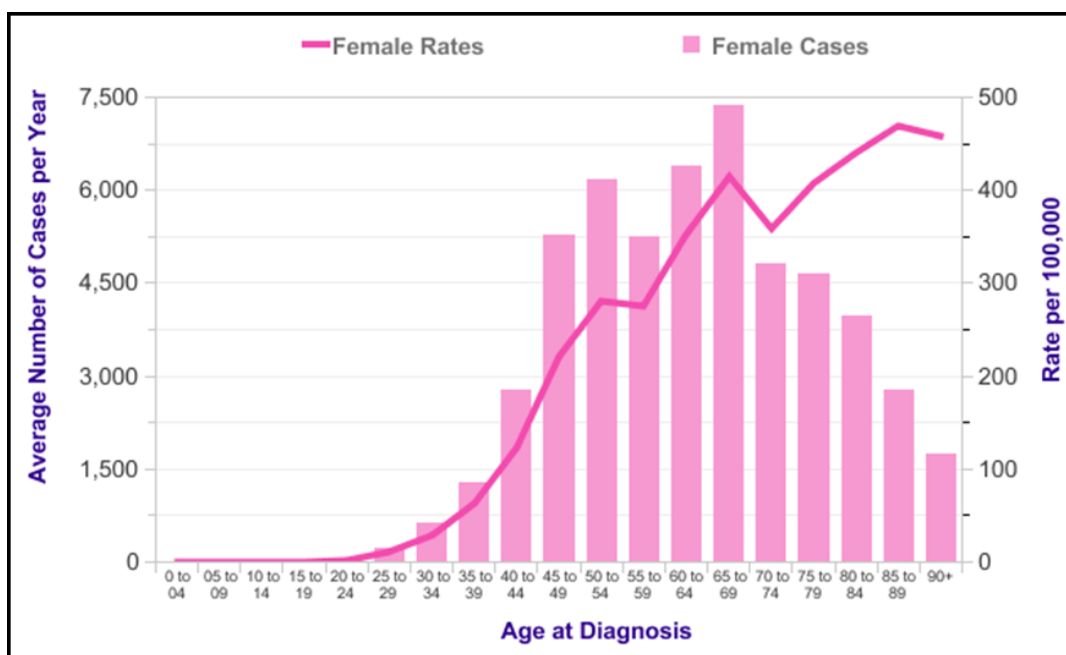


Figure 1-1: Average number of new cases per year and age-specific incidence per 100,000 population of female breast cancers in UK (CancerResearchUK, 2016b).

Limited breast health awareness and access to screening has a significant impact on the ability of health services to diagnose and treat the disease at a sufficiently early stage, and in many cases, by the time the patient presents at clinic, the cancers have already become aggressive. This problem is not unique to less developed countries as deaths also reached an all-time high during 2014 in the UK for women under 50 years old (CancerResearchUK, 2016b). The risks associated with breast cancer are well established and include genetic predisposition, reproductive factors (menarche, parity, lactation, menopause), environment (chemical exposures, drugs, infectious agents), demographic factors (age, race, sex, socio-economic status, geographic location), systemic factors (epigenetics) and lifestyle (smoking, alcohol use, diet, obesity, exercise) (Eccles et al., 2013),

The major risk factor is age and the majority of breast cancer cases occur in women after menopause. However, the correlation of risk with reproductive, environmental and lifestyle parameters, strongly indicate the origins of many breast cancers are concealed by a prolonged period of dormancy that only manifest in later life (Banys et al., 2012). Lifestyle parameters are unique among these factors, in that they fall within the scope of the individual to proactively reduce these risks by taking preventative measures, whereas reproductive, environmental and demographic factors are beyond the control of most individuals.

Breast cancer may be an inherited or an acquired genetic disease that is influenced by reproductive, environmental and behaviour factors. Hereditary breast cancers, account for 5-10% of cases, and in most cases, are related to mutations in BRCA-1 and BRCA-2 genes. Women with these genetic mutations

have 10- to 30-fold higher risk of developing breast cancer in life as compared to the general population (King et al., 2003). Genetic predisposition can be identified initially through investigation of family history of the disease and confirmed by screening for predictive mutations associated with specific high-penetrance genes, including BRCA1, BRCA2, TP53 and PTEN (Aloraifi et al., 2015).

Sporadic breast cancer developed due to consequences of somatic mutations over the life time account for 90–95% of cases. These mutations may be the result of polymorphisms in genes, encoded in the synthesis of proteins involved in cell signalling pathways, estrogen metabolism and DNA repair (Zhang et al., 2006).

The critical and poorly understood events in breast tumour development that have major effects on clinical management and outcome are (a) the changes from DCIS to invasive carcinoma and (b) the metastatic blowout of primary tumours to distant organs (Hu and Polyak, 2008). Heterogeneity of breast cancer is caused by different genetic and epigenetic mutations, which results in alterations of cell functions such as apoptosis, cell proliferation and angiogenesis (Reis-Filho and Lakhani, 2003). The major genetic mutations observed are high expression of proto-oncogenes, alterations of DNA repair or tumour suppressor genes, chromosomal instability and reactivation of telomerase, causing uncontrolled cell proliferation and tumour formation (Rodenhiser and Mann, 2006).

Because of extensive epidemiological and etiological studies, it is becoming easier to identify those women who are at highest risk of breast cancer. Unfortunately, none of these measures can predict when breast cancer

will occur. Regular screening for the appearance of breast cancer, as recommended for women reaching menopause, is critical in high risk women during their reproductive years. There are various screening modalities, each with advantages and disadvantages (Table 1-4), making them sub-optimal for more frequent monitoring that is required for this group. Consequently, the major means of preventative intervention for these women can be extreme, involving elective surgery to remove both breasts. Therefore, new and more sensitive approaches are required to detect the earliest stages of the disease, which will enable less drastic means of intervention.

1.2 Stages, Types and Grades of breast cancer

1.2.1 Histopathological types

Breast cancer is a complex and heterogeneous disease and understanding this diversity is important for diagnostic and prognostic applications. A wide range of histopathological types (Table 1-1) include: ductal carcinoma, tubular carcinoma, lobular carcinoma, mucinous carcinoma and some rare types like papillary and metaplastic breast cancer (Cancerresearchuk, 2016a).

Type of cancer	Frequency	Characteristics	Treatment options
Ductal Carcinoma	85-90%	Start in the ducts of breasts, may be pre-invasive, non-invasive, ductal intraepithelial neoplasia (DIN) or intra ductal	Surgery, chemotherapy and radiotherapy
Lobular Carcinoma In Situ (LCIS)	5-10%	Not a cancer (lobular neoplasia), Increased risk of breast cancer, Men can develop LCIS, can change into invasive lobular carcinoma	Breast examination every 6 to 12 months, Hormone therapy
Mucinous breast cancers	2%	Slow growing, likely to spread to the lymph nodes	Surgery and radiotherapy
Inflammatory breast cancer	1-4%	Breast tissue becomes inflamed; The cancer cells block the lymph nodes in the breast.	Surgery, chemotherapy and radiotherapy
Paget's disease	1-4%	A red, scaly rash of the skin over the nipple and areola, Similar to other skin conditions such as psoriasis or eczema	Surgery
Breast Angiosarcomas	1%	High incidence rate in younger women	Surgery and radiotherapy
Medullary Breast Cancers	5%	Few cells bigger than other breast cancer cells, common in women who have inherited a faulty BRCA 1 gene	Surgery, chemotherapy and radiotherapy
Tubular Breast Cancer	<1%	Cells look like tubes when seen under a microscope	Lymph nodes removed by surgery, Radiotherapy
Adenoid Cystic Carcinoma of the breast	<1%	Also called as cribriform cancer, slow-growing but locally aggressive,	Lymph nodes removed by surgery, Radiotherapy
Metaplastic Breast cancer	<1%	Mixture of two or more cell types,	Surgery, chemotherapy and radiotherapy
Lymphoma of the breast	<1%	Contain both lymphoid tissue and breast tissue	Chemotherapy, radiotherapy, and possibly surgery
Phyllodes or Cystosarcoma Phyllodes	<1%	Either cancerous (malignant) or Non-cancerous (benign),	Surgery, some cases radiotherapy
Papillary Breast cancer	<1%	Cells are like the shape of a fern, affect older women, slower growing.	Surgery

Table 1-1: Histopathological types and characteristics of breast cancer (DeSantis et al., 2014).

Breast cancer heterogeneity is due to the alteration of numerous oncogenic pathways, germ line mutations like breast cancer 1 early onset (BRCA-1) and breast cancer 2 early onset (BRCA-2) (Petrucci et al., 1993), as well as many non-genetic factors such as hypoxia and lactic acidosis which are involved in tumour micro-environmental stresses (Gatza et al., 2011). Furthermore, there is etiological heterogeneity of hormone-receptor defined subtypes of breast cancer (Redondo et al., 2012).

1.2.2 Breast cancer staging and grades

There are different staging systems for breast cancer but the most common one is the TNM (Tumour size, Lymph Nodes and Metastasis) system. Different stages of breast cancer are summarised in Table 1-2, using the TNM system (Cancerresearchuk, 2016a, Sobin and Fleming, 1997, Singletary et al., 2002). If cancer has not spread beyond the breast or the lymph nodes, it is considered as an early breast cancer, but if it has spread to another part of body such as the liver or bones (metastasis), then it is called advanced breast cancer.

Grading of breast cancer is performed by using tissue biopsies and observing the shape of cancer cells under the microscope (Singletary et al., 2002, Cancer, 2002). Breast cancers can be (Cancerresearchuk, 2016a, Cancer, 2002, Brierley, 2017):

- a. Grade 1; tumour cells are dividing at a slow rate and shape of nuclei are small and uniform.
- b. Grade 2; tumour cells are dividing at moderate rate and nuclei are medium-to-large size but of uniform size and shape within the tissue.

- c. Grade 3; tumour cells are dividing at a fast rate and nuclei are large and vary in size and shape.

Over Stage	T (Tumour)	N (Lymph Nodes)	M (Metastasis)
Stage 0	Tis	N0	M0
Stage 1	T1	N0	M0
Stage IIA	T0	N1	M0
	T1	N1	M0
	T2	N0	M0
Stage IIB	T2	N1	M0
	T3	N0	M0
Stage IIIA	T0	N2	M0
	T1	N2	M0
	T2	N2	M0
	T3	N1	M0
	T3	N1	M0
Stage IIIB	T4	Any N	M0
Stage IIIC	Any T	N3	M0
Stage IV	Any T	Any N	M1

T0; no primary tumour, **Tis**; ductal carcinoma in situ (DCIS), **T1**; Tumour size is 0.1cm across or less, **T2**; tumour is more than 2 centimetres but no more than 5 centimetres across, **T3**; tumour is bigger than 5 centimetres across, **T4**; Tumour of any size with extension to chest wall/skin or ulceration and inflammation, **N0**; no cancer cells in any nearby nodes. **N1**; cancer cells are in the 1-3 axillary lymph nodes, **N2**; cancer cells are in the 4-9 axillary lymph nodes, **N3**; cancer cells in infra or supraclavicular lymph nodes, or in >10 axillary lymph nodes, **M0**; no sign that the cancer has spread, **M1** means the cancer has spread to another part of the body.

Table 1-2: TNM (Tumour size, Lymph Nodes and Metastasis) system for different stages of breast cancer.

1.2.3 Genetic phenotypes

The genomic status can be a key aspect to understanding the biological characteristics of a tumour, where the DNA content is investigated to differentiate between subtypes. There are two types of epithelial cells in the human mammary glands; basal and luminal (Perou et al., 2000). Breast tumours are composed of a variety of associated cells alongside the cancerous cells and gene expression profiling of histology complex tissue has identified the cell-type-specific expression of certain clustered groups of genes. On the basis of microarray gene expression profiling, breast cancer can be classified into different molecular phenotypes. Genetic phenotyping initially identified 4 phenotypes; Luminal A, Luminal B, HER2 positive, and Basal type(Perou et al., 2000), later the basal type was more extensively classified as Basal-like and Claudin-low (Holliday and Speirs, 2011) (Table 1-3).

Genotypic Classification	Immunoprofile	Characteristics
Luminal A	ER+, PR+/-, HER2-	MKI67 low, endocrine responsive, low chemotherapy responsive
Luminal B	ER+, PR+/-, HER2+	MKI67 high, usually endocrine responsive, variable to chemotherapy. HER2+ are trastusumab responsive
Basal-like	ER-, PR-, HER2-	EGFR+ and/or cytokeratin 5/6+, MKI67 high, endocrine nonresponsive, often chemotherapy responsive
Claudin-low	ER-, PR-, HER2-	MKI67, E-cadherin, claudin-3, claudinin-4 and claudinin-7 low. Intermediate response to chemotherapy.
HER2	ER-, PR-, HER2+	MKI67 high, Trastuzumabresponsive, chemotherapy responsive

ER: Oestrogen receptors, PR: Progesterone receptors, HER2: Human epidermal growth factor receptor-2, MKI67; Proliferation marker protein Ki-67EGFR: Epidermal Growth Factor Receptors. (+); positive, (-); negative.

Table 1-3: Molecular classification of breast carcinoma (Holliday and Speirs, 2011).

The luminal category is characterised by endocrine receptor expression (estrogen - ER and progesterone receptors - PR), which was subdivided into luminal A and luminal B based on the expression of HER2. The second group was called the basal-type because cancerous cells' gene profiling is similar to the cells located outside the breast duct (basal cells). This category (basal) lacks the expression of hormonal receptors as well as HER2 so it is known as triple-negative breast cancer (TNBC), which has the poorest prognosis (Holliday and Speirs, 2011).

1.2.3.1 Luminal-type breast cancer

1.2.3.1.1 Luminal A and Luminal B

Luminal epithelia cells can be differentiated from basal epithelia cells by immunohistochemistry detection of keratin 8/18 (Perou et al., 2000). Furthermore, altered activity of antigen Ki-67 (MKI67) index has been shown to differentiate luminal breast cancers, with lower expression in luminal A compare to luminal B (Goldhirsch et al., 2011) Hence, the cell proliferation activity is lower in luminal A than other phenotypes of breast tumours (Cheang et al., 2009). Both luminal A and luminal B are ER-positive-luminal subtypes, but luminal A also exhibited high expression of GATA binding protein 3, X-box binding protein-1 and estrogen regulated LIV-1 transcription factors by hierarchical clustering (Sørli et al., 2001). The expression of HER2 differentiates luminal B from luminal A, however MKI67, cyclin B1 (CCNB1) and Myb-related protein B (MYBL2) are also proliferation signature genes for luminal B. The high expression of these genes in luminal B tumours clarifies the main difference between the two subtypes (Cheang et al., 2009). More-over increased expression of HER2 associated genes (HER2 and GRB7) were also observed

in luminal B subtype. The high cost of gene expression profiling makes it difficult to differentiate between the two luminal subtypes in the clinic but the MKI67 labelling index, which is based on immunostaining, can serve as a valuable labelling biomarker for the luminal B subtype (Cheang et al., 2009). The luminal B subtype is a worse type of breast cancer with poor outlook. Only 30% of luminal B tumours are HER2 positive, so a single marker is not sufficient to differentiate between luminal subtypes (Cheang et al., 2009).

1.2.3.1.2 HER2 enriched breast cancer

The ErbB family has four receptor tyrosine kinases members; HER1 (EGFR, ErbB1), HER2 (Neu, ErbB2), HER3 (ErbB3), and HER4 (ErbB4). The ErbB family is involve in triggering the cell proliferation and cell growth by binding to extracellular domain of tyrosine kinase receptors (O'Neill et al., 2012). The HER2 gene product (protein) has predictive and prognostic significance in invasive breast cancer (Seal et al., 2012). Approximately 15%-25% of breast tumours expressed high levels of HER2, yet are associated with poor prognosis (Seal et al., 2012). The PI3K pathway is most common up-regulated pathway in breast cancer. Among mechanisms of PI3K enhancement, PI3K3CA mutations are only observed in HER2+ breast cancers and are potentially a good prognostic biomarker (Mukohara, 2015). Targeted therapy of HER2 positive breast cancer by Trastuzumab; a monoclonal antibody and Lapatinib; a tyrosine kinase inhibitor, has increased survival significantly. The high expression of HER2 receptor in HER2 positive breast cancers activate cell proliferation (Paul and Mukhopadhyay, 2004), so early detection of the HER2 positive cancer is essential to prevent unnecessary cell growth.

1.2.3.2 Basal-type breast cancer

1.2.3.2.1 Basal-like (Basal A)

Triple negative (ER, PR and HER2 negative) breast cancers are the focus of much current research among the five different phenotypes because of poor clinical outcome. Basal-like tumours account for up to 15% of all breast cancers and mostly affect younger patients (Alluri and Newman, 2014). Histologically, the majority of basal-like breast cancers are Invasive ductal breast cancers of no special type (IDC-NST) type, with high histological grade, high mitotic indices and conspicuous lymphocytic infiltration (Livasy et al., 2006, Fulford et al., 2006). However, not all basal-like cancers are of the IDC-NST type; metaplastic carcinoma (Weigelt et al., 2009), adenoid cystic (Weigelt et al., 2008), secretory (Laé et al., 2009), and myoepithelial carcinomas (Azoulay et al., 2005) of the breast also show a basal-like phenotype. There is still no internationally accepted characterisation for basal-like breast cancers and how best to describe these tumours is a matter of controversy and ongoing debate (Telli, 2016). However, the proposed immunohistochemistry approach can be used to differentiate basal-like breast cancer by characterising the higher expression of cytokeratin 5/6 or lower expression of cytokeratin 8/18 or lack of expression of ER, PR, and HER2 (Nielsen et al., 2004). TP53 gene mutations are observed in up to 85% of breast cancers (Olivier et al., 2006), and alterations of the retinoblastoma protein (pRB) and cyclin-dependent kinase inhibitor 2A (CDKN2A), the G1/S cell-cycle checkpoint are remarkably dominant in basal-like breast cancers (Moore, 2016, Leidy et al., 2014). A recent study conducted by Moore et al 2016, showed that basal-like breast cancers lack pRB expression, and overexpress CDKN2A and p53 based on immunoreactivity (pRB-/CDKN2A+/p53+), whereas this immune-profile was rarely observed in

tumours of other molecular subtypes (Moore, 2016). Epidermal growth factor receptor (EGFR or HER1) and cytokeratin 5/6 are immunohistochemical biomarkers, used to identify the basal-like phenotype (Nielsen et al., 2004), but these markers are not specific because basal/myoepithelial cells of normal breast, almost uniformly express high-molecular-weight/basal cytokeratins (CK5/6, CK14, and CK17) (Mayer et al., 2014). Therefore, there is a need for a precise biomarker for diagnoses and prognoses of basal-like breast cancer.

1.2.3.2.2 Claudin-low (Basal B)

Gene expression studies of breast cancer provide better molecular understanding of the breast cancer and one of the latest molecular subtypes defined is claudin-low (Herschkowitz et al., 2007). Claudin-low tumours only represent 10% triple-negative breast cancers and are associated with young age of onset, larger tumour size and therefore higher tumour stage/grade, extensive lymphocytic infiltrate and a poorly defined tumour margin (Dias et al., 2017, Prat et al., 2010). Patients diagnosed with claudin-low breast cancer had a worse overall survival rate compare to patients with luminal breast cancer but interestingly, claudin-low cancers had low local recurrence rate in the area of the breast after treatment (Perou, 2011). Claudin-low subtype is also characterised with low expression of ER alpha, PR, HER2 and proliferation marker Ki67, but in addition by down-regulation of claudin-3, claudin-4, claudin-7 and E-cadherin; a glycoprotein, involved in calcium dependent cell-cell adhesion (Prat et al., 2010). A large-scale study, conducted recently by Dias et al 2017, on formalin fixed paraffin embedded 1600 human breast cancer tissues, showed typically negative expression of ER, PR, HER2, claudin 3, claudin 4, claudin 7 and E-cadherin in 90% claudin-low breast cancer (Dias et al., 2017).

1.3 Methods for detecting Breast cancer

1.3.1 Self-examination

Self-examination provides an effective way to encourage early detection of potentially harmful lumps that does not involve any specialist equipment or professional health care assistance (Güth et al., 2008). Self-examination can also encourage women to be involved in the maintenance of their own health, detecting up to 50% tumours and improving discovery of other breast-related diseases (Güth et al., 2008). Despite the advantages of early cancer diagnosis, engagement in breast self-examination by women is still low, particularly in educationally-deprived and developing countries (Hacihasanog˘lu and Gzm, 2008, Lechner et al., 2004).

1.3.2 Molecular imaging for breast cancer

1.3.2.1 Mammography

Mammography is by far the most common approach for screening breast cancer. In Europe, women aged 50 to 70 are invited for breast screening every two to three years, while in US annual mammograms are recommended from 50 to 74 years equating to nearly 39 million images in 2014. Women who have an increased risk of developing breast cancer, but are too young to join national screening programs, are offered annual screening using mammography and magnetic resonance imaging (MRI) based on a risk-benefit decision. Sensitivity of mammography can vary from almost 100% to 40% dependent on tissue composition (Drukteinis et al., 2013, DeSantis et al., 2011), and positive association for prediction of disease is only 50%. False-positive

rates in breast cancer screening are a significant limitation, as high call-back rates and unnecessary biopsies result in increased cost, radiation dose, and patient anxiety during re-screening (Stout et al., 2014, Loberg et al., 2015). Over a period of 20 years, based on 7 trials of more than 600,000 women, screening was shown to reduce breast cancer mortality by 15%, however over-diagnosis was estimated at 29% (Gotzsche and Jorgensen, 2013, Nelson et al., 2009). In particular, mammography is around 10% less sensitive in women under 50 than in postmenopausal women (Mushlin et al., 1998). Women in their 40s screened within the UK Age Trial had a 5% risk of a false-positive result at their first screen (Johns et al., 2010) and a random clinical trial of 161,000 women showed that women aged 39 onwards did not significantly reduce their risk of dying from breast cancer within 5-15 years (Moss et al., 2006).

The continuing development of mammography, including digitisation with improved resolution (full field digital mammography), has increased its position as the gold standard for screening but still requires the presence of a substantial mass in the breast for a diagnosis (Obenauer et al., 2002, Pisano et al., 2005). Many breast tumours may have already metastasised before detection by mammography. Equally, a mammogram cannot distinguish between DCIS that will remain benign (therefore not requiring treatment) and those that will progress to a malignant invasive tumour (Espina and Liotta, 2011). Furthermore, radiation exposure has been shown to contribute to an increased incidence in breast cancer, particularly in high risk populations (Drukteinis et al., 2013). Consequentially, some women receive a cancer diagnosis even though their cancer would not have led to death or sickness, resulting in unnecessary surgery to remove the lump or whole breast. At worst, over-diagnosis can lead to unnecessary treatment, and some healthy women

dying needlessly from heart disease induced by radiotherapy, with the overall conclusion that screening did not reduce total mortality.

One of the challenges for imaging methods is breast tissue density, which is higher in younger women, making their mammograms harder to interpret, whereas postmenopausal women with extensive fatty breast tissue or ductal atrophy are more likely to have an accurate diagnosis (Martin and Boyd, 2008). Breast density is strongly associated with tissue composition (collagen, epithelial cell and non-epithelial cell content, glandular area), genetic influences and hormonal regulation (prolactin, IGF-1). Furthermore, mammographic density is positively associated with alcohol consumption, smoking in postmenopausal women and with breast cancer (Martin and Boyd, 2008, Vachon et al., 2007). It is possible to quantitatively measure breast density using established and specialised imaging modalities in combination with bespoke algorithms (Yaffe, 2008), however this strategy has not been widely accepted.

1.3.2.2 OTHER IMAGING TECHNIQUES

Digital breast tomosynthesis (DBT) is a relatively new screening approach, which uses an X-ray beam in an arc around the breast to provide a 3D reconstruction of the tissue. In a comparison of DBT with digital mammography for 2,666 breast lesions, the former had improved performance (sensitivity and specificity of 90% and 79%) compared to the latter (89% and 72%, respectively)(Lei et al., 2014), but was also subject to the same limitations incurred by breast density. There were fewer recalls with DBT compared to digital mammography, particularly in younger women, and although approved by the food and drug administration (FDA) for breast screening, it is not widely available (Haas et al., 2013). Ultrasonography, using high-resolution linear

transducers, is readily accessible, similar in cost to mammography and moderately improves cancer detection, but has lower specificity, lower positive predictive value and was similarly obfuscated by tissue density (Giuliano and Giuliano, 2013).

Contrast-enhanced MRI, by comparison is not limited by breast density, nor does it use ionising radiation. In a review of 11 studies comparing MRI with digital mammography, the former achieved 94% sensitivity, but identified few cases of cancer in addition to those discovered by digital mammography alone (Warner et al., 2008). Furthermore, MRI exhibits relatively low specificity, is not as accessible and is expensive to run. Positron emission tomography with computer tomography (PET-CT) uses gamma-radiated emitting fluorine-18 fluorodeoxyglucose (FDG), a glucose analogue, injected into arm veins to accumulate in areas of high metabolic activity, such as tumours, and is then detected using a PET-CT x-ray scanner (Avril et al., 2000). Sensitivity of PET-CT ranged from 71% and 90%, which was improved further when combined with mammography, but results were affected by breast size (Schirrmeyer et al., 2001). However, PET-CT is more often used to assist surgery to remove diseased tissue rather than as a front-line routine screen for initial detection of breast cancer.

Wave elastography uses ultrasound or MRI techniques to measure tissue stiffness, which is positively associated with a tumour mass compared to normal tissue. The application of shear wave elastography, which uses the force of acoustic radiation produced by an ultrasound beam, demonstrated improved sensitivity and specificity (Krouskop et al., 1998). Electrical impedance scanning is another modality under development for breast cancer detection especially in young women with dense breasts, which is based on lower

electrical impedance in malignant tumours compared to the surrounding normal tissue (Hope and Iles, 2004). Recent technological advances have facilitated dynamic thermal analysis of the breast by recording of circadian rhythm variations and analysing the recorded data using highly sophisticated statistical algorithms, but the approach currently suffers from high false positive rates (Salhab et al., 2006). Although showing great promise, none of these innovative imaging approaches are in a position to replace mammography as the mainstream approach to routinely detect breast cancer and most importantly cannot differentiate between benign and malignant growths. Furthermore, the instrumentation for these types of screening is expensive and not readily available in all hospitals. The current molecular imaging techniques have been used for detection of breast cancer in clinical practise is summarised in Table 1-4 (Specht and Mankoff, 2012).

Despite this progress, there are still significant challenges, (i) detection of the earliest stages of cancer (differentiating between DCIS that will remain benign or become malignant) to enable early intervention for triple negative breast cancer patients (for whom there is no targeted therapy, poor prognosis, lower survival rate and higher chance of re-occurrence), (ii) avoiding unnecessary preventative surgery (particularly for BRCA-1/BRCA-2 mutation carriers) and (iii) identifying those patients who will not respond or become resistant to chemotherapy and hormone therapy.

Modality	Indication	Sensitivity	Specificity	Advantages	Disadvantages	Reference
Self-examination	Tumour detection	53.9%	54–59%	A free and easy way to look for cancer and reduce mortality,	Some breast lumps can be missed, cause unnecessary distress	(Güth et al., 2008)
Mammography	Tumour detection	73-86%	88-93%	Inexpensive, highly portable and does not necessarily require a contrast agent	Discomfort, limited depth penetration, challenging spatial localization, and radiation exposure, False positive and false negative results	(Smith et al., 2012)
Ultrasound, especially with contrast enhancement	Detection Tumour characterization	61.4%	82%	Highly portable, inexpensive Molecular microbubble agents possible	Operator dependence Contrast agents confined to vascular system.	(Houssami and Turner, 2014)
Magnetic resonance imaging (MRI)	Tumour characterization	77-99%	81-99%	Quantification of tumour perfusion and tumour capillary permeability	Confined space, Contrast design limited by need for magnetic atom.	(Bleicher and Morrow, 2007)
Magnetic resonance Proton spectroscopy	Tumour characterization	91%	93%	Can measure wide range of molecules No contrast necessary	Limited spatial resolution Challenging to obtain high-quality spectra in routine imaging	(Berg et al., 2006)
Positron emission mammography	Detection Tumour characterization	71-99%	66-69%	More sensitive for smaller tumours Higher spatial resolution	Increased radiation dose Visualization of posterior lesions Variable uptake of 18F-fluorodeoxyglucose (FDG) in small and less metabolically active tumours	(Sachelarie et al., 2005)
Positron emission tomography	Detection Response evaluation characterization	64-96%	73-99%	Wide range of molecular imaging probes Tracer imaging without perturbing biologic system	Limited spatial resolution (improved with use of non-contrast computed tomography) Some radiation exposure	(Specht and Mankoff, 2012)
Core biopsy	Detection, Tumour characterization	91.5%	87.7%	Differentiating benign and malignant,	Discomfort and painful because a surgical procedure required, Risk of complications like infection and bleeding.	(Bianchi et al., 2015, Yamaguchi et al., 2015)

Table 1-4: Methods for detecting breast cancer.

1.4 Biomarkers

A biomarker is defined as “A characteristic that is accurately measured and evaluated as an indication of normal biologic developments, pathogenic processes, or pharmacologic effect to a therapeutic intervention” (Aronson, 2005). The utility of a biomarker lies in its ability to provide an early indication of the disease, to monitor disease progression, to provide ease of detection, and to provide a factor measurable across populations. There are three functional types of biomarkers; (1) diagnostic, (2) prognostic and (3) predictive. Diagnostic markers are used confirm the presence or absence of disease, prognostic markers help to define a suitable clinical treatment for the disease and predictive markers forecast the likely outcome of treatment or disease progression in the absence of treatment (Azad et al., 2006).

1.4.1 Biomarkers of breast cancer

There have been extensive studies to identify breast cancer biomarkers, but with varying degrees of success. Genomics has already stratified the disease to identify high risk individuals and histopathological approaches are used to determine ER, PR and HER2 expression, helping to direct clinical intervention. A key challenge in breast cancer clinics is the resolution of tumour heterogeneity, particularly at the molecular level. The investigations, over the past few years, have identified key genes (e.g. BRCA-1 and BRCA-2) and specific proteins (e.g. ER, PR and HER-2) as markers for selection of therapy, through genomic techniques (Galvão et al., 2011). Differentiation of ER, PR and HER2 expression status, have played a valuable part in the prognosis and identification of the different subtypes of breast cancer. For example, women

diagnosed with estrogen receptor positive (ER+) and progesterone positive (PR+) tumours have a better prognosis and high response to hormonal treatment, using tamoxifen and toremifene, than those who are estrogen receptor negative (ER-) and progesterone receptor negative (PR-). An example of a biomarker in breast cancer is estrogen receptor that is expressed by some sub-groups of breast tumours. Various anti-estrogens molecules; tamoxifen or aromatase inhibitor (like Anastrozole) have been developed to target this receptor for breast cancer treatment (Ariazi et al., 2006).

Indeed, microarray-based technologies for routine prognostic screening of multigene signatures (for example, 70-gene MammaPrint for tumour aggressiveness/chemotherapy requirement /metastatic prognosis, Celera 14-gene metastasis score, Oncotype-DX 21-gene signature for measuring risk of re-occurrence, and 76-gene Veridex signature for tamoxifen therapy benefit (Zhang et al., 2009)) have been invaluable in supporting treatment of ER/PR positive breast cancers. However, the identification of a specific biomarker for the detection of breast cancer at the earliest stage that can be analysed in biopsies, has so far evaded the diagnostic industry. Proteins and metabolites have been discovered which were increased in malignant tissues compared to normal, but were subsequently found to be diluted beyond the level of detection in plasma or urine, or were found not to be specific to breast cancer (Brooks, 2009). In addition, depletion methods may be required to remove the most abundant proteins such as albumins and immunoglobulins from plasma, prior to analysis, which can result in losses and further reduced sensitivity. Suffice to say, a comprehensive review of breast cancer biomarkers in 2007 by the update committee of the American Society of Clinical Oncology failed to recommend any of the most promising candidates, including blood

levels of CA 15-3 and CA 27.29 (both forms of mucin-1) for diagnosis, detection of recurrence, decisions on therapy or metastasis, or circulating truncated extracellular HER2 for detection of breast cancer (Harris et al., 2007). Clearly, accessing tissue biopsies to look at tumour-associated biomarkers, where they are most concentrated, on a regular basis is not practical, though most of the accepted biomarkers ER, PR, HER2, uPA and PAI-1 are currently analysed in tissues by immunohistochemistry or ELISA (Harris et al., 2007). Hence, the challenge remains, how to provide a low cost, safe, simple, sensitive and specific method for detecting breast cancer, early enough, to improve the direction of current treatment regimen and avoid extreme preventative intervention such as elective bilateral mastectomy. In this context, molecular diagnostic approaches for the early detection of breast cancer remain largely untapped.

By far the best biofluids for developing a screening diagnostic are those readily accessible and closely aligned with the disease area, such as saliva for oral cancers or urine for bladder or renal cancer. In this respect, various methods have been developed to access ductal liquids that are associated with cells that are the origin of most breast cancers (see Table 1-5, and section 1.6 for further details on biopsy materials).

1.5 Proteomics

The term 'Omics' describes all the biological constituents of a system and common examples are genomics, epigenomics, transcriptomics, proteomics, metabolomics and lipidomics. The main aim of these is to characterise the profile of their given area in a comprehensive and unbiased way. These 'omics' approaches are based on whole biological system which defines the whole process from which DNA is transcribed to mRNA then translated to proteins, which in turn affect almost every omic system (Crick, 1970, Hegde et al., 2003). In the case of cancer, omics helps to differentiate patterns in biomolecules that are characteristic of neoplasia. Understanding these characteristics is important to identify the main differences between normal and cancerous cells, which can be used to target cancers or oncogenic mechanisms.

. Next-Generation Sequencing (NGS) allows sequencing of DNA and RNA much more quickly and cheaply, with high-throughput sequencing (van Dijk et al., 2014). The most common NGS technologies include; Illumina (Solexa) sequencing, Roche 454 sequencing, Ion torrent: Proton / PGM sequencing and SOLiD sequencing (Metzker, 2010). A single nucleotide alteration can induce altered translation, and can cause major changes to the function of a protein, which may have an important role in tumour suppression. For example, tumour protein p53 (TP53), has more than 2500 known mutations of which 52 have been identified in cancers (Leroy et al., 2013). TP53 mutations has been shown to be an important player in human cancers; activate angiogenesis and DNA repair proteins when DNA has sustained damage, cell proliferation, and inhibition of apoptosis (Leroy et al., 2014). Genomics also helps profile the effect of these mutations on promoter or enhancer regions and other noncoding DNA, by explaining the genetic cancer risk

factors, cancer screening and prevention, and targeted therapy, fulfilling the potential of personalized medicine(Weitzel et al., 2011).

The expression of total RNA molecules in a tissue or cell is characterised by Transcriptomics and is a useful approach to differentiate the mutated gene expression in a tumour (Hegde et al., 2003). The physio-chemical nature of mRNA makes the specific profiling of transcripts a rapid and reproducible method. Microarray technology, such as Affymetrix GeneChip microarray allows large scale investigation of thousands of transcripts simultaneously (Eddy et al., 2010). The shotgun sequencing 'RNA-seq' is a deep sequencing approach and presents the opportunity to obtain high-resolution total transcriptome characterisation (Maher et al., 2009). Protein translation process is controlled by numerous RNA binding factors, processing bodies, RNA secondary structures and ribosomal mechanisms (Kozak, 2007). So protein expression profiling does not directly correlate with transcript expression, presenting a need for the direct measurement of relative protein expression (De Godoy et al., 2008). Hence, to fully understand the phenotype of a cell, investigation of protein expression is very important.

Proteomics is the analysis of a complete set of proteins synthesized by a cell under a given set of physiological or developmental conditions (Anderson et al., 2000). The final product of a gene has even more complex functionality compared to the gene itself, so investigation of proteins is essential. Moreover, only through proteomics, can post-translational modifications be determined which have a diverse range of effects on the function of proteins. The emergence of proteomics is based on; developments in mass spectrometry, internet-based DNA and protein sequences databases and computer algorithms for database searching using mass spectrometric data (Lamond et al., 2012). A range of different disciplines such as

molecular biology, biochemistry and bioinformatics are required to achieve this ambitious target. The value of proteomics has broader applications: in medical research, drug discovery, forensics science, food quality control and microbial interactions Figure 1-2 (Lamond et al., 2012).

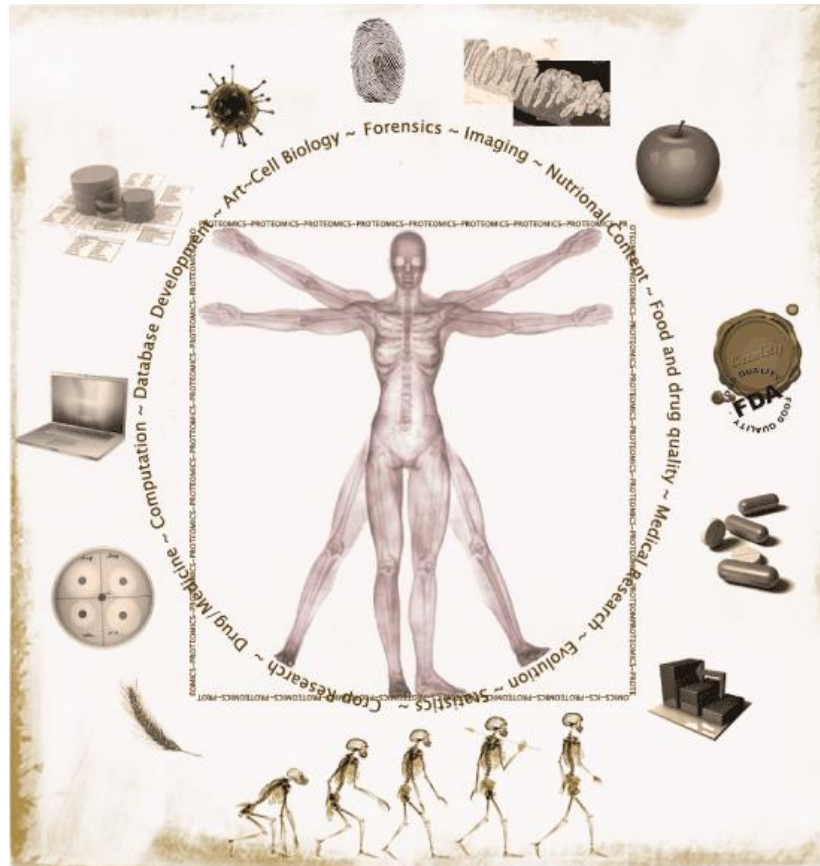


Figure 1-2: The significance of Proteomics on human life.

At the protein level, due to post-translational modifications (e.g. acetylation, glycosylation, phosphorylation), alternation in RNA splicing or proteolytic processing the 20,193 genes of the human genome inflates to expression of 500,000–1,000,000 variations (Galvão et al., 2011). Hence, proteins indicate the precise genetic mechanisms and their effect on the microenvironment of the cell making them more accessible therapeutic targets (Galvão et al., 2011).

Proteomics enables two main strategies for biomarker identification; (a) discovery approach and (b) targeted approach. Most biomarkers are proteins,

therefore investigation of protein expression that reflect changes healthy and diseased cellular behaviour using proteomics provides the opportunity to identify new biomarkers (Matt et al., 2008). The discovery approach is quantitative and qualitative proteome profiling of appropriate tissues, cell lines and bio-fluids to observe as many changes as possible between normal and disease protein profiles (Matt et al., 2008). The targeted approach is based on traditional hypothesis of a biological rational for investigation of a specific protein or pathway. Strategies for discovering a novel protein biomarker of breast cancer are shown in Figure 1-3 (adapted from (Aronson, 2005, Azad et al., 2006, Matt et al., 2008)).

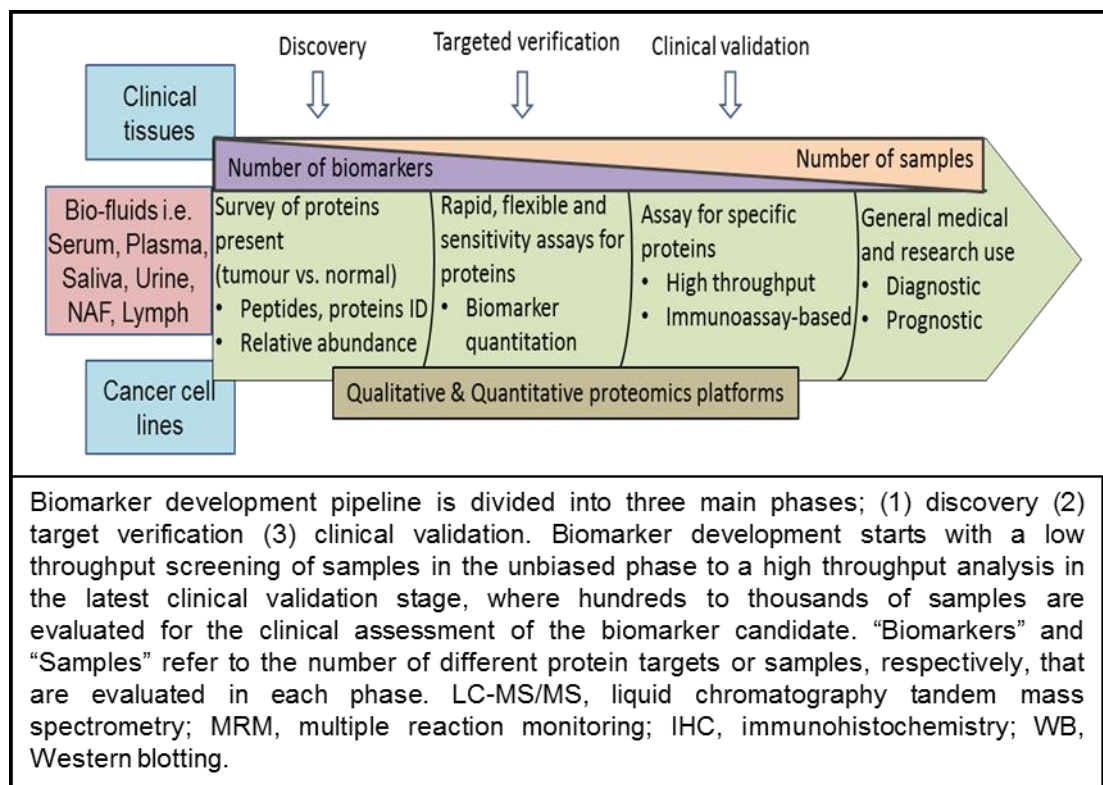


Figure 1-3: Strategies for discovering novel breast cancer protein biomarkers.

1.6 Proteomics and Mass Spectrometry

The profiling of the proteome is more challenging as compared to the genome, with potentially more than 5 times as many proteins than coding genes (Gstaiger and Aebersold, 2009). Proteins have enormous diversity due to different

combinations of the 20 amino acids from which they are built, amplified by many potential post-translation modifications. Where DNA and RNA investigation can be facilitated through their logarithmic amplification using methods such as the polymerase chain reaction (PCR), proteins don't have such properties and their detection is determined by the sensitivity of available proteomics methods and techniques (Altelaar et al., 2013). Most proteomics approaches are based on mass spectrometry, which allows precise mass measurements by MS and fragmentation by MS/MS of a peptide confident identity of a protein (Karas and Hillenkamp, 1988, Hunt et al., 1986). There are two main strategies for MS based proteomics; top-down, to characterise intact proteins and bottom-up, which first converts proteins into peptides by amino acid-specific protease digestion or chemical cleavage. The use of bottom-up proteomics for more complex samples, is called as 'shotgun proteomics', which is the conversion of 1000's of protein to 10,000's of peptides (Wolters et al., 2001). One of the most important challenges in proteomics is the separation of these complex peptide digests for characterisation, and is essential prior to sensitive and accurate analysis by MS (see section 1.6.3 for MS analysis).

The proteomics approach process consists of a number of steps; preparation of biological samples, separation of peptides mixture, ionization of peptides, acquisition of full spectrum, selection of specific precursor ions, fragmentation and identification of peptides or proteins through databases search (Figure 1-4) (Domon and Aebersold, 2010).

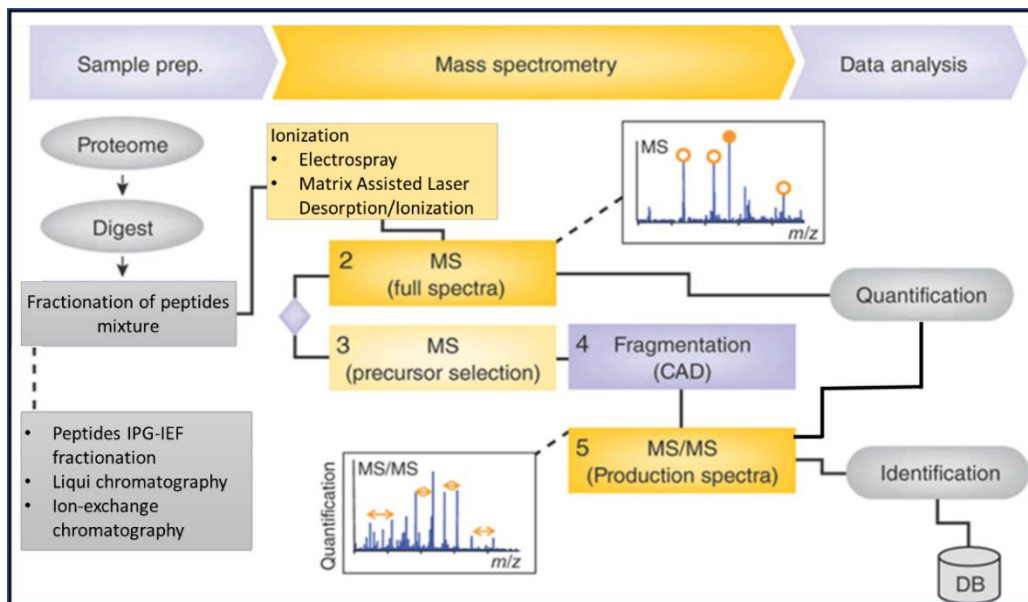


Figure 1-4: Workflow illustrating typical proteomic experiment (Domon and Aebersold, 2010).

1.6.1 Sample preparation

In bottom-up Proteomics, the sample preparation is very important and includes three steps; cell/tissue disruption, protein solubilisation in a buffer, and removal of interfering substances. A well-controlled procedure is required to extract high quality and adequate amounts of proteins from sample for proteomics analysis, to maximize reproducibility of data. Protein extraction can be performed on cultured cell lines and tissues by three methods; mechanical, chemical, enzymatic, individually, and/or in combination. Chemical based extraction methods involve the use of organic solvents or detergent and are very popular because of low cost and high protein yield. Detergents (for example, SDS, CHAPS, Triton X-100) disrupt the cell membrane by breaking the lipid-protein interactions and result in solubilisation of proteins in the extraction buffer. The main challenge of detergent-based extraction methods, are suppression of peptide derived signals in MS analysis by residual detergent (Wisniewski et al., 2009). Therefore, extraction buffer must have

optimised amount of detergent and/ or be followed by detergent removal step before MS analysis.

Mechanical extraction methods have an advantage over chemical methods, because they can keep the activity of the protein of interest but yield less protein amount compare to chemical based method (Moore et al., 2016). There are different type of mechanical methods such as; freeze-thaw (Kumar et al., 2004), cryo-pulverization (Moore et al., 2016), sonication (Rial-Otero et al., 2007), and electrical homogenizers or blenders (Bodzon-Kulakowska et al., 2007), but the choice of method depends on different factors; type of biological material, cellular location, stability and post extraction analysis.

The most common protein extraction method for whole proteome analysis, is a combination of detergent and sonication (Bodzon-Kulakowska et al., 2007, Wisniewski et al., 2009). In this assay, samples are first solubilised in extraction buffer (detergent and protease inhibitor; to prevent protein degradation), followed by probe sonication to disrupt the cells or tissues by high-frequency sound waves. These sound waves can increase the temperature of sample, so this process is conducted in multiple short bursts and an ice bath to maintain the samples at low temperature (Wisniewski et al., 2009).

1.6.2 Preparation of peptides mixture

After protein extraction, proteins are broken into peptides by chemical or enzymatic digestion in an approach called Mudpit (Multidimensional protein identification technology) or Shotgun proteomics. Extracted proteins are desalted using cut off filters or precipitation with organic solvents, to remove contaminants such as; salts, nucleic acids, detergent, and lipids (Feist and Hummon, 2015).

Protein samples are treated with a reducing reagent (e.g. β -mercaptoethanol or dithiothreitol (DTT)), to cleave disulphide bonds within and between protein subunits, result in opening of 3D structure of proteins. This step is further supported by alkylating reagents (e.g. iodoacetamide, or vinyl pyridines), to prevent reforming disulphide bonds (Bai et al., 2005, Righetti, 2006). After reduction and alkylation, proteins are broken into peptides (protein digestion) with chemicals or enzymes. The most common method of proteins digestion is based on trypsin, a protease with well-defined cleavage specificity. Trypsin cleaves at the C terminal side of the basic amino acid residues lysine (K) and arginine (R) unless the next residue is a proline (P) (Olsen et al., 2004). The MS and MS/MS analysis is conducted on the digested peptides and the information acquired then transferred to a search engine to identify and/or quantify the original proteins. Hence, careful sample processing at the protein and peptides level is the key for high quality MS data.

1.6.3 Fractionation of peptides mixture

Different fractionation or separation methods are used in proteomics, to improve proteome coverage and signal-to-noise and to reduce interference between peptides in MS analysis. These fractionation methods also provide some additional information about analytes, such as molecular weight, hydrophobicity or isoelectric point. The most common fractionation methods are; strong cation exchange (SCX) chromatography, peptide isoelectric focusing (pIEF) and high-pressure liquid chromatography.

1.6.3.1 Peptides IPG-IEF fractionation

Peptide isoelectric focusing (IEF) in immobilized pH gradients (IPG) fractionations can be performed in two ways: in-gel or off-gel, but in both

approaches, a peptide mixture is applied on to an IPG strip by rehydration and focused within the gel matrix by IEF, which separates peptides based on the isoelectric point (pI), a physicochemical feature of charged analytes. The fractionated peptides are extracted from wells of an IPG strip (off-gel) or excised sections of the strip (in-gel) (Pernemalm and Lehtiö, 2012, Cargile et al., 2005, Hörth et al., 2006). The most rapid effective off-gel separation system for state-of-the-art proteomics, is the Agilent Technologies 3100 OFFGEL Fractionator (Pernemalm and Lehtiö, 2012).

1.6.3.2 Liquid chromatography

Liquid chromatography (LC) provides a source of separation for proteins or peptides on the bases of their physicochemical properties. The separation is carried out using columns packed with chromatography beads with specific binding properties, within which different molecules have different levels of interaction, described as a retention. The two major components of LC; a stationary phase, the material with which the analytes can interact and the mobile phase, the continuously flowing solvent transporting the analytes (Di Palma et al., 2012). There are a number of types of chromatography; ion exchange chromatography, normal phase chromatography, reverse phase chromatography, flash chromatography, size exclusion chromatography and affinity chromatography (Snyder et al., 2011).

Reverse-phase (RP) LC is used in proteomics, which enables the separation of peptides based on hydrophobicity. The stationary phase is a hydrophobic resin, consists of long hydrophobic alkane molecules, typically 8 or 18 carbon atoms in length, termed C8 and C18, respectively (Di Palma et al., 2012). Digest peptides are injected onto the stationary phase, which cause partition of peptides into the stationary phase, resulting in retention. RP LC also provides

desalting capability because buffers are not partitioned on the stationary phase. The hydrophobicity of the mobile phase is then gradually increased as a 'gradient', which results in release of peptides into the mobile phase. The retention time of each peptide is therefore proportional to its hydrophobicity (Di Palma et al., 2012).

1.6.3.3 Ion-exchange chromatography

Ion-exchange chromatography (IEC), is based on electrostatic attraction between the analytes and the functional groups of the stationary phase, which results in the fractionation or separation of peptides according to differences in their charge (Phyllis, 1997). IEC has two types; (a) anion-exchange chromatography (SAX), positive functional groups have affinity for negatively charged peptides at basic pH and (b) cation-exchange chromatography (SCX), negative functional groups attract positively charged peptides at acidic pH. In IEC techniques, the elution of peptides is performed by changing the pH or ionic strength of mobile phases to disrupt the peptide interaction with the stationary phase. Strong cation exchange (SCX) chromatography with a salt gradient or incremental salt steps such as potassium chloride (KCl), is frequently used in peptides mass spectrometry, with combination of reversed-phase (RP) chromatography as a second dimension (Manadas et al., 2010, Essader et al., 2005).

1.6.4 Types of Mass spectrometry

Mass Spectrometry based characterisation of peptides is performed by several techniques, depending on the type of instrument, and means of ionisation. By definition, a mass spectrometer consists of three parts; (i) an ion source, (ii) a mass analyser that measures the mass-to-charge ratio (m/z) of the ionized analytes, and (iii) a detector that reads the number of ions with each m/z value. Matrix-

assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI), are the two common techniques to volatilise and ionise proteins or peptides for mass spectrometric analysis (Fenn et al., 1990, Karas and Hillenkamp, 1988). In the MALDI approach, samples are ionised via laser pulse, on a dry mixture of sample and crystalline matrix, while ESI ionises the analytes out of solution and is therefore coupled to liquid based separation tools such as liquid chromatographic instruments. ESI-MS systems (LC-MS) is normally used to analyse more complex samples, whereas MALDI-MS systems are preferred for the analysis of simple peptide mixtures (Walther and Mann, 2010). After ionisation, charged molecules are focused with lenses then separated with the mass analyser. There are different types of mass analyser, such as; Ion trap, Orbitrap, Quadrupole, time of flight (TOF), and the mass-to-charge ratio (m/z) determined with a detector (Walther and Mann, 2010).

1.6.4.1 MALDI-MS

Matrix Assisted Laser Desorption Ionization (MALDI) instrument is usually coupled with TOF analysers but instruments with quadrupole ion-trap and two types of TOF mass analysers are also available, to allow the fragmentation of MALDI-generated precursor ions (Krutchinsky et al., 2001). In MALDI-TOF/TOF instruments, two TOF sections are separated by a collision cell (Medzihradszky et al., 2000), whereas in the hybrid quadrupole instrument, the collision cell is placed between a quadrupole mass filter and a TOF mass analyser (Loboda et al., 2000). TOF or quadrupole analysers are used to select ions of a particular m/z , then fragment these molecules in a collision cell and generated fragments separated in TOF analyser. The MALDI-TOF/TOF mass spectrometer (Ultraflex TOF/ TOF, Bruker) consists of a MALDI ion source (King et al., 1995, Juhasz et al., 1996, Holle

et al., 1997), a high-resolution timed ion selector (TIS), a “lift” device for raising the potential energy of the ions (analytes), a further subsequent post-acceleration (source 2) to focus the ions, a post-lift metastable suppressor (PLMS), an energy focusing reflector, and fast ion detectors for the linear and reflector mode (Figure 1-5, adopted from Suckau et al 2003) (Suckau et al., 2003).

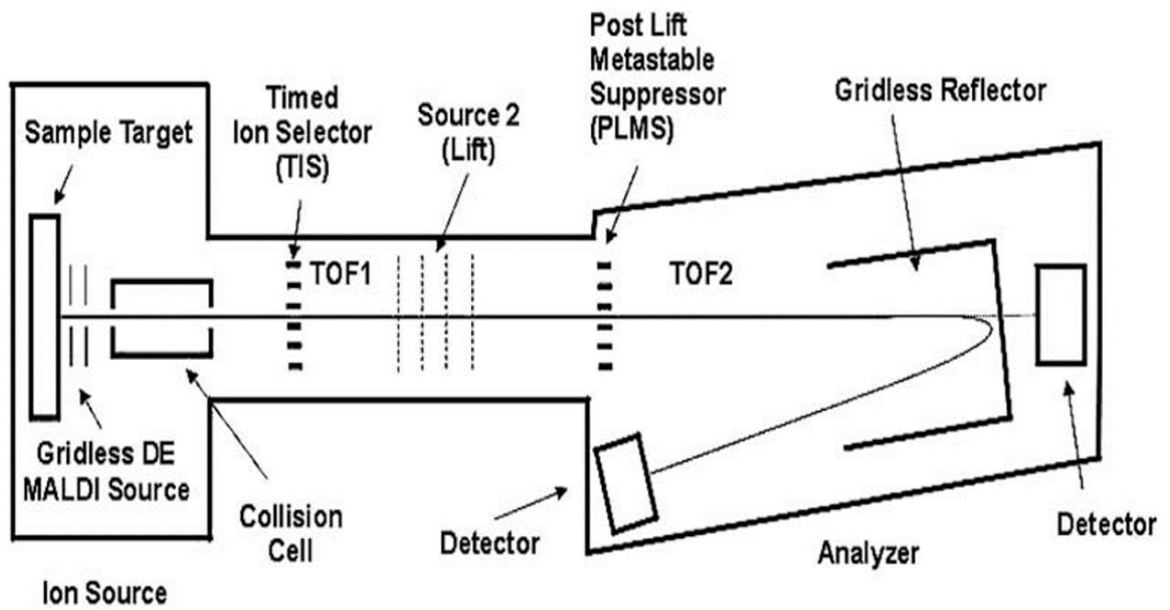


Figure 1-5: Schematic view of a MALDI-TOF/TOF mass spectrometer (Suckau et al., 2003).

1.6.4.1.1 MALDI-TOF/TOF ion source

The MALDI ion source accepts target plates of the same shape and size as microtiter plates. MALDI target plates (AnchorChip), equipped with 384 hydrophilic anchors on a hydrophobic surface, which enhance the sensitivity for peptide samples at least 10-fold in MS analysis (Schürenberg and Franzen, 2001, Schuerenberg et al., 2000). All AnchorChip plates have a bar code, which is read by the mass spectrometer, to specify for sample-specific analytical tasks. The AnchorChip targets are automatically introduced into the source chamber, where a vacuum system is maintained through a vacuum lock. The MALDI source has an image recognition system, to find the exact locations of the analyte spots. This feature allows automated MS and MS/MS analysis in the instrument without manual interference. However, MALDI-TOF/TOF is not directly coupled to liquid chromatography, so peptides must be deposited on a sample plate and mixed with the MALDI matrix, a process that has thus far proven difficult to automate MALDI-MS analysis.

1.6.4.1.2 Laser

The most common type of laser used in MALDI-TOF/TOF, operates at 200 Hertz. When the laser is fired on matrix crystals, causing ionisation of matrix (addition of a proton), then this proton is transferred to proteins or peptides, thus charging the analytes. A typical MS spectrum with good dynamic range and high signal-to-noise ratio can be achieved with 50–200 laser shots. In MS/MS mode typically 300-600 shots are required to get better spectra when analysing biological samples (Suckau et al., 2003).

1.6.4.1.3 MS mode for analytes acquisition

In the MS mode, the instrument operates as a MALDI-TOF spectrometer in linear, reflector, positive and negative modes. In linear MS mode, a gated MCP (micro channel plate) detector is applied to prevent detector saturation from low-mass ions and neutrals and increases the sensitivity MS, especially for high-mass proteins. In reflector mode, a fast dual MCP detector is used with micro channels (5µm bore diameter and a bias angle of 12°), to get symmetrical peak shapes for determining the exact mass of isotopic peaks in the spectrum by software algorithms such as SNAP (Sjoberg, 2014). The instrument is prepared for a full mass-range high resolution mode, from 500Da to 4000Da and lower mass range is limited by the detection system which enhances a constant signal width to the ion signal (Suckau et al., 2003, Han et al., 2008).

1.6.4.1.4 MS/MS mode for analyte identification

After full MS scan, acquisition conditions must be modified to generate high quality MS/MS scan; fragmentation of parent peptides. This is achieved by increasing laser power to provide a larger number of precursor and fragments ions per shot. Once these fragments ions are generated, the velocity is maintained with the corresponding precursor, so this 'ion family' (precursor and its fragments), will reach the timed ion selector (TIS) together. The TIS will only allow selected ion family to enter "LIFT" device and deflect all other ion families. The LIFT technology consists of three steps; (a) increasing the potential on the two adjacent grids forming the cell, (b) a focusing cell to modulate the speed of the ions, and (c) accelerating the ions at full speed by post-acceleration cell and time-focused onto the detector. The gridless two-stage reflector also plays its

role by focusing the divergent ion beam onto the detector. The high sensitivity and high signal-to-noise ratio can be achieved on MALDI-TOF/TOF by utilising “post lift metastable suppressor” (PLMS) and gridless reflector. Both features prevent undesired fragment ion formation after post-acceleration and the PLMS works as an ion deflector same like TIS before LIFT device (Jurinke et al., 2004, Suckau et al., 2003, Han et al., 2008).

1.6.4.2 ESI-MS

Over the last decade, electrospray ionisation mass spectrometry (ESI-MS) has become an important analytical technique in scientific laboratories. It provides a sensitive, robust and reliable method to investigate at femtomole quantities in micro-litre sample volumes that are impossible to analysis by other conventional techniques. Mass spectrometers coupled with high performance liquid chromatograph (HPLC), provides very powerful approach for analysing both small and large molecules of various polarities in a complex biological sample (Han et al., 2008). Automated sample introduction system; HPLC-ESI-MS, make this a technique for rapid analysis and high sample throughput.

1.6.4.2.1 Electrospray Ionisation Process

Electrospray ionisation (ESI) is carried out through electrical energy to assist the transfer of ions (analytes) from solution into the gaseous phase before they are subjected to MS analysis. The transfer of analytes from solution into the gas phase by ESI involves three steps: (a) dispersal of a fine spray of charge droplets, (b) solvent evaporation and (c) ion discharge from the highly-charged droplets (Desiderio, 2013, Fenn et al., 1990, Bruins et al., 1987). ESI source coupled with mass spectrometer, is maintained at a high voltage; 1.8 - 5.0 kV,

while a continuous stream of sample solution is passed through a stainless steel or quartz silica or glass emitter. A mist of highly charged droplets; positive or negative, are generated according to polarity of the capillary voltage. ESI process can be further improved by an elevated ESI-source temperature and using a nebulising gas; such as nitrogen, if a solution is being injected at a higher flow rate(Ho et al., 2003). These factors; electric field, elevated temperature and nitrogen gas, result in ejection of ions from charged droplet into the gaseous phase. Only charged ions are sampled by skimmer cone and analysed by mass spectrometer (Ho et al., 2003).

1.6.4.2.2 Orbitrap Fusion

The Orbitrap Fusion is a tribrid architecture mass spectrometer (Figure 1-6) with three mass analysers; quadrupole, linear ion trap, and Orbitrap, with multiple fragmentation techniques (CID, HCD, ETD) at any stage of MS_n, maximise the amount of high-quality data acquisition(Senko et al., 2013). Ultra-high-field Orbitrap mass analyser with resolving power up to 500,000 FWHM at m/z 200 and dual pressure linear ion trap mass analyser, are especially useful when dealing with complex and low-abundance samples in proteomics, glycomics, metabolomics, lipidomics and similar applications. Synchronous MS³ precursor selection (SPS) significantly increases number of peptides and quantification accuracy of iTRAQ or TMT experiments (Senko et al., 2013).

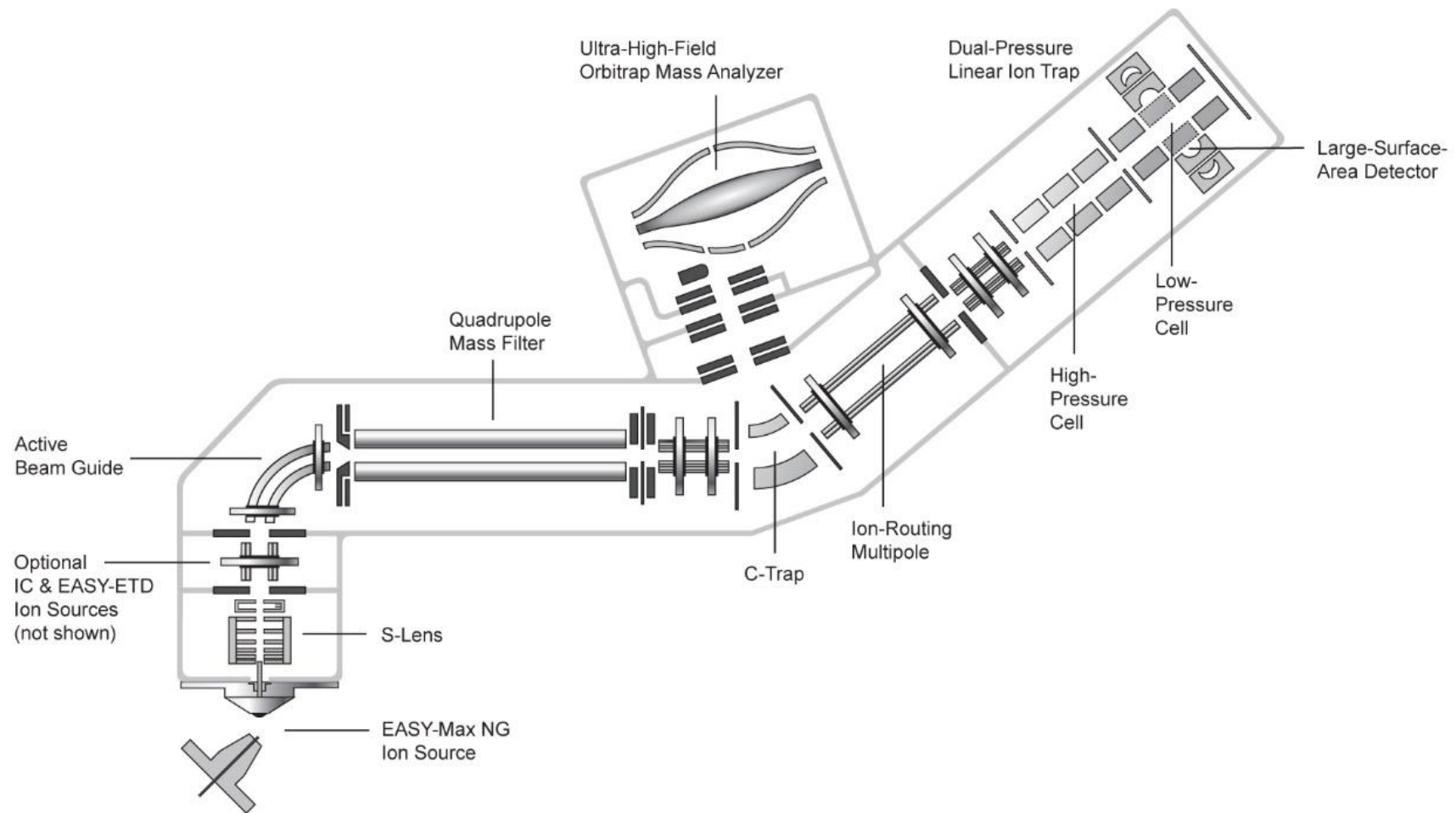


Figure 1-6: Schematic view of Orbitrap Fusion Tribrid mass spectrometer.

1.6.4.2.2.1 Quadrupole mass analyser

A quadrupole mass analyser is an assembly of 4 parallel metal rods and is kept at equal distance. When ions pass through these metal rods, they travel forward in the z direction with oscillatory motion in the x-y plane because an equal but opposite DC voltage and radio frequency (RF) AC voltage is applied to the diagonally placed pair of rods. The amplitude of oscillation contains a unique relationship with the m/z ratio and can be controlled in pre-fixed ratio, by changing the DC and RF voltages simultaneously. These DC and RF voltages values, allow only desirable m/z ratios (target ions) to travel along the z-axis without hitting the quadrupoles, while removing undesirable ions which are neutralised and fail to reach the detector (Ho et al., 2003, Marmet, 1971).

1.6.4.2.2.2 Ion trap mass analyser

Ion trap mass analyser consists of three hyperbolic electrodes: (a) the ring electrode, (b) the entrance end cap electrode, and (c) the exit end cap electrode, to form a cavity in which ions are trapped (stored). Ions travel through a small hole in the centre of both end cap electrodes and the ring electrode is located in the middle between the end cap electrodes. The ions are trapped and ejected from the mass analyser by applying various voltages. A 3-dimensional quadrupolar potential field within the trapping cavity is produced by ring electrode RF potential. This AC potential of constant frequency and variable amplitude stores ions in a stable oscillating trajectory within the trap cell. These trapped ions are ejected to detector system in order of increasing m/z ratio, by altering the potentials on electrodes. During MS/MS acquisitions, the precursor ion is selected inside the trap cell where an inert gas; helium is introduced for collision-induced dissociation (CID). After fragmentation of precursor ion, the product ions are ejected for detection or kept inside the trap cell for further CID

reactions (MS^n); to differentiate molecules with similar structures (Yoshinari et al., 2000, Makarov et al., 2006).

The main limitation of ion trap instrument is cycle time for a single MS/MS scan; how fast the ions are transferred out of the analyser to the detectors. But novel dual-pressure linear ion trap mass analyser in Orbitrap Fusion instrument, provides increased ion transmission because of a two-chamber ion trap mass analyser regulated at different pressures (Senko et al., 2013). Dual-pressure linear ion trap mass analyser, provides higher resolution, more efficient isolation and dissociation of ions, and much faster scanning rates than the basis linear ion trap instrument (Pekar Second et al., 2009).

1.6.4.2.2.3 Orbitrap Mass analyser

The orbitrap analyser is a member of Fourier transform ion cyclotron mass spectrometry (FTMS), based on Fourier transform ion cyclotron resonance (FTICR) principle (Comisarow and Marshall, 1974). In FTMS instruments, ions are trapped in a strong magnetic field combined with weak electric trapping plates, where they are excited by an oscillating electric field orthogonal to the magnetic field and detected, digitized, and converted in mass spectrum using Fourier transform into the frequency domain. The Orbitrap mass analyser has three electrodes; two outer electrodes and a central ring electrode. The outer electrodes have the shape of cups, facing each other and electrically isolated by a hair thin gap but secured by the central electrode (Makarov et al., 2006, Michalski et al., 2012). A linear electric field is maintained between the outer and the central electrodes by voltage current, to create purely harmonic oscillations. At the same time, the centrifugal force of the field strongly attracts ions to the central electrode but with Orbitrap configuration, the ions remain on a nearly circular spiral inside the trap cell, much like planets in a solar system.

In the Orbitrap Fusion, ions are stored in the C-trap after the quadrupole, then a high-voltage pulse is applied across the trap, each m/z being ejected as small packets between the outer and central electrodes through a specially machined slot in one of the outer electrodes. The special conical shape of electrodes, push the ion toward the widest part of the trap cell, initiating harmonic axial oscillations then outer electrodes are used as image current detectors for these axial oscillations. The digitised image current in the time domain is Fourier-transformed into the frequency domain to give a mass spectrum (Michalski et al., 2012, Zubarev and Makarov, 2013).

1.6.4.2.3 Triple quadrupole mass spectrometer

A triple quadrupole mass spectrometer consists of three quadrupoles (3 x 4 parallel metal rods), arranged in a linear fashion. A typical work flow on this instrument has following steps; (a) selection of analyte ion in first quadrupole (Q1), (b) fragmentations of the precursor ion in second RF-only quadrupole collision cell (Q2) by collision gas (usually argon) and (c) third quadrupole mass analyser (Q3) is used to monitor generated fragments to determine the quantitative information of the analyte ions (Hunt et al., 1986).

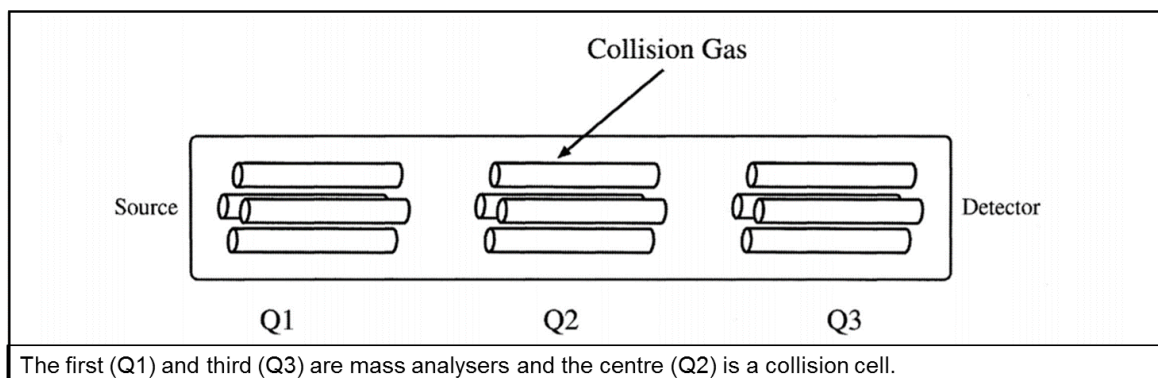


Figure 1-7: Schematic view of a triple quadrupole mass spectrometer system (Ho et al., 2003).

The process of fragmentation of precursor ion is known as collision-induced dissociation (CID). This tandem system is commonly denoted as “QQQ” mass spectrometer. When Q1 is applied for specific m/z ratio, it filters out all other precursor ions with different m/z ratios. This step is like a “purification” inside the mass spectrometer, eliminating time-consuming and complicated sample purification or enrichment procedures prior to MS analysis. The QQQ-MS system is commonly used for product ion scanning to study molecular structure of molecules drug metabolites (Hunt et al., 1986). Nowadays, triple quadrupole mass spectrometers are commonly used in clinical laboratories for multiple reaction monitoring (MRM) assay, to quantify disease related markers in complex biological sample (Ho et al., 2003).

1.6.5 Peptide fragmentation

Each precursor (peptide) is isolated by applying a mass filter and is then subjected to fragmentation. This process is called MS to MS/MS or MS^2 , which generates daughter ions that are representative of fragmentation across the peptide backbone (Washburn et al., 2001). Fragmentation of peptides is

performed inside the mass spectrometer by using a gas (nitrogen or helium) or a reagent (Fluoranthene). The resulting fragments ions have a mass difference corresponding to the residue masses of respective amino acids. If charge is retained on the N- terminal of peptide residue, the fragment ions are called a, b or c transitions and x, y, z transitions if charge is maintained at C-terminal (Figure 1-8). This information is subsequently used to identify the peptide, in a process known as peptide mass fingerprinting.

There are three common types of peptides fragmentations; Electron-transfer dissociation (ETD), collision-induced dissociation (CID) and high-energy collisional dissociation (HCD). ETD is performed in an ETD ion source (Fluoranthene 202 m/z), on the Orbitrap Fusion, which induces fragmentation of cations (e.g. peptides or proteins) by transferring electrons to them. Electrons are transferred via an ion-ion reaction between the peptide cation and the radical reagent anion (electron carrier) (Shin et al., 2003). HCD is performed by introducing nitrogen gas in the ion trap producing smaller peptides fragments (Olsen et al., 2007) while CID is performed with helium gas, produces large fragments of the peptide being analysed (Wells and McLuckey, 2005). ETD induces fission at the C-N-R bond of O=C-NH-R peptide bond while CID and HCD induce at the C-N bond of the O=C-N-H peptide bond (Figure 1-8) (Olsen et al., 2007, Wells and McLuckey, 2005). Using HCD or CID, mainly two fragments types are produced; the N-terminus fragment is termed a 'b' ion and the C-terminus fragment is termed a 'y' ion.

Under different fragmentation methods like Negative electron-transfer dissociation (niETD) and Electron detachment dissociation (EDD), fission can occur between the carbon of an amino acid and the carbon of the peptide bond

generating 'a' and 'x' ions or negative-ion electron capture dissociation (niECD) can induce between the N-C α bond of the peptide, generating 'c' and 'z' ions (Figure 1-8) (Senko et al., 2013).

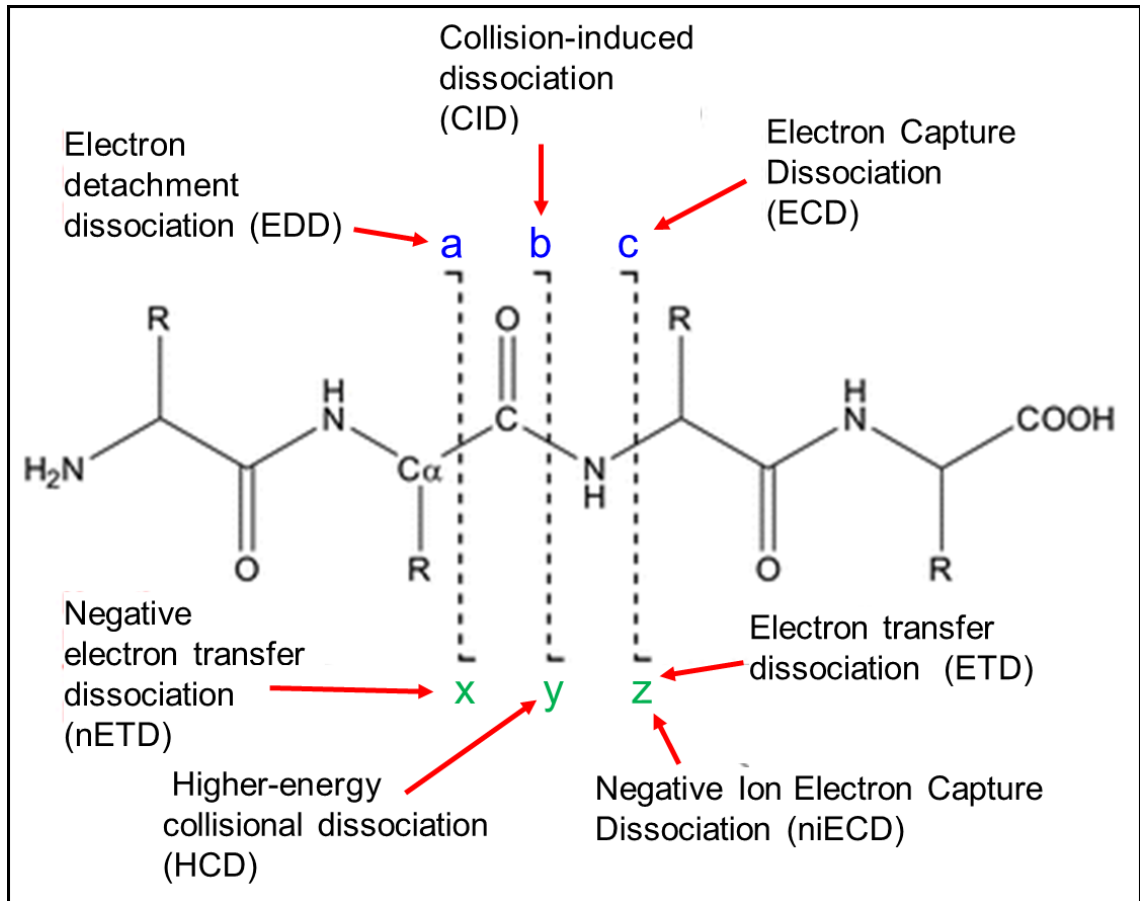


Figure 1-8: Peptide backbone fragmentation: peptide structure, selected fragmentation techniques, and fragment ion nomenclature (Zhurov et al., 2013).

1.6.6 Shotgun Proteomics

There are two main types of bottom-up proteomics; Data dependent acquisition (DDA) and data independent acquisition (DIA) (Doerr, 2015). DDA is the more common process, in which a full spectrum of all ionised peptides is obtained (MS) then all precursors are fragmented (MS/MS) in order of intensity, starting with the most abundant. This process can also be called as 'top N' or 'top S' where N is the number or speed of peptides characterised before another

full spectrum scan is performed and the process is repeated (Lopez, 2007). Another important feature for of DDA process is dynamic exclusion, which excludes those peptides which were fragmented in the previous MS scan, to focus on the less abundant and newly eluting peptides from the LC (Senko et al., 2013). DIA is newly emerging technique, where all MS and MS/MS data are acquired on mass analysers autonomously by scanning specific mass windows, then data is processed with DDA spectral libraries (Doerr, 2015).

1.6.7 Database searching

After MS analysis, the next step in the proteomics pipeline is to conduct a database search to identify the peptides and proteins. A number of parameters should be considered during database searching; for example, the nature of samples, organism, reduction, alkylation, and type of digestion (chemical or enzymatic) (Gundry et al., 2009).

A typical protein database has a collection of protein amino acid-based sequences in FASTA format, used to generate the in-silico peptide database against which the raw MS data file will be searched. So, it is very important for protein database to have sufficient protein coverage, otherwise the search engine will not find suitable match. SwissProt (Boutet et al., 2016), TrEMBL (Boeckmann et al., 2003), IPI (Kersey et al., 2004), RefSeq (Pruitt et al., 2007) and PRIDE (Vizcaíno et al., 2013) are commonly used protein databases for searching proteomic MS data. The protein database can be restricted at the species level (e.g. Homo sapiens) or class level (e.g. mammals), to get high confidence protein data. The peptide database is generated by the same enzyme, the way peptides were prepared prior to MS analysis, because of cleavage specificity of the enzyme, trypsin cleaves mainly the lysine or arginine

in the peptide chain, except when either is followed by proline. There is the possibility of missing the cleavage site with enzymatic digestion, therefore a missed cleavage value between 0-2 is also considered during database search. The main challenge for bottom-up proteomics is post-translational modifications; directly related to sample processing. The modifications can be either fixed; for example, carboxyamidomethylation of cysteine, which occurs after treatment with iodoacetamide or variable; for example, oxidation of methionine. These modifications increase the MS data processing time and false positive matches, which require considerable manual validations. There are software packages to deal with these issues such as Percolator (Brosch et al., 2009) and Fixed value PSM validator (Elias and Gygi, 2007). Instrumentation-specific parameters are also considered during database search; for example, mass and charge of precursors, fragmentations method (CID, ETC and HCD), mass accuracy for parent and fragments, masses acquisition is monoisotopic or average (Gundry et al., 2009).

MS data processing is performed through search algorithms (Mascot (Matrix-Science, 2016), SEQUEST (Diament and Noble, 2011), OMSSA (Geer et al., 2004), MS Amanda (Dorfer et al., 2014), Comet (Eng et al., 2013) and Morpheus (Wenger and Coon, 2013)) and the choice is orientated around ease of use, cost, features, accessibility, vendor compatibility, performance with certain data types, or a combination of all the above. These search engines (search algorithm) work by using a scoring algorithm and widely depends on type of data searched.

Mascot is a powerful search engine which uses MS data to identify proteins from protein sequence databases. There are a number of similar tools

available, but Mascot is unique in that it supports all type of MS instruments, fragmentations, and methods of searching; such as peptide mass fingerprint, sequence query and MS/MS ion search from both Fasta files and spectral libraries. The logarithm of Mascot is based on probability scoring; a simple rule to be used to define whether a result is significant or not. Mass values of peptides and or MS/MS fragment ions match to theoretical masses in database, are always treated on a probabilistic basis (Matrix-Science, 2016). The probability of the observed match is the total score and reported in $-10 \cdot \text{LOG}_{10}(P)$, where P is the absolute probability and significance threshold ($p < 0.05$) expected to occur at random with a frequency of less than 5%. If a match has probability of 10^{-20} , then Mascot score will be 200. The search parameters like mass tolerance and post-translational modifications, affect the significance threshold (Matrix-Science, 2016).

1.6.8 Target-Decoy peptide searching

Because both the MS spectra and theoretical spectra number many thousands, the probabilities of false discovery matches occurring by chance are very high. To fix this false discovery, a second 'decoy' search is performed which reverses the sequences of the target proteome and outlines the probabilistic score of those spectra matching by chance. The probabilistic scores of both searches; target and decoy, present an overlap which is frequently set at 1% overlap, termed a false discovery rate (FDR) of 1%. This can also be called as a 'q-value', and for a <1% FDR equates to $q < 0.01$. The process of target decoy searching is performed in some search engines; Andromeda, SEQUEST and MASCOT (Aebersold and Mann, 2016, Zhang et al., 2014b) based on different algorithms but the basic target decoy process is the same; assigning each PSM

with a false discovery rate. Then FDR of PSMs is defined by different ways but 'Percolator' is a semi-supervised machine learning for peptides identification, which is based on many features such as Mascot score, precursor mass error, fragment mass error, number of variable modifications, of the target and decoy PSMs providing a far more accurate and efficient method of separating the true matches in the data (Käll et al., 2007).

Several thousand MS/MS, acquired by shotgun proteomics experiment, must be searched to identify peptides and the original protein. This process is very simple for unique peptides because they map to single proteins. However, this process is more complicated for protein isoforms where many of the searched peptides match to two distinct protein sequences. The proteome of higher eukaryotic organisms has more redundancy, therefore a process of protein's grouping on the base of unique peptides is applied to get high probability of both the identified peptides and the aligned proteome (Nesvizhskii et al., 2003).

1.6.9 Quantitative Proteomics

The qualitative analysis of a protein present in complex biological sample is useful but quantitative information is more important for the perspective from the perspective of understanding a biological problem or discovery of biomarkers. The quantitative proteome approach allows the characterisation of disease samples quantitatively, relative to control or healthy proteomes. These quantitative experiments help to compare different biological hypotheses within two or more samples. There are two main types of quantitative experiments; label-based and label-free approaches. Label-free approaches based on data derived from separate LC-MS experiments, whereas

label-based approaches are based on combining quantitation into a single LC MS workflow (Bantscheff et al., 2012).

1.6.9.1 Label-free quantitation

Label-free quantitation is performed by using peak area or intensity of precursors, or peptides spectral counting. The relative intensity of comparable precursors is calculated on the base of mass and retention time of analyte (Bondarenko et al., 2002). Spectral counting approach works on the principle that there is a correlation between the number of peptides spectrum matching to a protein and the protein's abundance (Old et al., 2005). Label-free quantitation is least accurate means of proteins quantitation, requiring multiple replicate analyses to determine intra- and inter-experimental variation, therefore required more sample and instrument time (Bantscheff et al., 2012).

1.6.9.2 Label-based quantitation

There are two major types of label-based proteomics; incorporate stable isotopes into the protein during its synthesis in a cell and covalently modify peptides with isotope-coded tags as part of the proteomics workflow (Aebersold and Mann, 2016, Bantscheff et al., 2012). Peptides labelled with stable isotopes are quantified by using the intensity of precursors with MS1 while isobaric tag-labelled peptides are quantified by the intensity of reports ion at MS² or MS³ level. The other main purpose of labelling approach is to eliminate the variability seen when chromatography is performed iteratively (Bantscheff et al., 2012).

The advantage of Stable isotope labelling by amino acids in cell culture (SILAC), is that cell lines are grown under identical conditions. The only difference between two conditions, is supplementing one of the cell lines (most

commonly the control) with amino acids incorporating stable isotopes, typically arginine and lysine (Ong et al., 2002). The relative spectral intensities of stable isotope labelled and unlabelled peptides provide quantitative information between two cell lines (Ong et al., 2002). The application of metabolic labelling is limiting, especially in the context of primary human tissue (Aebbersold and Mann, 2016). The basic principle of SILAC is that the cell line under investigation is cultured with isotopic labelled amino acids, which are consumed by cell lines and incorporated into its proteome. Once fully labelled, the protein extract from a SILAC cell line is combined with that of another cell line; grown in normal medium. This enables the same preparative procedures on the two combined samples, eliminating quantitation inaccuracies introduced during sample handling. Finally, the proteome of two cell lines is acquired on same LC-MS analysis and mass shift introduced by the isotopic labelled and natural amino acids, provides the quantitative information between two biological conditions. But the SILAC approach has few limitations; (a) not straight forward for tissue biopsies and only possible for simple organisms like cell lines (Pan and Aebbersold, 2007), (b) only two conditions can be compared in a single experiment (Ong, 2012), (c) diet containing isotopic labelled peptides could effect on development, growth or behaviour of cells (Cutillas and Timms, 2010), (d) inaccuracies in quantification due to conversion of isotopic labelled arginine to proline in cells (Van Hoof et al., 2007).

Isotope-coded tags contain an amine reactive group, which modify the primary amines at either lysine residues or the N-terminus of peptides. These tags are called isobaric labels because they maintain the identical mass by different combinations of a reporter group and a balancing group (Ross et al., 2004). The identical peptides labelled with these tags maintain same

physiochemical properties and elute in same time from LC and consider as single precursor in MS analysis. This provides a single MS/MS spectrum of peptide for identification and relative quantitation. The two most common reagents are isobaric tags for relative and absolute quantitation (iTRAQ) and tandem mass tags (TMT) (Aebersold and Mann, 2016). iTRAQ offers the simultaneous relative quantitation of 4 and 8 samples, termed as 4-plex and 8-plex respectively, with quantitation occurring at MS² or MS³ level. iTRAQ 4-plex reagent is based on the relative intensities of reporter ions at 114, 115, 116 and 117 Da (Ross et al., 2004) while 8-plex has 4 further labels; 113, 118, 119 and 121 Da, in the low mass region of the MS/MS spectrum (Choe et al., 2007) (for further details about iTRAQ, see Chapter 3). iTRAQ based MS quantification has a number of advantages; (a) simple workflow for peptide labelling and data interpretation for relative quantitation, (b) quantitation of up to 8 samples simultaneously in a single LC-MS experiment and (c) high confident identification and quantification data by tagging multiple peptides per protein (Ow et al., 2009). On other side, isobaric workflows has a few limitations as well; cost, labelling inefficiencies and reporter ion ratio compression due to precursor ion co-isolation (Karp et al., 2010). The labelled peptides with similar mass-to-charge ratios and chromatographic retention times, result distorted ratios of reporter ions due to the inability to differentiate co-isolated precursors. Therefore, ratio suppression effect is more common due to co-isolation and co-fragmentation. New advances in MS-based approaches have been proposed to minimise or eliminate the effects of co-isolation. Some of these approaches are multi-notch MS³ fragmentation (see Chapter 3), traveling wave ion mobility separation, gas phase purification, improved chromatographic resolution and

optimisation of DDA settings (McAlister et al., 2014, Ow et al., 2011, Savitski et al., 2011, Ting et al., 2011).

1.6.10 Targeted mass spectrometry (MRM/PRM)

Both multiple reaction monitoring (MRM) and parallel reaction monitoring (PRM), are highly sensitive mass spectrometry based techniques to quantify selected analytes in complex samples. This assay is commonly performed on a mass spectrometer with quadrupole (triple quadrupole or Orbitrap) to select the target ion corresponding to the compound of interest, then a selected ion is fragmented to produce a range of daughter ions (Figure 1-9)(Harlan and Zhang, 2014, Wasinger et al., 2013). These daughter ions are then collected at the MS detector for quantification purposes. Only analytes of interest are isolated and all other ions that flow into the mass spectrometer ignored, giving high sensitivity, whilst maintaining exquisite accuracy. MRM assay was initially developed for quantification of small molecules but nowadays is also used for quantitation of proteins, peptides, metabolites and lipids from plasma, serum and other biological samples (Harlan and Zhang, 2014, Wolf-Yadlin et al., 2007). MRM or PRM assays have several significant advantages; (a) short assay development time (4 to 10 weeks), (b) highly multiplexed (from 10 to 100 peptides/proteins can be quantified in single experiment), (c) absolute quantification, (d) 100% specificity (because of unique peptides), (e) antibodies are not required for analysis, (f) same assay can be applied on any type of biological materials such as cell line, tissues and serum, and (g) minimal amount of sample is required (1-50 μ g) (Liebler and Zimmerman, 2013). The one disadvantage of MRM assays is the high cost of stable isotope labelled standard peptides (Percy et al., 2014).

A PRM assay is normally performed on hybrid Quadrupole-Orbitrap instruments. The selection of precursor and fragmentation is similar to MRM but with the high-resolution, accurate-mass (HRAM) Orbitrap detection system, which allows the measurement of all fragment ions of a given precursor in parallel (Figure 1-9)(Ronsein et al., 2015).

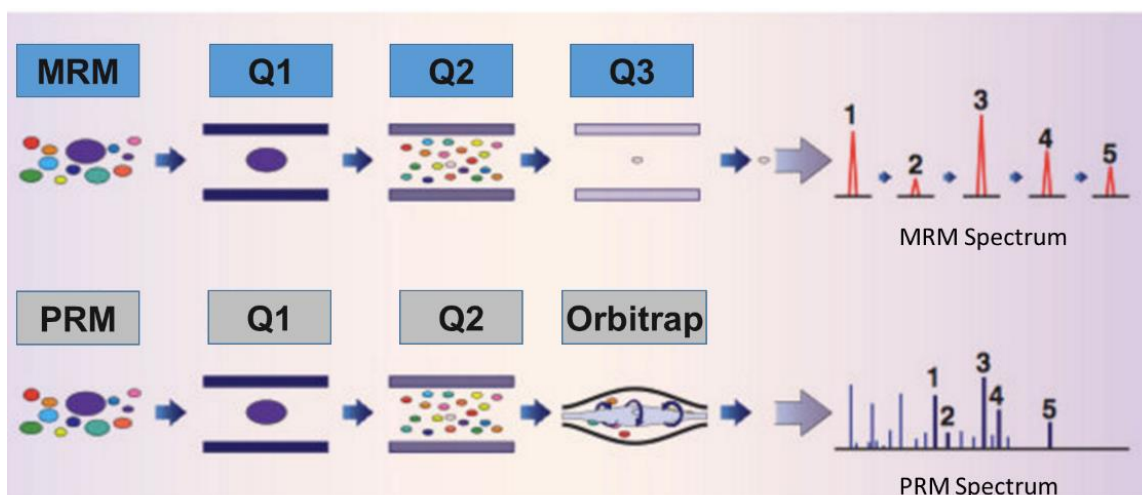


Figure 1-9: Schematic diagram of MRM/PRM.

MRM/PRM assay ideally involves signature peptides, which are used as a reference for the identification and quantitation of the target proteins. Initially standard peptides of target proteins are used to optimise LC-MS conditions and to create the calibration curves for absolute quantification (for further details for proteins/ peptides selection, light and AQUA peptides, see Chapters 5). MRM/PRM assays provide reliable and robust quantitation of peptides or proteins with good accuracy and precision (Lesur and Domon, 2015).

1.7 Biological samples for identification of breast cancer biomarkers

There are many strategies for identifying biomarkers and in the last decade, proteomics is one such approach that has been used extensively to identify changes in biological samples that correlate with breast cancer diagnosis. Proteome research of breast cancer has involved the analysis of a diverse range of samples, including tissues, cell lines and biological fluids such as saliva, serum, plasma, cerebrospinal fluid and nipple aspirate fluid, in order to understand the disease, identify diagnostic and prognostic biomarkers, and discover new targets for drug development. The biological samples used for biomarkers of breast cancer through proteomics are shown in Table 1-5.

Biological samples	Advantages	Dis-advantages	Research Methods	References
Tissue biopsies	Original disease source	Complex nature of biological materials.	TMA, LC/MS, IHC, WB, MRM RPPM	(Hondermarck et al., 2008, Gujral et al., 2012, Shaheed et al., 2013a, Gast et al., 2009)
Cell culture	Easy to use, relatively inexpensive, and reproducible, Ideal growth conditions,	Lack of extracellular components	WB, LC/MS, ELISA, IHC, MRM	(Shaheed et al., 2013a, Geiger et al., 2012)
Plasma	Minimally invasive, the samples are technically and psychologically easy to process	Abundance of Plasma proteins, sample preparation	LC/MS, ELISA, WB, MRM	(Hu et al., 2006b, Hanash et al., 2008)
Tear fluid	Accessible	Limitation of volume. dependent on fluid flow rates	LC/MS,	(Tiffany, 2003, Hu et al., 2006b, Böhm et al., 2012)
Saliva	Easy sample collection and processing, low cost,	Dependent on fluid flow rates	LC/MS	(Zhang et al., 2013)
Xenografts	Easy to use, relatively inexpensive, and reproducible	Not representative of the respective human tumour, cross-species	IHC, MRI, TMA, LC/MS, WB	(Pavlidis et al., 2010, Drake et al., 2011)
Breast Cyst Fluid (BCF)	Original disease source	Discomfort	IHC, ELISA,	(Mannello et al., 2006, Celis et al., 2006)
Nipple Aspirate Fluid (NAF)	Quick, Painless, and collected non-invasively.	Limitation of volume and expression.	LC/MS,	(Pavliou et al., 2010, Villa Flor Brunoro et al., 2014)
Nipple Discharge (ND)	Frequently discharge from breasts.	Limitation of expression.	LC/MS,	(Sauter et al., 2005)
Pathologic Nipple Discharge (PND)	Pathological spontaneous nipple discharge with a suspicious Lesion.	Limitation of expression.	LC/MS,	(Sauter et al., 2005)
Ductal Lavage (DL)	More cells as compare to normal aspiration	Discomfort and difficult to recannulate the same ducts.	QM-PCR, IHC,	(Fackler et al., 2006, Khan et al., 2009)
Random Periareolar Fine Needle Aspiration (RPFNA)	A snap-shot of whole breast, can be performed in majority of women, high number of cells.	High cost, discomfort, low patient turn over, reproducibility of the method	IHC, Reverse Phase Proteomic Microarray (RPPM)	(Ibarra-Drendall et al., 2012)
Milk and Colostrum	Easy sample collection	Sample preparation	LC/MS	(Hu et al., 2006b, O'Donnell et al., 2004)

Table 1-5: Biological samples used for biomarker proteome of breast cancer.

1.7.1 Cell culture

Breast cancer cell lines provide models to investigate the molecular mechanism underpinning the disease as well as identification of biomarkers. The quantitative and qualitative proteome of cell lines and conditioned medium (CM) is well established and provides a rich source of potential biomarkers. The advantages of cell lines are; (1) cost effective, (2) easily propagated, (3) versatility, and (4) non-invasive (Hanash et al., 2008). The cell culture-based approach also provides a platform to validate the potential efficacy of a novel biomarker. The presence of potential biomarkers in CM, either due to secretion or cell death, provides promise to extrapolate biomarker detection in body fluids. Cell lines are grown in serum-free medium in order to enrich the secretome, an approach which cannot be used for tissues (Paul et al., 2013). Different breast cell lines represent different stages and type of tumour helping to provide relevant information of disease. But immortalised cell lines lose the control to divide after a certain amount of time and also contain mutations or genetic alterations introduced by transfection. Primary cell can be better option but frequently require special growth conditions, risk of contamination and generate only a few passages (Paul et al., 2013). Designing experiments on different phenotypes of breast cancer cells can be helpful for better understanding of cell signalling, migration, proliferation and survival. Cell lines can also be used for creating 3D tumour models (spheroids) and xenografts in preclinical models to test drug candidates, prior to clinical trials (Wilding and Bodmer, 2014).

1.7.2 Tissue biopsies

Breast tissue biopsies take a number of forms, including core biopsies collected under local anaesthetic and larger samples collected during surgery/mastectomy. Breast tissue biopsies are heterogeneous cellular structures, with different proportions of epithelial, fibroblast, endothelial and myo-epithelial cells, accompanied by stroma, nerve fibres, macrophages, adipocytes, circulating cells, vasculature and lymphatics, representative of healthy and tumour environments (Hondermarck et al., 2008). Proteome profiling of tumour tissues can reveal the changes caused by disease progression that lead to breast cancer. Tissues offer the advantage of relating to the origin of the disease and the concentration of potential biomarkers may be high compared to surrounding microenvironment. There are more than 5,000 papers on tissue proteomics for identification of breast cancer specific biomarkers (NCBI, 2017) but the differentially expressed proteins identified varies considerably due different objectives, patient cohorts and experimental approaches, (Hanash et al., 2008). Tissue proteomic analysis from a diagnostic perspective, however is confounded by the presence of blood, is highly invasive and causes discomfort for the patient. Sample acquisition, also requires co-ordination with a clinician, a pathologist, snap-freezing to prevent tissue degradation and it can be very difficult to get matched healthy breast tissue [40]

1.7.3 Plasma/Serum:

There is a high probability for detection of a biomarker in biological fluids, but the link between expression of potential biomarkers at the tissue level and detectable levels in serum for disease specific profiles is not fully clear. The easy sample collection, the dynamic nature and constituents of serum, makes

it a logical choice for understanding the physiological or pathological states of a disease by biomarker applications (Hanash et al., 2008). Thousands of publications have reported a single protein or panels of targets as cancer biomarker(s) and have produced some encouraging data. For example, the level of prostate-specific antigen (PSA) in the blood to detect prostate and breast cancer or carcinoma embryonic antigen for therapy response in colon cancer have resulted from plasma proteomics (Hanash et al., 2008, Nikolenko et al., 2015, Chang et al., 2014, Steward et al., 1974). However, MS-based plasma/serum proteomics is extremely challenging for example; (i) the concentrations of individual proteins span 10–12 orders of magnitude and (ii) presence of few proteins in very high concentration such as albumin, represent more than 99% of the total bulk mass of protein content, and removal of these by, for example, by immunodepletion, requires additional steps with possible losses of less abundant proteins (Feist and Hummon, 2015).

1.7.4 Milk and Colostrum

There have been a number of proteomics studies on milk purely from a functional perspective, unrelated to cancer, with a recent study identifying up to 1600 proteins (Roncada et al., 2013, Beck et al., 2015). Colostrum, produced 2 or 3 days prior to lactation, prepares the infant's digestive system for milk as a food source and immunisation against infection, has also been characterised. In a study of 100 samples using 2D liquid chromatography mass spectrometry (LC MS), 151 proteins were identified after immunodepletion to remove the most abundant proteins, including 83 found in colostrum but not milk (Palmer et al., 2006). Although available in useful volumes for analysis, the period of production during the reproductive phase of life is relatively narrow and in most

cases will not overlap with breast cancer development. Consequently, there has been very little research on proteomic profiling of breast cancer in milk or colostrum because of the relatively small proportion of women in which the disease and postnatal breast-feeding, coincide. Nevertheless, Schneider et al were able to profile samples from a small cohort and identified proteins that were uniquely present in milk from women diagnosed with breast cancer (Schneider et al., 2014).

1.7.5 Breast Cyst Fluid

Several epidemiological and prospective studies indicate there may be a relationship between cystic breast disease and cancer (Celis et al., 2006). There are two types of breast cyst (Type I or apocrine cysts and Type II) differentiated on the basis of morphological and cellular characteristics. Apocrine cysts differ from Type II cysts in having a higher K^+/Na^+ ratios while Type I cysts are more strongly related with breast cancer (Mannello et al., 2006). According to Mannello et al, more than 100 studies have identified 81 proteins in breast cyst fluid, with the major components identified as albumin, prolactin inducible protein, zn- α_2 -glycoprotein, and apolipoprotein D (Mannello et al., 2006). A 2D gel electrophoresis/mass spectrometry study on apocrine macrocyst fluid collected identified that 15-hydroxy-prostaglandin dehydrogenase and 3-hydroxymethylglutaryl-CoA synthase were associated with cysts and tumour tissue but not in matched normal tissue (Celis et al., 2006).

1.7.6 Ductal Lavage (DL)

DL is a non-surgical breast epithelial sampling procedure that was developed to identify high risk breast cancer women and to detect malignant lesions in breast epithelial cells. In the DL procedure, a microcatheter is cannulated to the ducts, infused by a saline solution, and then aspirated by a suction device to collect cells from the lining of ducts. More cells are acquired through ductal lavage than from nipple aspiration (Dooley et al., 2001), and biopsy fluids can be obtained in women that do not yield fluid by passive nipple aspiration or discharge or from breast massage. However, ductal lavage can cause considerable discomfort and requires a specialist device which has prevented widespread clinical use (Mitchell et al., 2005). Ductal lavage containing sufficient cells was collected from 31 women (out of 36 volunteers) diagnosed with breast cancer and analysed for atypical cytology as a possible diagnostic indicator, however only 13% produced a significant positive indication (Khan et al., 2004). A separate study of 30 samples found only 23.3% of women with atypical lavage cytopathology but these women had normal mammogram screening of breast, indicating the potential for improved sensitivity (Hartman et al., 2004). The isolation of cells from ductal lavage opens up the opportunity for applying molecular biology approaches. Quantitative multiplex methylation-specific polymerase chain reaction (QM-PCR) was used to quantitate cumulative gene promoter hypermethylation in multiple genes, which are markers for breast cancer, and found to double the sensitivity of detection of cancer cells compared with cytology (Fackler et al., 2006). An attempt to identify biomarkers of tamoxifen treatment (estrogen receptor α , Ki-67 and cyclooxygenase-2) in ductal lavage, however, found no significant cytological or molecular biomarkers in patients (Khan et al., 2009). Recently, an improved

method of ductal aspiration, collecting multiple aliquots, considerably increased the cell recovery with 45/50 subjects yielding more than 1000 cells and 50% of those producing more than 20000 cells with 80-100% epithelial cell purity. This provided genomic DNA, RNA and miRNA samples for analysis however, to date only qualitative observations of the molecular profiles have been reported (Danforth et al., 2015).

1.7.7 Random Peri-areolar Fine Needle Aspiration (RPFNA)

RPFNA, developed by Dr. Carol Fabian in 1980, provides a snap-shot of the breast by sampling cells from the entire breast of asymptomatic women. The major advantage of RPFNA is that it can be performed in the majority of women and the cell yields vary from 72-85%, considerably higher than ductal lavage. (Zalles et al., 1995). After anesthetizing the breast with 1 % lidocaine, five needle aspirations are made on the lateral breast site and four from the middle skin of each. The aspirated fluid consists of epithelial, immune, stromal and adipose cells (Fabian et al., 2000).

A clinical trial of 480 women indicated that RPFNA increased cytological atypia associated with breast cancer in high-risk women (based on family history, a prior diagnosis and precancerous biopsy) (Khan et al., 2009). Of the cohort, 20 women developed breast cancer after 45 months (7 DCIS and 13 invasive), indicating the promise for very early diagnosis. RPFNA was used for a chemoprevention study of alpha-difluoromethylornithine (DFMO) in 119 high risk women, but found no change in cytology or other RPFNA-based molecular markers such as expression of proliferating cell nuclear antigen, p53 or epidermal growth factor receptor (Fabian et al., 2002). A proteomic microarray study found that up to 60 phosphoproteins can be verified in triplicate from 5,000

to 10,000 micro-dissected RPFNA epithelial cells, suggesting the potential to track signalling pathways in order to understand the molecular changes occurring in mammary carcinogenesis (Ibarra-Drendall et al., 2012). The heterogeneous nature of the cell populations being tested for specific molecular markers and considerable discomfort to obtain the samples, are key limitations of the RPFNA approach. Furthermore, the difficulty in reproducing the method may preclude a role in screening of high risk women that involves repeated harvesting of material.

Proteomic strategies have been developed for the analysis of complex samples such as serum and plasma, however due to the wide dynamic range of protein concentrations (10^{12} order of magnitude), still only the more abundant components are detected. Factors that may contribute to these incongruities are; (1) heterogeneity within tumours or patients, (2) obtaining adequate controls, (3) non-specific disease-associated changes in blood proteins (e.g. inflammation associated proteins) (Hanash et al., 2008).

1.7.8 Nipple Aspirate Fluid (NAF)

The breasts of adult non-lactating women secrete a small volume of fluid, called "nipple aspirate fluid" (NAF) into the breast ducts (Pettrakis, 1993). The fluid passes down the main ducts and ampullae through alveolar glands of the breast, from which it enters the lymphatic and blood circulation (Pettrakis, 1986). Under normal conditions, the breast fluid cannot escape from nipple because the nipple ducts are blocked by viscous and dried secretions or constriction bands of smooth muscle and keratinized epithelium (Pettrakis, 1986). To maintain stable physiology of the breast, an equilibrium exists between fluid secretion and re-absorption. Several factors are associated with

NAF expression; age, ethnicity, early menarche, history of lactation, high dietary fat consumption and dietary intake of lactose (Zhao et al., 2009a). The concentration of proteins can be higher in NAF compared to plasma, enriched for proteins originating from epithelial cell lining the duct (Djuric et al., 2005). The potential biomarkers detected in NAF are summarised in Table 1-6. There is also a direct relationship between ear wax and NAF because both are produced by ceruminous glands. Women with wet ear wax yield more NAF s compared to women with dry ear wax (Hawke, 2002). Premenopausal women with lactation experience, aged 30 to 50 years and had early onset of menarche produce more NAF compared to those who have not had children (Zhao et al., 2009a).

Biomarkers	Expression in NAF	Characteristics	Reference
Prostate specific antigen (PSA)	Down	Inversely proportional to disease stage, size of tumour, node status and distant metastases.	(Sauter et al., 1996, Sauter et al., 2004a)
Thomsen-Friedenreich (TF)	Up	Predictive for the presence of breast cancer or atypia.	(Kumar et al., 2005, Deutscher et al., 2010a)
Testosterone	UP	Only in postmenopausal women is predictive.	(Sauter et al., 1999, Sauter et al., 2002, Eliassen et al., 2006)
Superoxide Dismutases (SOD-1)	Down	Involved in cancer initiation and progression by ROS related damages	(Sinha et al., 2009, Mannello et al., 2010)
Protein DJ-1	Up	mRNA level increased but protein level decreased in tissue.	(Oda et al., 2012)
Cytokines/chemokines	Up	High level of pro-inflammatory C-C and CXC chemokines.	(Mannello et al., 2013)
Plasminogen activator inhibitor-1 (PAI-1)	Up	Promotes breast cancer invasion and metastasis	(Qin et al., 2012)
Transferrin protein (TFR) and ferritin (FTN)	Up	Proliferation of cancer cells	(Mannello et al., 2011)
C-reactive protein (CRP)	Up	Serum biomarker for metastasis of different type of cancers	(Lithgow et al., 2007)
Aluminium (Al)	Up	Varied concentrations between different tissue and fat of the breast.	(Mannello et al., 2009, Mannello et al., 2011)

Table 1-6: Potential biomarkers in nipple aspirate fluid.

A study of 25 to 49-year-old premenopausal nulliparous women found that proportionately, Asian-descendant women were less likely to express NAF compared to White American women [53]. NAF collection has been achieved with varying degrees of success dependent on the method and the practitioner and, in some cases, has deterred researchers from further investigation. Electronic and manual breast pumps (normally used for lactation), massage, warming and combinations of each have been used to acquire NAF samples (Sauter et al., 1997b). Most promising has been the use of oxytocin nasal spray which helps the release of already existing fluid in the ducts increasing collection in 95% of patients and volunteers (Zhang et al., 2003). On the base of unique characteristics of NAF, the proteomic analysis of NAF for potential breast cancer biomarkers may serve as useful approach to understand the physiology of breast cancer.

NAF is composed of a variety of endogenous substances such as lactose, proteins, fatty acids, hormones (estrogens, androgens, progesterone), sterols, but may also contain exogenous substances such as nicotine and cotinine from cigarette-smoking (Petrakis, 1993). The colour of NAF varies from clear to brown, bloody, black, pale yellow, dark yellow, white or green (Sartorius, 1973), and is associated with the concentration of cholesterol, estradiol, estrone, cholesterol epoxides and peroxidated lipids (Petrakis et al., 1988). The colour of NAF is more an epidemiological factor than indicator of the risk of breast cancer, however one study found that women have a higher risk breast cancer with bloody or brown nipple discharge compared to those which were white, cream, yellow or green (Dietz et al., 2002). Another study of 327 women found that the frequency of red or brown colour was increased with progression

of disease from pre-cancer to cancer and surgical biopsy had more influence on NAF colour compared to needle biopsy (Sauter et al., 2006).

NAF production, nutritional aspects and estrogen level have been found to be related with breast cancer risk. A large scale study of 1496 participants (1347 white and 153 black women) found a positive association between higher dietary fat and NAF secretion in the group aged 30-44 years (Lee et al., 1992). As obesity is associated with a high fat diet and is a major risk factor for breast cancer investigation, the correlation of fat intake and NAF expression and composition may be helpful for breast cancer prevention and prognosis (Huang et al., 2008). A link between lactose and soy intake has also been reported, however contrary results from a randomized crossover trial discovered no influence of soy on NAF volume and circulating estrogen level (Maskarinec et al., 2011). On the other hand, a fruit-and-vegetable diet was inversely related with NAF production while decreasing the circulating hormone concentration (Djuric et al., 2006) and the concentration of micro-nutrients, such as carotenoids and soy isoflavones in NAF, was related to dietary intake (Maskarinec et al., 2008).

Some components of NAF have been investigated as biomarkers of breast cancer (Table 1-6). Based on the differential levels of testosterone in serum from pre- and post-menopausal women, Sauter et al measured testosterone levels in NAF samples and found it would be a suitable biomarker to predict breast cancer risk (Sauter et al., 1999). A separate study, which measured the level of free and albumin-bound testosterone in NAF, found high levels of the former in premenopausal women with breast cancer (Sauter et al., 2002).

Proteins are major constituents of NAF with concentrations higher than plasma, typically averaging 71–170 mg/ml. NAF, however and most importantly, is enriched for proteins originating from epithelial cells lining the duct (Djuric et al., 2005), some of which have been evaluated as potential biomarkers of breast cancer. Prostate specific antigen (PSA), also known as kallikrein hK3, first identified in seminal plasma and prostatic tissue, produced by the epithelial cells lining the acini and ducts of prostate gland, has also been identified in female breast tumours. A study of NAF found that women with no risk factors or family history of breast cancer had high levels of PSA, but women with precancerous or invasive cancer had reduced levels (Sauter et al., 1996). Furthermore, PSA levels were inversely proportional not only to disease stage, but also tumour size, node status and distant metastases (Sauter et al., 2004a).

The concentration of superoxide dismutase [Cu-Zn] (SOD-1) in NAF was decreased in breast cancer patients compared to healthy individuals (Mannello et al., 2010). SOD-1 is involved in cancer initiation and progression caused by reactive oxygen species-related damage. Therefore, it was proposed that measuring the concentration of SOD-1, a key antioxidant enzyme in breast microenvironment, may be helpful to differentiate between the normal and tumour breast. The expression of the anti-oxidant oncogene DJ-1 mRNA is increased in ductal carcinoma tissues but the opposite effect was observed at the protein level, where expression is decreased and contrarily was elevated in blood of breast cancer patients. A study on NAF collected from 136 patients identified high levels of DJ-1 protein in NAF from breast cancer patients, but low levels in benign papilloma cases (Oda et al., 2012).

NAF samples collected from non-cancer and cancer women for cytokine profiling found no difference in anti-inflammatory cytokines (IL-4, IL-9, IL-10 and

IL-13), pro-inflammatory cytokines (IL-2 and interferon- γ), immuno-modulatory interleukins (IL-5, IL-7) or chemokines (RANTES, IP-10, eotaxin). However, NAFs from cancer patients with high levels of aluminium in the breast microenvironment, had higher concentrations of pro-inflammatory cytokines (IL-1 β , IL-6, IL-12 p70, and TNF- α), and C-C (MCP-1 and MIP-1 α) and CXC-type chemokines (IL-8) compared to those cancer patients with low aluminium levels. This indicated a significant correlation between pro-inflammatory cytokines (IL-6), monocyte/macrophage chemo-attractant chemokines (MIP-1 α and MCP-1), oxidative stress and aluminium content in cancerous NAFs (Mannello et al., 2013).

1.8 NAF Proteomics

On the basis of the unique characteristics of NAF, proteomic analysis should serve as a useful approach to understand the physiology of breast cancer and for biomarker discovery. However, early proteomic profiling of NAF samples collected from cancerous and non-cancerous breast of patients using surface-enhanced laser desorption ionization mass spectrometry (SELDI-MS), revealed no significant differences in the SELDI-MS peak profiles (Paweletz et al., 2001). Use of more powerful separation techniques, however, started to reveal differences. Varnum et al identified 64 proteins in immune-depleted NAF samples, using an ion trap mass spectrometer, among which 15 had previously been reported to be altered in tumour tissue and serum from women with breast cancer, including osteopontin and cathepsin D (Varnum et al., 2003). Two-dimensional PAGE separation of proteins, followed by in-gel digestion with trypsin and matrix-assisted laser desorption ionization time-of-flight mass spectrometer (MALDI-TOF) analysis, identified 41 components in NAF (Alexander et al., 2004). Among these, levels of prolactin-inducible protein,

apolipoprotein D, and α 1-acid glycoprotein, were observed to be changed in cancer NAF samples. Further validation by ELISA, indicated that expression of these proteins correlated with pre-/post-menopausal status and cancer stage. Pawlik et al (Pawlik et al., 2006) used Isotope-coded affinity tag (ICAT) tandem mass spectrometry (MS) for qualitative and quantitative analysis of tumour specific proteins in NAF, identified 353 peptides from 39 proteins in NAF samples from 12 women with breast cancer and 15 healthy volunteers. alpha-2-HS-glycoprotein, was found to be decreased, whereas lipophilin B, beta-globin, hemopexin and vitamin-D binding protein were increased in breast cancer NAF samples. A recent study on six NAF samples (3 healthy individuals and 3 patients) analysed by using an Orbitrap[®] mass spectrometer, identified more than 854 unique proteins, including established putative breast cancer biomarkers candidates, cancer antigen 15.3, tissue plasminogen activator, uPA, and cathepsin-D (Pavlou et al., 2010). Recently, in a series of experiments to optimise protein separation from a NAF sample, Brunoro et al identified 557 different proteins (Brunoro et al., 2015). The different protein profiles identified in NAF samples clearly highlights the potential for identifying biomarkers that could be related to breast cancer.

1.9 Aims and objectives

The overall aim of this study was to implement the latest advances in quantitative proteomics to characterise, as comprehensively as possible, protein biomarkers of breast cancer by using cell lines, tissue biopsies, serum and nipple aspirate. While several published researches have explained proteomic characterisations of breast cancers, none have successfully reached the full potential of quantitative proteomics in biological samples for early detection of breast cancer.

Breast cancer cell lines (BCCL) and tissue biopsies are well studied models for investigation of cancer related markers because they provide a good representation of human breast cancer. Breast cancer cell lines provide the opportunity to investigate the phenotypic signatures and full development course of breast cancer. BCCL models are available for different phenotypic variations of breast cancer like luminal A, luminal B, HER2, basal-like and claudin-low. These transfected cell lines are partially artificial but allow predictable tumour development under spontaneous conditions in the space of months rather than decades. To date, no published studies have used human mammary epithelial cells (HMEC), to quantitatively characterise breast cancer cell lines or their respective phenotypic proteomes using iTRAQ based MS proteomics.

In the clinical environment, the verification and validation of these signature proteins in breast cancer cell lines or tissue biopsies, is not an ideal approach for early detection of disease. Therefore, biofluids such as serum and nipple aspirate fluid (NAF) provide better options because of accessibility of

biopsies material. If the signatures are identifiable in the serum and NAF they can potentially reveal insights into how a whole organism progressively responds to development tumours. The contrast between serum and NAF, poses interesting questions as to the similarities and differences in proteome of both models.

New advancements in quantitative proteomics, both in sample preparation techniques and instrumentation, have the potential to provide quantitation for thousands of proteins simultaneously and non-biasedly. Such near comprehensive quantification of protein, rather than mRNA, expression has far greater clinical applicability to understanding the phenotype of breast cancer. Protein based biomarkers have great diagnostic values because of their functional roles in promoting the hallmarks of cancer. Isobaric tags (iTRAQ) quantification provides an analytical approach for relative quantification of up to 8 biological samples in a single LC-MS experiment, however, it still remains limited by the dynamic range compression because of the effects of precursor co-isolation. New Orbitrap Fusion mass spectrometer, with three mass analysers, provide an ideal proteomics workflow to deal with dynamic range compression of Isobaric tags quantitation and most effectively utilise quantitative MS data. Multiple reaction monitoring (MRM) assay allows the absolute quantification of protein-based biomarkers in complex biological samples. MRM assays are highly accurate for quantitation of hundreds of target proteins in a single LC-MS experiment and require very small amount of sample. MRM-MS assays are highly sensitive (targeted) because they are based on the tryptic peptides as stoichiometric representatives of the target proteins. The concentration of target peptides is quantified against the calibration of curves of

these tryptic peptides, to get absolute amounts of protein in biological sample. But selection of the right instrument and right tryptic peptides is very important for sensitivity and accuracy of multiple reaction monitoring mass spectrometry (MRM-MS). This investigation therefore comprises 3 key aims:

1. To use an iTRAQ quantitative proteomics strategy to determine and compare phenotype-specific signatures in a panel of breast cell lines. In so doing, I will also explore the dynamic range compression of iTRAQ based quantitation by comparing the two different MS platforms; Ultraflex II and Orbitrap Fusion.

2. Through label-free proteomics analysis to identify stage-related protein changes in nipple aspirate fluid samples from different stages of disease. From this, I will determine whether the proteomic composition of NAF can be used for biomarker discovery and breast health screening.

3. To develop a multiplex MRM-MS assay for target proteins from the phenotype- and stage-specific data for evaluation in breast tissue and liquid biopsies. This will provide a platform for an expanded validation study of selected biomarkers that can be utilised for highly specific, early detection of breast cancer.

CHAPTER 2. MATERIALS AND METHODS

2.1 Chemical and reagents

Wherever possible, the highest quality reagents were used for this project; sodium dodecyl sulphate (SDS), Tris (hydroxymethyl) aminomethane hydrochloride (Tris-HCl), Glycine, Sodium bicarbonate (NaHCO_3), Magnesium chloride (MgCl_2), Sodium chloride (NaCl), Sodium deoxycholate (DOC), 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate (CHAPS), Tetra methyl ethylenediamine (TEMED), urea, thiourea, glycerol, Bromophenol blue, ammonium persulfate, Triethylammonium bicarbonate (TEAB), Iodoacetamide, PhastGel® Blue R, Dulbecco's modified Eagle's medium (DMEM), Roswell Park Memorial Institute medium (RPMI), glutamine, pyruvate, penicillin, streptomycin, Foetal bovine serum (FBS), HPLC and LC-MS grade water, methanol, acetonitrile (ACN), acetone and formic acid (FA) were purchased from Sigma-Aldrich UK.

Tween 20 (Tween), 2-mercaptoethanol, sodium azide, acetic acid, dimethyl sulfoxide (DMSO) and dithiothreitol (DTT), 30% w/v acrylamide, phosphate buffered saline pH 7.4 (PBS), trifluoroacetic acid (TFA) were purchased from Fisher Scientific UK. Multi-grade Developer and Rapid Fixer (Ilford Photo, Cheshire, U.K). Proteomics grade trypsin and protease inhibitor cocktail were purchased from Roche Diagnostics GmbH, Germany. iTRAQ 8-plex isobaric labelling reagents were purchased from Sciex.

2.2 Cell culture

2.2.1 Breast cell lines

Breast cancer cell lines (MCF-7; ZR-75; MDA-MB-468; MDA-MB-231; MDA-MB-453) and Non-tumorigenic breast cells (MCF-10A) were obtained from American Type Culture Collection (Manassas, VA). Human mammary epithelial cells (HMEC) were isolated from biopsy material collected from a patient undergoing a double mastectomy for BRCA1 risk reduction at Bradford Royal Infirmary and determined by pathology to be essentially normal. Ethical approval was given by Leeds (East) Research Ethic Committee, reference 07/H1306/98+5. HB2 non-tumorigenic cells were kindly provided by Professor Valerie Speirs from The Leeds Institute of Cancer and Pathology (University of Leeds).

All cell lines were tested for mycoplasma contamination using MycoProbe mycoplasma detection kit (R&D Systems, Abingdon UK). Breast cancer cell lines (MCF-7, ZR-75; MDA-MB-468; MDA-MB-231; MDA-MB-453) were maintained in RPMI medium supplemented with 5% v/v L-glutamine, 5% v/v sodium pyruvate and 10% v/v fetal bovine serum (FBS). HB2 breast cell lines were grown on DMEM medium containing 10% FBS supplemented with hydrocortisone (5 µg/mL) and insulin (10 µg/mL). MCF-10A normal breast cells were cultured in Mammary Epithelial Cell Growth Medium (MEGM) Bullet Kit (Lonza Walkersville, USA) and HMEC cells were cultured in MEGM™ Mammary Epithelial Cell Growth Medium (Lonza Walkersville, USA), referred to as standard growth medium. All cells were cultured at 37°C and 5%CO₂ and were harvested. Cell cultures were maintained for 8-10 passages, with cells typically

being harvested for analysis between 5 and 10 passages, then collected at log phase of growth, with density of 5×10^6 cells/ml.

2.2.2 Cells washing for proteomics

Cells were harvested by trypsin (Sigma-Aldrich, Poole, U.K.)-treatment once reaching 80-90% confluence. Cells were washed at least 3 times with sterile ice cold PBS, centrifuging at 1000 rpm for 5 minutes each time the supernatant was decanted and discarded after centrifugation at 1000 rpm (revolutions per minute) for 5 min, and the resulting cell pellet (5×10^6 cells) was stored at -20°C until required. Washing in PBS was repeated to remove cell debris, predominant FBS proteins and contaminants that may be bound to the cells.

2.3 Breast tissues procurement

The study protocol and patient consent forms were approved by the Cyprus National Bioethics Committee. Patients underwent surgery for removal of breast lesions and subsequent histopathological diagnosis. Following inspection by a histopathologist, resected specimens were snap frozen in isopentane cooled by liquid nitrogen and stored at -80°C . Tissues from the breast lesion and areas identified as normal, at least 5 cm apart, were obtained. Biopsies from patients with fibroadenoma, DCIS or invasive carcinoma were selected for proteomic analysis. Frozen sections were cut from matched blocks of normal and disease breast tissue using a Bright cryostat as described previously (Sutton et al., 2010).

2.4 Collection of serum samples

Blood samples were collected as part of population-based case–control study of breast cancer in Cyprus after approval from the Cyprus National Bioethics Committee. Controls were women with no prior history of breast cancer and were active members of national mammography population screening programme. Each healthy volunteer and case gave written consent before sample collection. Blood samples were collected in no additive-vacutainer tubes and kept on ice during transported to the laboratory. Then samples were centrifuged at 2000 rpm for 15 minutes at 4°C, to collect serum and stored at -80°C until further use.

2.5 NAF collection

NAF samples were obtained from healthy volunteers and breast cancer patients, who presented to Bradford Teaching Hospitals NHS Trust, between 2013 and 2016. All participants gave written informed consent to undergo bilateral nipple aspiration. The study protocol was approved by University of Bradford's Independent Scientific Advisory Committee (reference: application/13/051). Ethical approval was given by Leeds (East) Research Ethics Committee, reference 07/H1306/98+5. Before aspiration was attempted, the nipple was initially cleansed with an alcohol pad to remove any keratin plugs. NAF collection from cancer patients was performed under general anaesthetic by the clinical team, prior to surgery for lumpectomy or mastectomy, assisted by massaging the breast and the liquid collected from the nipple using a sterile pipette. After collection, the samples were transferred to chilled, pre-labelled tubes containing a freeze-dried protease inhibitor cocktail mixture [Roche

Diagnostics, Germany], and frozen within 30 minutes of collection, at -20°C in dedicated fridge-freezers. Where possible, NAF samples were collected separately from both breasts to provide matched normal and disease samples. NAF from healthy volunteers was collected in a similar manner by the individual whilst conscious. A bank of more than 100 patient samples has been collected to date.

2.6 Protein extraction

2.6.1 Cell lines

Each cell pellets (5×10^6 cells) were thawed, and 200 µl of urea extraction buffer (7 M urea, 2 M thiourea, 0.4% w/v CHAPS, 50 mM DTT in PBS pH7.4 containing protease inhibitor cocktail was added, and incubated on ice for thirty minutes and sonicated for 20 seconds on ice using a Status US70 sonicating probe (Philips Harris Scientific, UK). The samples were centrifuged at 13,400 rpm for 20 minutes, 4°C and the liquid phase extracted to new tubes. The sonication and centrifugation steps were repeated if the sample remained cloudy. Lysates were stored at -20°C short term or -80°C long term.

2.6.2 Tissue Biopsies

Cryo-sections for matched normal and diseased tissues for four patients were prepared simultaneously using a dual lysis buffer method. For each sample, RIPA lysis buffer (50 µL, PBS pH 7.4, 0.1% w/v SDS, 0.25% w/v sodium deoxycholate containing EDTA-free protease inhibitor cocktail (Roche diagnostics GmbH, Germany) was added, subjected to vortexing for 30 minutes

at room temperature and sonicated for 20 seconds on ice using a Status US70 sonicating probe (Philips Harris Scientific, UK). The samples were centrifuged at 13,400 rpm for 20 minutes, 4°C and the liquid phase extracted to new tubes. Urea lysis buffer (50 μ L, 7M urea/2M thiourea/4% w/v CHAPS/50mM DTT in PBS containing EDTA-free protease inhibitor cocktail) was added to the pellet twice, each time treated with vortexing, sonication and centrifugation, and the resulting supernatant combined with the RIPA buffer protein extract.

2.7 Protein quantification

The protein concentration of each sample (NAF, serum and lysate from cell lines and tissues), was measured using the Bradford assay (Bradford, 1976) protein quantification kit (Bio-Rad Laboratories, Hemel Hempstead, U.K.), according to the manufacturer's instructions. The kit was used to calculate the protein concentration of unknown samples relative to a standard curve of bovine serum albumin (BSA – 0 to 2000 μ g/ml). All samples were allowed to thaw on ice, before calculating the protein concentration, followed by vortexing for 1 minute at room temperature. Samples were typically diluted 1-in-20 with HPLC grade water, and where appropriate, cell lysis buffer or 8M urea, was also used to determine background absorbance in the assay due to non-protein interferences. Bradford reagent (1.5 ml) was mixed with 50 μ L of samples and standards, and incubated for 10 minutes at room temperature. After incubation, 1ml of reaction mixture was transferred to plastic cuvettes to measure the absorbance at 595nm by using a Multiskan Spectrum plate reader (Thermo Scientific), operated with SkanIt software. The equation of optical density relative to standard BSA concentration was calculated, and applied to the

unknown samples to determine the protein concentration. All measurements were performed in triplicate wherever possible.

2.8 SDS-PAGE and Western blotting

2.8.1 SDS-PAGE

Samples, equivalent to 20 µg of protein, were mixed with 10 µL of SDS reducing buffer (Laemmli buffer; 63mM Tris-HCl pH 6.8, 10% glycerol, 2% SDS, 0.0005% bromophenol blue and 0.1% β-mercaptoethanol) heated to 70°C for 15 minutes. Denatured lysates were diluted with HPLC grade water to derive a consistent volume before loading to SDS polyacrylamide gels, with a 4% (v/v) stacking gel and 12% (v/v) separating gel in electrophoresis buffer (25mM Tris, 193mM glycine, 0.1% SDS). Denatured lysates and 10 µL of Precision Plus Protein Standard (Fisher Scientific UK) 10-250 kDa were resolved with a Mini-PROTEAN 3 Cell system (Bio-Rad) for 10 min at 80 V and then 1 hr for 150 V.

2.8.2 Coomassie blue staining

After electrophoresis, the apparatus was disassembled and the gel transferred to a container and covered with 10 ml of the Coomassie blue reagent, for 1 hour at room temperature with gentle agitation. Then, the staining reagent was discarded and the gel was covered by 10 ml of destain solution (50% v/v methanol, 40% v/v water and 10% v/v acetic acid). The de-staining continued with gentle agitation until the protein bands became clear. Finally, the gels were scanned by Canon CanoScan 9000F MKII A4 Flatbed Colour

Scanner, operated through MP Navigator EX Ver. 5.0.2 software (Canon Tokyo, Japan).

2.8.3 Western blotting

Proteins were transferred by 'wet transfer' to nitrocellulose membrane (GE Healthcare) at 300 volts for 60 minutes at 4°C in transfer buffer (SDS-PAGE running buffer with 20% v/v methanol). The nitrocellulose membrane was first pre-activated for 5 minutes in HPLC grade water, then a protein transfer sandwich was made as follows: Positive electrode, sponge, filter paper, membrane, gel, filter paper, sponge, negative electrode (Figure 2-1).

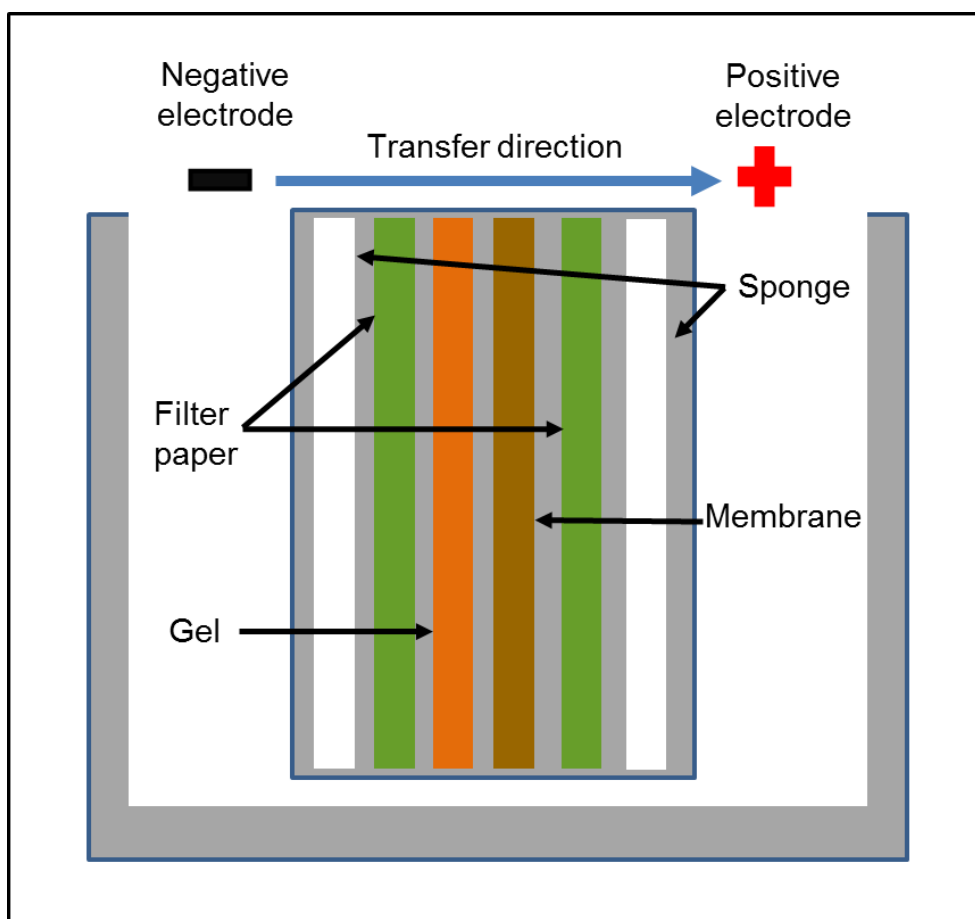


Figure 2-1: Representation of transfer "sandwich" assembly for Western blotting.

Coomassie blue staining was performed on residual gels to demonstrate successful protein transference. After transfer, membranes were washed in distilled water for 5 minutes.

Nitrocellulose membranes were then blocked with 5% (w/v) non-fat milk in TBS-Tween-20 (0.05%) for 1 hour to prevent further, unwanted protein association. All immunoblotting was performed with primary antibodies diluted in blocking buffer (5% (w/v) non-fat milk in TBS Tween-20. Table 2-1 identifies the specific dilutions, manufacturer and the animal in which the antibody was produced. After incubation with primary antibodies, membranes were washed 3 times for 5 minutes with TBS Tween-20.

The appropriate secondary antibodies (Table 2-1), in blocking buffer, were incubated with the membrane for 45 min at room temperature. The residual secondary antibodies were removed with three TBS-Tween washes. All secondary antibodies were conjugated to an enzyme "horseradish peroxidase" (HRP) and proteins detected by the addition of ECL Plus Western Blotting reagent, then exposed to X-ray film (GE Healthcare, Amersham, U.K.). Relative quantification of band intensity was calculated using GelAnalyzer 2010a software (<http://www.gelanalyzer.com>).

Target Gene	Description	Antibody	Clonality	Amount (µg)	Catalogue no.	Supplier
ACTB	Actin, cytoplasmic 1	Anti-ACTB	monoclonal	0.05	A2228	Sigma, Dorset, U.K
ALDH2	Aldehyde dehydrogenase, mitochondrial	Anti-ALDH2	polyclonal	0.40	WH0000217M1	Sigma, Dorset, U.K
CALR	Calreticulin	Anti-CALR	monoclonal	0.10	WH0000811M1	Sigma, Dorset, U.K
CD44	CD44 antigen	Anti-CD44	polyclonal	0.05	ab157107	Abcam, Cambridge, U.K
CFL1	Cofilin-1	Anti-CFL1	polyclonal	0.10	ab54532	Abcam, Cambridge, U.K
CYP3A4	Cytochrome P450 3A4	Anti-CYP3A4	polyclonal	0.20	ab3572	Abcam, Cambridge, U.K
EGFR	Epidermal growth factor receptor	Anti-EGFR	polyclonal	0.10	ab131498	Abcam, Cambridge, U.K
HSPE1	10 kDa heat shock protein, mitochondrial	Anti-HSPE1	monoclonal	0.03	ab108600	Abcam, Cambridge, U.K
PIP	Prolactin-inducible protein	Anti-PIP	polyclonal	0.03	ab15453	Abcam, Cambridge, U.K
PTGES3	Prostaglandin E synthase 3	Anti-PTGES3	monoclonal	0.10	ab2814	Abcam, Cambridge, U.K
VIM	Vimentin	Anti-VIM	polyclonal	0.20	SAB4503081	Sigma, Dorset, U.K
N/A	Secondary antibody-conjugated with horseradish peroxidase	Anti-Rabbit IgG-HRP	Polyclonal	0.03	P0448	Dako, Glostrup, Denmark
N/A	Secondary antibody-conjugated with horseradish peroxidase	Anti-Mouse IgG-HRP	Polyclonal	0.03	ab6789	Abcam, Cambridge, U.K

Table 2-1: Antibodies used for Western blotting.

2.9 Peptide preparation for LC-MS

2.9.1 Trypsin digestion

Each protein extract from cell lines and tissue biopsies (200 µg of protein) was precipitated overnight with 100% acetone at -20°C and centrifuged for 20 min at 13 400 rpm at 4°C . The pellet was resuspended in 8M urea in 400mM ammonium bicarbonate (AMBIC) and protein concentration was again measured by Bradford assay (above section 2.7 Protein quantification) to check the efficiency of acetone precipitation.

Sample amounts for proteomics were selected based on sample type and strategy; cell lines (80µg for iTRAQ and 50µg for MRM-MS), NAF (200µg for label-free quantification and 50µg for MRM-MS), tissue biopsies (50µg for MRM-MS) and serum (50µg for MRM-MS) Cell lines and tissue extracts were already in 8M urea, then reduced with 50 mM dithiothreitol (DTT) at 60°C for 15 mins and alkylated with 100mM iodoacetamide (IAA) at ambient temperature, in the dark, for 15 min. MS-grade trypsin (Fisher Scientific UK) was used to digest proteins at a protease-to-protein ratio of 1:10 (w/w) at 37°C for 20 hrs. After digestion, each sample was desalted on an Isolute C₁₈ desalting column and lyophilized.

2.9.2 Checking the efficiency of tryptic digestion

To check the efficiency of in-solution tryptic digestion, a 1µL aliquot of each digest was diluted 10-fold in 9µL 10% v/v acetonitrile (ACN). Then 0.5µL of each diluted sample was spotted between 2x 0.5µL saturated solution of α -cyano-4-hydroxy-cinnamic acid (CHCA) in 30% v/v acetonitrile (ACN), on a

MALDI Target Plate (MTP Anchorchip 800/384 massive target T, Bruker Daltronics) for manual analysis by Ultraflex II. Peptide calibration mix II (Bruker Daltronics) was also spotted on to the target plate to calibrate the instrument. The dried droplets of calibrant and digested samples, were analysed on Ultraflex II, MS1 spectrum were acquired through FlexControl v3.4 and visualised in FlexAnalysis v3.4 (please see section 2.10 for further details of Ultraflex II). A complex MS1 spectrum indicated good efficiency of tryptic digestion.

2.9.3 Isobaric tag peptide labelling

All lyophilised peptides were resuspended in 1M TEAB (Sigma-Aldrich, Poole, U.K.), 0.1% SDS, and were incubated with iTRAQ 8-plex (Sciex, U.K.), according to the manufacturer's instructions. The labelled peptides were then combined together, desalted on an Isolute C₁₈ desalting column and the eluate was lyophilized (48°C, aqueous mode) until fully dry.

2.9.4 OffGel fractionation

The total iTRAQ-labelled sample was resuspended in OffGel peptide sample buffer (containing pH 3–10 ampholytes) and applied to an OffGel 3100 (Agilent Technologies, Wokingham, U.K.) isoelectric focussing system using a pH 3–10 high-resolution strip, for 50 kV hours. Twenty-four fractions were collected, desalted on Isolute C18 RP cartridges, and then lyophilized until full dry and stored at -20°C.

2.9.5 SCX fractionation

The combined sample was resuspended in 600µL SCX loading buffer (10mM KH₂PO₄ in 25% ACN, 0.01% w/v sodium azide, adjusted to pH3), added to an Isolute SCX column (Kinesis Ltd, UK) pre-wetted with HPLC grade water and allowed to flow through under passive hydrostatic pressure. Peptides were then eluted stepwise in 12 fractions with successive 500µL volumes of elution buffer, using a potassium chloride from 0 mM to 1000mM (Table 2-2). Sample fractions were diluted with 1.5ml Solvent A (2% v/v ACN, 0.05% v/v FA), desalted on an Isolute C18 desalting column and lyophilised (48°C, aqueous mode) until fully dry and stored at -20°C.

Elution Fraction Number	Potassium Chloride Concentration
E01	0 mM
E02	30 mM
E03	60 mM
E04	90 mM
E05	120 mM
E06	150 mM
E07	180 mM
E08	250 mM
E09	300 mM
E10	350 mM
E11	500 mM
E12	1000 mM

Table 2-2: Strong Cation Exchange Elution Buffer (KCl) Concentrations.

2.10 LC-MS analysis

2.10.1 Ultraflex II

Each lyophilized sample was resuspended in 50 μL of 10% acetonitrile and 0.05% TFA (mobile phase A), 5 μL injected on to an LC Packings UltiMate 3000 capillary HPLC system (ThermoFisher, Bremen, Germany), washed on a C₁₈, 300 μm \times 5 mm, 5 μm diameter, 100 Å PepMap pre-column (LC Packings, Sunnyvale, CA) before transfer to a C₁₈, 75 μm \times 15 cm, 3 μm diameter, 100 Å PepMap column (LC Packings). Peptides were eluted with a linear gradient of 10-90% mobile phase B (80% acetonitrile, 0.05% TFA) over 95 min run time (Table 2-3). A total of 384, 75 nL fractions were co-deposited with 0.5 μL of a saturated α -cyano-4-hydroxy cinnamic acid (CHCA) matrix (Bruker Daltonik, Bremen, Germany) solution onto a MTP AnchorChip 800/384 target plate (Bruker Daltonik) using a Proteineer FC fraction collector (Bruker Daltonik) and allowed to air-dry. Peptide Calibration Standard II (Angiotensin I, Angiotensin II, Substance P, Bombesin, ACTH clip 1–17, ACTH clip 18–39, Somatostatin 28, Bradykinin fragment 1–7 and Renin Substrate Tetradecapeptide porcine; covering the mass range 700–3200 Da, Bruker Daltonik) was applied between each group of four fractions. Mass spectrometric analysis was carried out using a MALDI–TOF/TOF UltraFlex II instrument (Bruker Daltonik) with a 200Hz Smartbeam laser (>250 μJ /pulse) in reflector mode. A fully automated workflow was performed using WarpLC software (version 1.3), which encompassed, (i) data acquisition (FlexControl v3.4), data-processing (FlexAnalysis v3.4 - TopHat baseline subtraction, Savitzky-Golay smoothing and SNAP peak detection algorithms), (ii) compilation of a non-redundant list of peptides from the 384 HPLC fractions, (iii) data-dependent MS/MS of each peptide using LIFT

mode, and (iv) compilation of the MS/MS fragment mass lists into a batch (WarpLC v1.3). Duplicate LC–MALDI analyses were performed for each OffGel fraction.

2.10.2 Orbitrap Fusion

Lyophilized peptide fractions were individually reconstituted in 10-30 μ l of loading mobile phase (2% ACN, 0.1% FA) and 2-3 μ l loaded by a Dionex Ultimate 3000 (Thermo Scientific) at 25 μ l/minute for 4 minutes onto a C₁₈ PepMap100 trapping cartridge (5 mm \times 300 μ m ID, 5 μ m particle) (Thermo Scientific) in loading mobile phase. After peptide loading, the trapping cartridge was brought in line with an Acclaim PepMap 100 column (25 cm or 50 cm \times 75 μ m ID, 2 μ m particle) at a flow rate of 300 nl/minute with 5% mobile phase B (80% CAN or 100% ACN, 0.1% FA) in mobile phase A (2% ACN, 0.1% FA). Several reverse phase elution gradient lengths were used, proportionally extrapolated from the 120-minute gradient from 5-90% solvent B (Table 2-3).

Peptide elution was directly coupled to electrospray ionisation (ESI) at 2.0 to 2.4 kV using a steel emitter (Thermo Scientific), and characterised with an Orbitrap Fusion mass spectrometer (Thermo Scientific). MS analysis of eluting peptides was conducted through Xcalibur 4.0 with Foundation 3.1 SP1(Thermo Scientific) on Orbitrap Fusion, between 350 and 1500 m/z at 120,000 mass resolution, with the maximum injection time was 100 ms. All MS/MS acquisition was performed on the Ion-trap, in top speed mode with 3s cycle time, a dynamic exclusion (\pm 5 ppm) of 50-60 seconds, intensity threshold 5000, with charge states 2+ to 7+ were sequentially fragmented by collision-induced dissociation (CID) with a normalized collision energy (NCE) of 35%. A

maximum of 200 ms ion injection time was allowed. Additionally, the Polysiloxane (C₂H₆SiO) at 445.12003 was used as a MS lock-mass.

Automated Synchronous Precursor Selection (SPS) for MS³ setting was used for quantification of iTRAQ, a special feature of Orbitrap fusion which improves quantitative accuracy when using isobaric mass tags. All MS³ was performed on Orbitrap at 30000 resolutions, scan range 100-500 m/z, maximum injection time 105ms and with 65% high-energy collisional dissociation (HCD).

2.10.3 MRM-MS

All experiments were performed on a standard-flow LC-MRM/MS platform with Xevo TQD, Quattro Premier XE and Quattro Ultima triple quadrupole mass spectrometers (Waters, Manchester, UK). Xevo TQD was coupled to a Waters Acquity I-Class UPLC separation system, Quattro Premier XE with Waters Acquity UPLC Separation system and Quattro Ultima with on-line Waters Alliance 2695 HPLC Separation system. Instrument control was performed using MassLynx data system (details in Chapter 5). A two-stage UPLC gradient using solvent A and solvent B (Table 2-3) on Luna C₁₈ column (2 mm ID x 25 cm length, 5µm particle size, Phenomenex Inc., Macclesfield, UK), was used for the separation of the target peptides. A linear gradient of 3 to 90 % solvent B was applied for 40 minutes, followed by a column wash for 10 minutes using solvent B and then a 10-min equilibration of column, total run time was 60 minutes (Table 2-3). The UPLC retention time, cone voltage and collision energy were optimized using synthetic reference peptides for target proteins. Calibration curves were prepared with the synthetic peptides (Chapter 5) before quantitative analysis of protein extracts from cell lines and NAF samples.

Instrument	Experiment	Trap-Column	Analytical-Column	Solvent A	Solvent B	Total run time (Min)	Gradient time (Min)	Gradient % Solvent B	Column cleaning	Equilibration of column
Ultraflex II	iTRAQ-8plex (Chapter 3)	C18, 300 µm × 5 mm, 5 µm diameter, PepMap	C18, 75 µm × 25 cm, 2 µm diameter, PepMap	2% CAN, 0.05% TFA	80% ACN, 0.05% TFA	120	0	10	10 minutes with 90% solvent B	10 minutes with 10% solvent B
							80	35		
							90	45		
							95	90		
Orbitrap Fusion	iTRAQ-8plex (Chapter 3)	C18, 300 µm × 5 mm, 5 µm diameter, PepMap	C18, 75 µm × 25 cm, 2 µm diameter, PepMap	2% ACN, 0.1% FA	80% ACN, 0.1% FA	120	0	5	15 minutes with 90% solvent B	15 minutes with 5% solvent B
							5	10		
							65	25		
							85	45		
	1D-LC-MS, label-free (Chapter 4)	C18, 300 µm × 5 mm, 5 µm diameter, PepMap	C18, 75 µm × 50 cm, 2 µm diameter, PepMap	2% ACN, 0.1% FA	100% ACN, 0.1% FA	220	0	5	20 minutes with 85% solvent B	15 minutes with 5% solvent B
							5	7		
							165	25		
							180	45		
	2D-SCX-LC-MS, label-free (Chapter 4)	C18, 300 µm × 5 mm, 5 µm diameter, PepMap	C18, 75 µm × 50 cm, 2 µm diameter, PepMap	2% ACN, 0.1% FA	100% ACN, 0.1% FA	120	0	5	20 minutes with 85% solvent B	15 minutes with 5% solvent B
							5	7		
							65	25		
							80	45		
Xevo TQD	MRM-MS (Chapter 5)	N/A	C18, 2mm × 25 cm, 5 µm diameter, Luna	2% ACN, 0.1% FA	100% ACN, 0.1% FA	60	0	3	10 minutes with 85% solvent B	10 minutes with 3% solvent B
							30	25		
							38	40		
							40	85		
Quattro Premier XE	MRM-MS (Chapter 5)	N/A	C18, 2mm × 25 cm, 5 µm diameter, Luna	10% ACN, 0.1% FA	90% ACN, 0.1% FA	60	0	10	10 minutes with 90% solvent B	10 minutes with 10% solvent B
							30	25		
							38	40		
							40	90		
Quattro Ultima	MRM-MS (Chapter 5)	N/A	C18, 2mm × 25 cm, 5 µm diameter, Luna	10% ACN, 0.1% FA	80% ACN, 0.1% FA	60	0	10	10 minutes with 90% solvent B	10 minutes with 10% solvent B
							30	25		
							38	40		
							40	90		

PepMap; product of Thermo Scientific, **Luna**; product of Phenomenex, **N/A**; not applicable, **ACN**; Acetonitrile **TFA**; Trifluoroacetic acid **FA**; Formic acid,

Table 2-3: List of instruments and columns optimised gradients for each experiment.

2.11 MS data processing

MS/MS fragment mass lists were searched, via ProteinScape v3.0 for Ultraflex II (Bruker Daltonik) and Proteome Discoverer 2.1 for Orbitrap Fusion (Thermo Scientific), using Mascot software version 2.4 (Matrix Science, U.K.) against Swiss-Prot version 2016 containing 552,259 human protein sequences. A decoy search (based on automatically generated random sequences of the same length) was employed to determine the rate of false-positive identifications (Figure 2-2). Non-redundant protein profiles for each experiment were created in ProteinScape or Proteome Discoverer by combining the corresponding LC-MS datasets. The list of protein identifications was assessed manually, and all proteins defined as Master Protein Candidates, that could not be differentiated from the Master Protein (i.e. no unique peptides), were omitted for post processing of MS data. All stages involved in MS data processing on Proteome Discoverer and ProteinScape, from raw MS data file to final proteins list, are summarised in Figure 2-2.

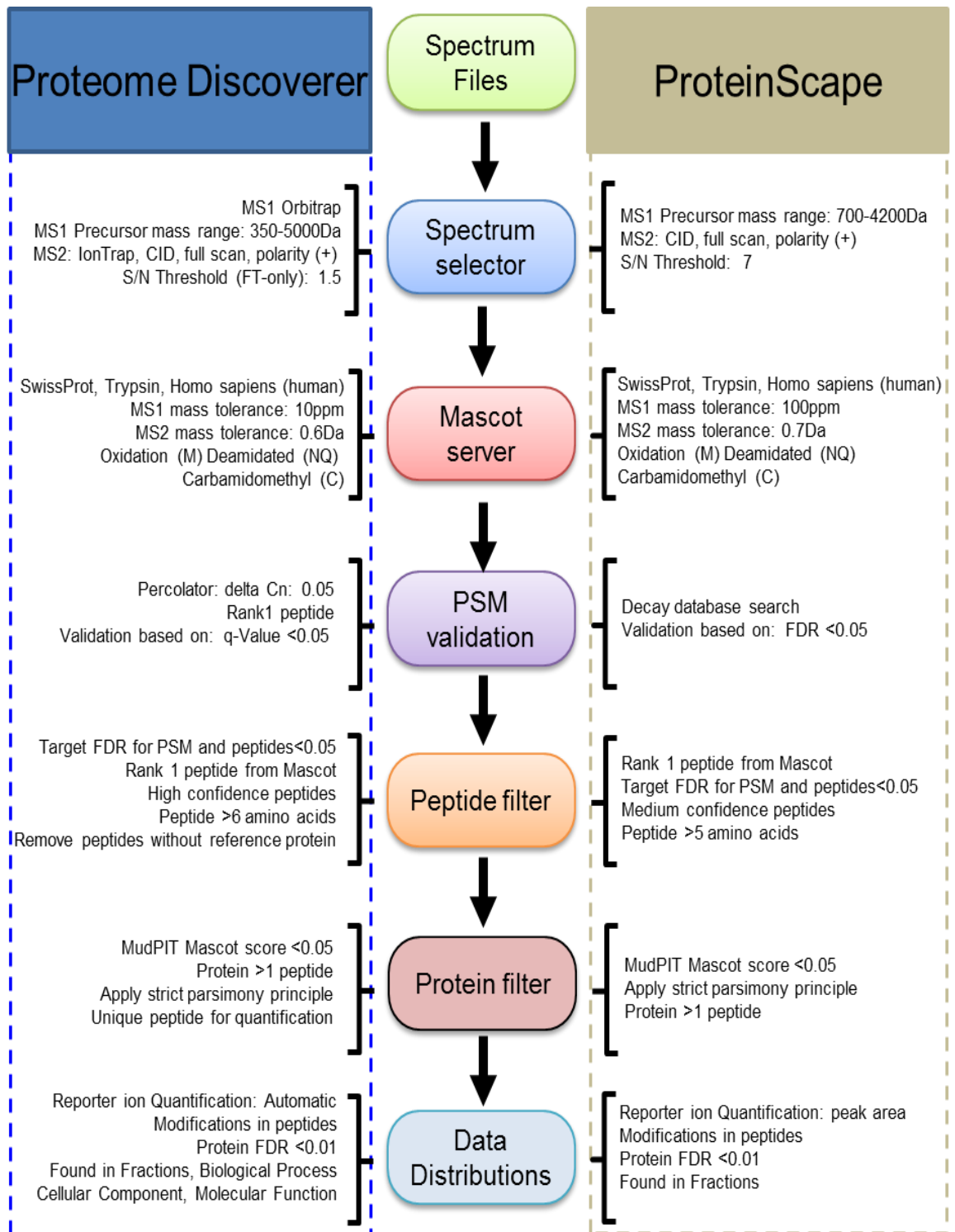


Figure 2-2: Overview of the different stages of Proteome discover and ProteinScape for LC-MS data processing.

2.12 Post-processing of MS data

Peptide spectrum match data was exported from ProteinScape and Proteome Discoverer with the associated features: number of peptides, unique peptides, number of PSM, coverage, score Mascot, posterior error probability (PEP), charge states, precursor intensities and the reporter ion intensity values.

2.12.1 Statistical analysis

All the statistical analyses were undertaken using R Studio version 0.99.903 (R Studio, Boston, MA) or PRISM version 6.0 software (GraphPad Software, San Diego, CA).

2.12.2 Hierarchical clustering

R studio was used for hierarchical clustering of quantified proteome, using a similarity metric of Euclidian distance and complete linkage. The R script is given in Figure 2-3;

```
1. library(Biobase)
2. library(gplots)
3. getwd()
4. setwd("C:/Users/Sadr/Documents")
5. getOption("max.print")
6. options(max.print = 99999999)
#calling data file
#####
1. mydata <- read.table(file = "Proteome.txt", sep = "\t", header = T,row.names=1,
  as.is=TRUE,check.names = F)
2. mydata
#How to summarise data

1. head(mydata)
# heatmap.2 works only with matrix, convert the dataframe to matrix
1. m<-as.matrix(mydata[,1:7])
2. rownames(m)<- mydata$genes # add the gene names as the row lable
3. png(filename = "myData.png", width=600, height = 800) #save the heatmap to a png or a
  pdf by pdf(filename=...)
4. bk = unique(c(seq(-5.73,-1.4, length=100),seq(-1.4,1.4, length=100),
  seq(1.4,5,length=100)))
5. hmcpls<- colorRampPalette(c("red","black","green"))(length(bk)-1)
6. heatmap.2(m,breaks=bk,col=hmcpls,trace="none",main = "BCL-Proteome",
  density.info="none",cexCol=1.0,labRow=NA,symm=F,symkey=F,symbreaks=T,
  scale="none",keysize=1.2)
7. dev.off()
#####
```

Figure 2-3: R script for Hierarchical clustering.

2.12.3 Limma t-statistics

Significantly up and down regulated proteins were defined by using Limma t- statistics in R- studio. The script for t-statistics is given in Figure 2-4.

```
1. library(Biobase)
2. library(limma)
3. getOption("max.print")
4. options(max.print = 99999999)
5. #calling data file
6. mydata
7. mydata <- read.table(file = "Proteome.txt", sep = "\t", header = T,row.names=1,
  as.is=TRUE,check.names = F)
8. #How to summarise data
9. head(mydata)
10. sml <- c("G0","G1","G1","G1","G1","G1","G1")
11. table(sml)
12. # set up the data and proceed with analysis
13. fl <- as.factor(sml)
14. fl
15. design<-model.matrix(~0+fl)
16. design
17. colnames(design)=levels(fl)
18. design
19. #boxplot(mydata)
20. #differential expression
21. fit <- lmFit(mydata, design)
22. cont.matrix <- makeContrasts(G1-G0, levels=design)
23. cont.matrix
24. fit2 <- contrasts.fit(fit, cont.matrix)
25. fit2 <- eBayes(fit2, 0.01)
26. # dim means dimension of data
27. dim(fit2)
28. tT <- topTable(fit2, adjust="fdr", sort.by="p", number=nrow(fit2))
29. head(tT)
30. #how to select all those proteins which have p < 0.05
31. #tT <- tT[tT$P.Value <0.05,]
32. dim(tT)
33. write.table(tT, file="Proteome.txt", sep="\t")
34. #####
```

Figure 2-4: R script for Limma t-statistics.

2.12.4 Gene Ontology (GO) enrichment analysis

All differentially expressed proteins were classified broadly into several groups according to the Gene Ontology (GO) annotation (www.geneontology.org). The over-representation analyses of GO terms, including the biological process, molecular function and enriched pathway analysis was performed using FunRich; Functional Enrichment Analysis Tool (www.funrich.org). Protein-protein interactions and network associations for significantly changed components were performed using STRING (version 9.05, www.string-db.org). Proteins identified in NAF samples were subjected to analysis in TMHMM Server v. 2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>, (Krogh et al., 2001)), to determine those with transmembrane helical regions.

2.13 Comparison of Western Blotting, iTRAQ and MRM-MS quantitation methods

2.13.1 Processing of immune-blot

Cell line extracts, equivalent to 20 µg of protein, were applied to SDS polyacrylamide gels. The blotted membrane was incubated in blocking buffer and then the appropriate primary antibodies; ACTB, CFL1, PTGES3 and HSPE1, in blocking buffer was incubated overnight at 4 °C. After removal of the primary antibody with three TBS-Tween buffer washes, the appropriate secondary antibodies and proteins detected by the addition of ECL Plus Western Blotting reagent, then exposed to X-ray film. For comparison, all immune blots were analysed by GelAnalyzer to calculate the intensity of bands then normalised with ACTB. The fold-change expression of proteins, was

calculated against HMEC then converted into log₂ value to simplify the interpretation of the dynamic range of expression.

2.13.2 Processing of iTRAQ labelled peptides

From iTRAQ Orbitrap Fusion data (Chapter 3), 4 proteins with 8 peptides; Cofilin-1 (CFL-1-1, CFL-1-2, CFL-1-3), 10 kDa heat shock protein (HSPE1-1, HSPE1-2, HSPE1-3), Prostaglandin E synthase 3 (PTGES3-1) and Actin Cytoplasmic (ACTB-1); were selected because these peptides were common in both quantitation assay (iTRAQ and MRM). All peptides (high confident data $p < 0.05$) were exported with non-normalised iTRAQ ratios then normalised with peptide ACTB-1 (GYSFTTTAER-iTRAQ labelled) and converted into log₂ values to compare with MRM-MS and Western blotting.

2.13.3 Processing of MRM peptides

The amount of all eight peptides (CFL-1-1, CFL-1-2, CFL-1-3, HSPE1-1, HSPE1-2, HSPE1-3, PTGES3-1), was normalised by using peptide ACTB-1 (GYSFTTTAER). Only one peptide of PTGES3 was selected because of the limited availability of unique peptides. The fold-change expression of peptides, was calculated against HMEC then converted into log₂ value.

2.13.4 Calculating the amount of protein

The average of three peptides (MRM-MS and iTRAQ data) was used to calculate the expression of CFL-1 and HSPE1 while expression of PTGES3 was determined from one peptide (PTGES3-1).

CHAPTER 3. PROTEOME PROFILING OF BREAST CANCER CELL LINES

3.1 Introduction

The aim of this study was two-fold (i) to establish the basis for precise protein quantification by iTRAQ (isobaric tags for relative and absolute quantification) mass spectrometry, and (ii) to understanding the heterogeneity of breast cancer at a molecular level and refine the taxonomy based on the presence of productive markers such as ER, PR and HER2. For these objectives, we applied iTRAQ based quantification approach on breast cancer cell lines. The iTRAQ labelled peptides were fractionated by isoelectric focusing and liquid chromatography and analysed by two mass spectrometry platforms; Orbitrap Fusion and MALDI-TOF/TOF (Ultraflex II). We have investigated how variance and bias in the iTRAQ reporter ions data are affected by MS platforms. Breast cancers which have similar histopathology behave differently in terms of disease progression and response to treatment. As such, there is still an unmet clinical need to develop biomarkers that can differentiate between different phenotypes of breast cancer. Proteomics provides the means to investigate the protein complement of biological systems and identify quantitative changes in specific components that can be further investigated by a range of independent methods as potential biomarkers.

The advancement in proteomics techniques; methods and instruments, enabled quantitative analysis of complex samples with high sequence coverage of proteins. Several methods for quantification by mass spectrometry exist, both using isotopic labelling and label free approaches (Ong and Mann, 2005,

Bantscheff et al., 2007, Bantscheff et al., 2012). Quantification by isotopic labelling can be performed on precursor ion level like stable isotope labelling by amino acids in cell culture (SILAC) (Ong et al., 2002), or by quantifying isobaric label fragments like Isotope-coded affinity tag (Thompson et al., 2003) or isobaric tags for relative and absolute quantification (iTRAQ) (Ross et al., 2004), in fragment spectra.

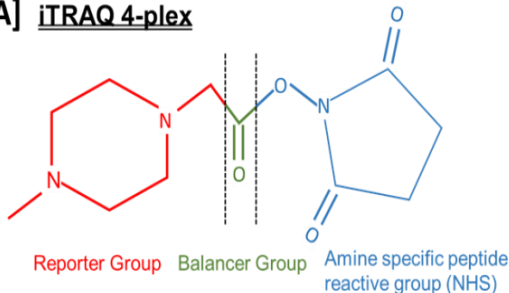
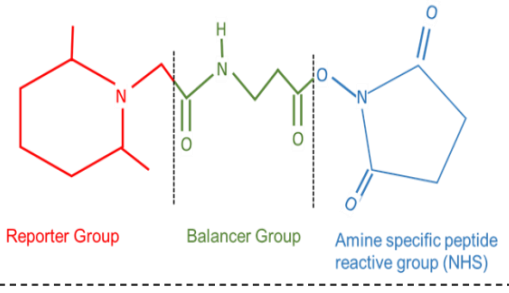
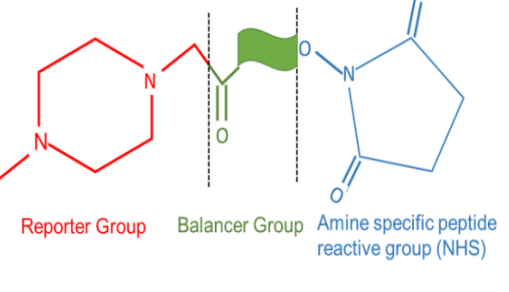
Isobaric mass tags are used for labelling of peptides and generate relative quantitative information in an isobaric labelling-based quantification approach. Isobaric labelling-based quantification, provides the ability to perform high-throughput quantification by sample multiplexing compared to other stable isotope labelling techniques (Rauniyar and Yates III, 2014). A single multiplex experiment eliminates the need to compare multiple LC-MS² data sets, and reduce overall analytical time and run-to-run variation. This type of analytical approach is ideal for studies; such as different stages of cell differentiation, comparisons of normal and diseased environment, comparisons of multiple drug treatments, measurement of inhibitor dose response, or time course comparisons, where whole proteome profiling is important to understand a biological condition (Zhang et al., 2005).

Another advantage of isobaric labelling is dynamic range of identification and quantifications proteins; both high- and low-abundance, and with wide array of physiological properties (Bouchal et al., 2008). The identified and quantified proteome covers diverse molecular weight and pI ranges, functional categories, and cellular locations (Aggarwal et al., 2006, Trotter et al., 2010). MS²/MS³ based quantification is better compared to MS based quantification, because the co-elution of light and heavy peptides in the latter,

can compromise sensitivity and occurrence of multiple precursor ion species create redundancy in MS² data (Aggarwal et al., 2006, Trotter et al., 2010). The isobaric labelled peptides decreased the sample complexity during MS analysis. All peptide labelled with isobaric tags, have same chemical structure and molecular weight, same chromatographic time and with the same peptide mass, which improve the overall signal-to-noise ratios by merging the data of same peptide from different samples, producing good-quality MS² data from low-copy-number proteins (Hardt et al., 2005, Choi et al., 2009). In isobaric labelling based quantification strategy, the peptides are labelled in vitro and can be applied in variety of sample; cell lines, human tissues, biofluid and animal tissue, which make it more efficient approach (Choi et al., 2009, Garbis et al., 2008, Zhou et al., 2009, Zhong et al., 2010, Sutton et al., 2010, Shaheed et al., 2013a).

An isobaric mass tag has three parts; reactive group, balance group and reporter group. The reactive group mostly react with amine but tags react with cysteine residues and carbonyl groups are also present. To achieve the same mass of all tags with different isotopic variants, the balancing group adjusts for the mass difference in the reporter groups. The overall mass of reporter group and balance group of a tag, are kept constant using differential isotopic arrangements with ¹³C, ¹⁵N, and ¹⁸O atoms. The relative intensities of the reporter group provide the quantitative information about peptides between the different samples (Rauniyar and Yates III, 2014). Tandem mass tag (TMT) and isobaric tags for relative and absolute quantification (iTRAQ), are two common commercially available mass tags. TMT reagents (Thermo Scientific), are available with different multiplex options; TMTzero, TMT duplex, TMT 6-plex, and TMT 10-plex, while iTRAQ reagents (AB Sciex) are available as set of 4-plex and 8-plex mass tags.

The 4-plex iTRAQ reagents have reporter ion masses from m/z 114 to 117 and a corresponding balance groups of 28–31 Da, to give a summed tag mass of 145 Da (Ross et al., 2004). The 8-plex reagents have reporter ion masses at m/z 113–119 and 121 with a balance groups to get final tag mass of 305 Da (Sciex, 2017). The structure of the iTRAQ 8-plex balance group has not been published and tag mass 120 is excluded in iTRAQ 8-plex to avoid contamination from phenylalanine ammonium ion (m/z 120.08) (Rauniyar and Yates III, 2014). The Figure 3-1 represents the chemical structured of iTRAQ 4-plex, TMT 6-plex and iTRAQ 8-plex, with isotopic enrichment of ^{13}C , ^{15}N and ^{18}O (Ross et al., 2004, Pichler et al., 2010, Rauniyar and Yates III, 2014, ThermoFisher, 2016, Sciex, 2017). iTRAQ 8-plex allows relative quantification of up to 8 samples within a single experiment.

[A] iTRAQ 4-plex		Tag m/z	Reporter group	Balance group	Reporter + Balance
 <p>Reporter Group Balancer Group Amine specific peptide reactive group (NHS)</p>	114	$C_5^{13}CH_{13}N_2$	$^{13}C^{18}O$	145	
	115	$C_4^{13}C_2H_{13}N_2$	$C^{18}O$	145	
	116	$C_4^{13}C_2H_{13}^{15}NN$	^{13}CO	145	
	117	$C_3^{13}C_3H_{13}^{15}NN$	CO	145	
[B] TMT 6-plex		Tag m/z	Reporter group	Balance group	Reporter + Balance
 <p>Reporter Group Balancer Group Amine specific peptide reactive group (NHS)</p>	126	$C_8H_{16}N$	$^{13}C_4H_4^{15}NO_2$	229	
	127	$C_8H_{16}^{15}N$	$^{13}C_4H_4NO_2$	229	
	128	$C_6^{13}C_2H_{16}N$	$C_2^{13}C_2H_4^{15}NO_2$	229	
	129	$C_6^{13}C_2H_{16}^{15}N$	$C_2^{13}C_2H_4NO_2$	229	
	130	$C_4^{13}C_4H_{16}N$	$C_4H_4^{15}NO_2$	229	
	131	$C_4^{13}C_4H_{16}^{15}N$	$C_4H_4NO_2$	229	
[C] iTRAQ 8-plex		Tag m/z	Reporter group	Balance group	Reporter + Balance
 <p>Reporter Group Balancer Group Amine specific peptide reactive group (NHS)</p>	113	$C_6H_{13}N_2$	ND	305	
	114	$C_5^{13}CH_{13}N_2$	ND	305	
	115	$C_4^{13}C_2H_{13}N_2$	ND	305	
	116	$C_4^{13}C_2H_{13}^{15}NN$	ND	305	
	117	$C_3^{13}C_3H_{13}^{15}NN$	ND	305	
	118	$C_3^{13}C_3H_{13}^{15}N_2$	ND	305	
	119	$C_2^{13}C_4H_{13}^{15}N_2$	ND	305	
	121	$^{13}C_6H_{13}^{15}N_2$	ND	305	

(A) Chemical structure of iTRAQ 4-plex reagent. The complete molecule consists of a reporter group, a mass balance group, and a peptide reactive group (NHS ester). The reporter group ranges in mass from m/z 114–117, whereas the balance group ranges in mass from 28 to 31 Da, such that the combined mass remains constant (145 Da) for each of the four reagents of the iTRAQ 4-plex set. (B) Chemical structure of TMT 6-Plex with reporter group mass range from m/z 126-131. (C) Chemical structure of iTRAQ 8-plex reagent, with tags m/z from 113-121. The balancer group of 8-plex reagents, was never released. **ND**: not defined, **Red** colour fonts; differential isotopic enrichment with ^{13}C , ^{15}N , and ^{18}O atoms.

Figure 3-1: Chemical structures of iTRAQ 4-plex, TMT 6-plex and iTRAQ 8-plex reagents.

Isobaric labelled peptides can be analysed on different type of mass spectrometers; MALDI-TOF/TOF (Wiese et al., 2007, Sutton et al., 2010) Quadrupole (Wolf-Yadlin et al., 2007), LTQ-FT (Köcher et al., 2009), LTQ-Orbitrap (Bantscheff et al., 2008) and Orbitrap fusion (Williamson et al., 2016). The detection of low m/z fragment ions on TOF instruments, are ideal for acquisition of reporter ions but large ion selection window can result compression of signals from reporter ions because of background chemical noise (Keshamouni et al., 2006b). To avoid this type of compression effect on

reporter ions, MS³ mass spectrometry is performed on a hybrid ion trap Orbitrap platform (Ting et al., 2011). In this approach, after CID-MS² fragmentation on an ion trap, the most intense product ion is selected from MS², then analysed on Orbitrap by HCD-MS³ fragmentation (Senko, 2015). This type of instrument, provides an experimental solution to remove background interference, thus eliminating the ratio distortion problem.

The goal of an iTRAQ-based biomarker discovery approach, is to measure quantitative expression of protein level between two or more clinical conditions. It is important to achieve accurate and precise quantitative information from the biological data, and correctly estimate the limitations of the quantification. This is particular relevant when selecting a protein as biomarker for a clinical condition, because further validation in other clinical material is time consuming and costly (White, 2011).

Quantitative analysis of human proteome has some challenges; (a) large biological variations in expression of proteins in a cellular system (b) unknown complexity of proteome under a set of environmental conditions, and (c) large dynamic range for concentration of proteins. The factors contribute to large variety of peptides that can cause bias in the mass spectrometry analysis (Duncan et al., 2010, Hultin-Rosenberg et al., 2013). The precision and accuracy of a quantitation method is effected by systematic errors, that is, differences between true and observed values (bias). Several studies have proved that iTRAQ labelling is related with bias; proteome fold changes are compressed because of background chemical noise and this uncertainty in quantification is caused by co-eluting peptides with similar m/z values that are analysed together, creating mixed iTRAQ intensities in complex samples

(Mahoney et al., 2011, Karp et al., 2010, Ow et al., 2009). The coefficient of variance (CV) of the iTRAQ signal depends on the intensity and results higher CV for low intensity peaks (Mahoney et al., 2011, Karp et al., 2010, Bantscheff et al., 2008, Griffin et al., 2007). The relative quantifications of proteins are calculated from iTRAQ intensities, which are obtained from MS2 spectra of peptides. There are different approaches to calculate a reliable protein ratio by combining the iTRAQ peptide data. Some of these methods to decrease uncertainty in protein quantification, are based on excluding low intensity peptide data (Hu et al., 2006a, Lin et al., 2006), stabilizing the variance (Karp et al., 2010) or weighting the peptide data according to intensity (Hu et al., 2006a, Gan et al., 2007, Li et al., 2012).

Breast cancer is a heterogeneous disease which has more complex histological subtypes in term of disease progression, response to treatment and patient outcomes (Cadoo et al., 2013). Phenotyping by gene expression microarray has provided a powerful classification of breast cancer (Mackay et al., 2011) that correlate with the origins of normal breast cell development, elegantly portrayed by Prat and Perou in 2011, Figure 3-2 (Prat and Perou, 2009, Prat and Perou, 2011).

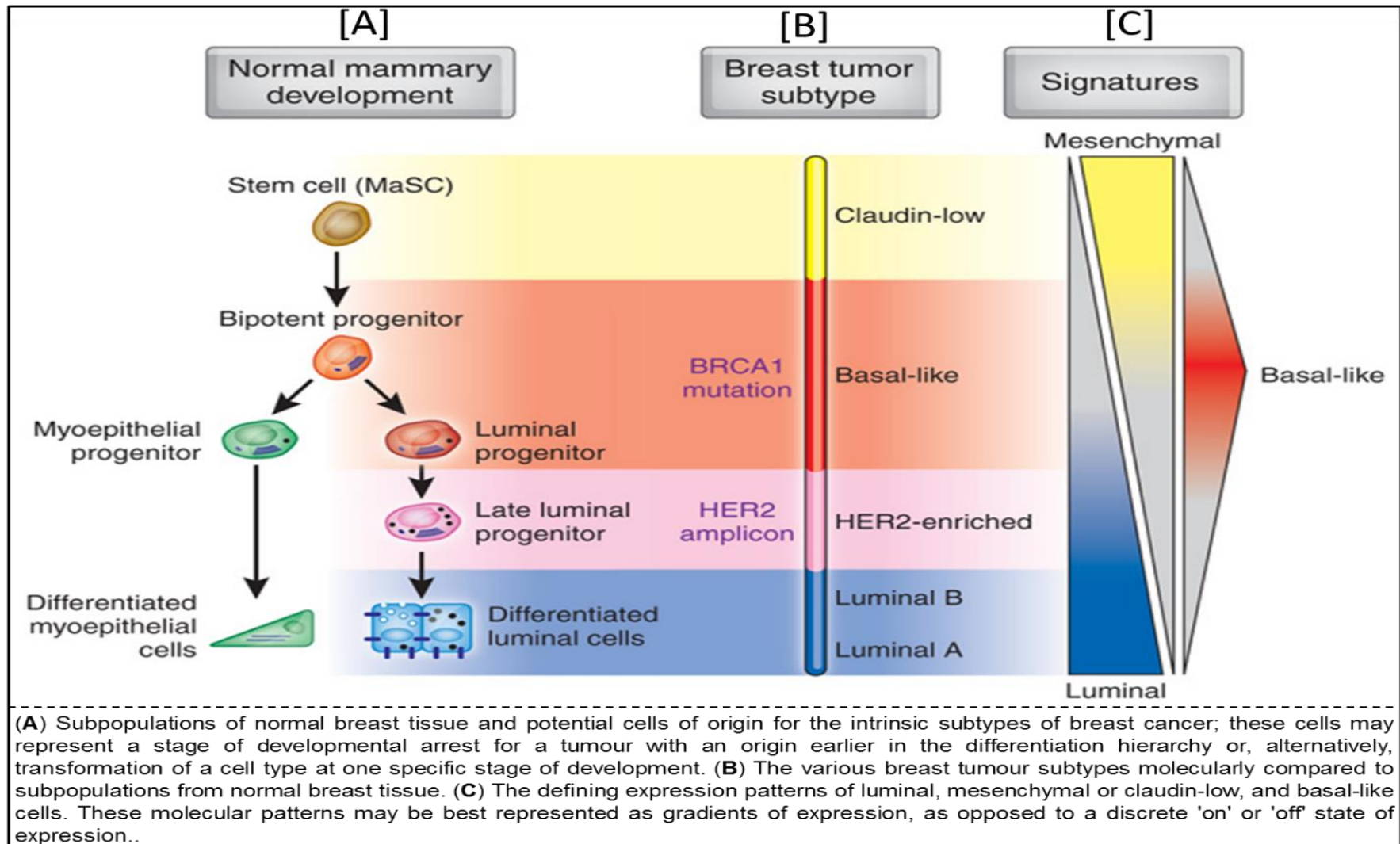


Figure 3-2: Model of the human mammary epithelial grading linked to breast cancer subtypes (Prat and Perou, 2009).

Stratification has provided researchers and clinicians with valuable information to predict the risk of developing the disease, treatment regimen for patients, resistance to treatment and susceptibility to relapse thereby contributing towards improved personalised treatment with reduced side effects. Phenotyping, based on characteristic gene product profiles including unique classifier genes, initially identified four groups (Perou et al., 2000), which was subsequently increased to five classifications – luminal A, luminal B, HER2 positive, basal-like and claudin-low (Holliday and Speirs, 2011).

This not only applies to patient's tumour profiles but categorisation of established breast cancer cell lines (Holliday and Speirs, 2011). Cell lines are an essential tool for the study of breast cancer biology, enabling development of in vitro (e.g. monolayers, 3D spheroids, co-cultures, drug resistance), in vivo (e.g. xenograft) and in situ (e.g. knock-in, knock-out) models, particularly in deciphering our understanding of molecular mechanisms. The ability to grow cell lines, under control conditions, provides researchers with a valuable resource for biological material. However, the artificial transformation/selection of cells for their ability to proliferate indefinitely is not without consequence in altering their molecular make-up, which may not fully reflect the original primary cells. At the time of publication/writing, there are approximately 55,084 published articles (PubMed, (NCBI, 2017)) papers that reference breast cancer cell lines, some of which have played a fundamental part in disease stratification using comprehensive genomic and proteomic strategies. For example, the Cancer Cell Line Encyclopaedia includes comprehensive mRNA microarray profiling for 58 breast cancer cell lines (<http://www.broadinstitute.org/ccle/home>) (Barretina et al., 2012). Nevertheless, mRNA provides only one part of the picture in terms of the molecular events that describe cancer. Proteins, the

product of mRNA translation, play an integral role in the oncogenic processes that determine transformation, proliferation, survival, migration, invasion, and hence have been a rich seam of investigation to identify markers for early detection of the disease, disease stage, targets for drug development and indicators of response to treatment (toxicology). However, there has been considerable debate regarding the relationship between expression at the mRNA and protein levels (Gry et al., 2009, Maier et al., 2009, Cifani et al., 2015), which for the most part has indicated limited quantitative correlation. The work of Cifani et al, in particular, concluded that there was no correlation between mRNA profiles that delineate the established phenotypes and the equivalent proteins using breast cancer cell lines, especially those genes defined as classifiers (Cifani et al., 2015). However, one challenge of this study was to use one cell type, MDA-MB-231, (a claudin-low subtype of the basal-like phenotype) as the reference cell line, which would distort potential profile aberrations constituted by cell line immortalisation.

A recent study by Calderón-González and his co-workers identified 1,020 proteins in breast cell lines by iTRAQ 4-plex labelling and tandem mass spectrometry approach (Calderón-González et al., 2015). They compared the proteome of MCF7 and T47D (luminal A), MDA-MB-231 (claudin low) and SK-BR-3 (HER2+) breast cancer cell lines with MCF-10A (normal-like) and proposed BAG6, DDX39, ANXA8 and COX4 as putative biomarkers in breast cancer. A label free proteomics approach conducted by Laurence in 2015, identified 12,000 different proteins in 20 breast cancer cell lines and 4 tissues biopsies (triple negative) using mass spectrometry (Lawrence et al., 2015). This study was focused on comparison of proteomes of triple negative breast cancer tissues with cell lines (tumour and normal-like). A recent study on

88 luminal-type breast cancer tissues, collected at different stages of breast progression, identified average of 5,439 proteins identified in each tissue (Pozniak et al., 2016). In this study, they also compared the quantitative proteome of MCF-7 (ER+) with normal mammary epithelial cells (HMEC), to understand the synthesis, degradation, and turnover of proteins in ER-positive cancer cells.

To date, there has been no study to compare the proteome of different phenotype of breast cancer cell lines with normal mammary epithelial cells (HMEC), to identify phenotype specific signatures. This study was an iTRAQ experiment aimed at (a) identifying protein-specific profiles in human breast cancer cell lines representative of established phenotypes (luminal A, luminal B, HER2, basal-like and claudin-low), and (b) comparing the dynamics of iTRAQ quantification on two MS platforms (Ultraflex II and Orbitrap Fusion).

3.2 Materials and methods

3.2.1 Breast cell lines

Breast cancer cell lines (MCF-7; ZR-75; MDA-MB-468; MDA-MB-231; MDA-MB-453) and Non-tumorigenic breast cells (MCF-10A and HB2) were grown according recommended conditions (please refer to chapter 2 full details). Human mammary epithelial cells (HMEC) were isolated from biopsy material and ethical approval was given by Leeds (East) Research Ethic Committee, reference 07/H1306/98+5. All cell lines were tested for mycoplasma contamination using MycoProbe mycoplasma detection kit. The authenticity of MCF-7, MCF-10A, MDA-MB-468 and MDA-MB-231 cell lines, was confirmed by DNA typing (DDC Medical). All breast cancer cell lines were maintained in standard growth medium at 37°C and 5% CO₂.

3.2.2 Protein extraction from cell lines

Cells were harvested by trypsin treatment on reaching 80-90% confluence and washed three times with PBS and resulting cell pellet (5 × 10⁶ cells) was resuspended in 100 µL of urea extraction buffer (7 M urea, 2 M thiourea, 4% w/v CHAPS, 50 mM DTT in PBS pH7.4 containing protease inhibitor cocktail) was added, vortexed, sonicated, and then centrifuged. The protein concentration of each cell line extract was measured using the Bradford assay (Bradford, 1976).

3.2.3 LC-MS analysis

3.2.3.1 Trypsin digestion and iTRAQ labelling

Each extract (200 µg of protein) was precipitated overnight with 100% acetone at -20°C and centrifuged for 20 min at 13 400 rpm at 4°C. The pellet was resuspended in 8M urea in 400mM ammonium bicarbonate (Ambic). After protein measurement by Bradford assay, each protein sample (80 µg of protein) was reduced with 50 mM dithiothreitol (DTT) for 15 min at 60°C, alkylated with 100 mM iodoacetamide (IAA) at ambient temperature for 15 min, and digested with 3 µL of a 1 mg/mL solution of modified sequencing grade trypsin at 28°C for 20 hrs. After digestion, each sample was desalted on an Isolute C18 RP LC column, lyophilized, resuspended in 1M TEAB, 0.1% SDS and an iTRAQ reagent added as outlined in Table 3-1, for 2 hrs at room temperature. The labelled peptides were then combined, desalted on an Isolute C18 RP LC column, and the eluate lyophilized.

3.2.3.2 Peptide fractionation

The total iTRAQ-labelled peptide sample was resuspended in OffGel peptide sample buffer and applied to an OffGel 3100 isoelectric focussing system using a pH 3-10 high-resolution strip for 50 kV hours. Twenty-four fractions were collected, desalted on Isolute C18 RP cartridges and lyophilized. Figure 3-3 illustrates the workflow for the 2D Gel-LC-MS analysis of breast cell lines on two MS platforms.

Cell Line	Cell Type	Description	iTRAQ Label
MCF-7	Luminal A	Adenocarcinoma, ER+, PR+/-, HER2-	113
ZR-75	Luminal B	Ductal Carcinoma, ER+, PR+/-, HER2+	114
MDA-MB-468	Basal-like	Adenocarcinoma, EGFR+, TGF α +, ER-, PR-, HER2-	115
MDA-MB-231	Claudin-Low	Adenocarcinoma III, EGFR+, TGF α +, ER-, PR-, HER2-	116
MDA-MB-453	Luminal-HER2	Metastatic Carcinoma, FGR+, ER-, PR-, HER2+	117
MCF-10A	Normal-Like	Non-tumorigenic mammary epithelial cell line	118
HB2	Normal-Like	Non-tumorigenic mammary epithelial cell line	191
HMEC	Healthy	Normal human mammary epithelial cells	121

ER; estrogen receptor, **PR**; progesterone receptor, **HER2**; human epidermal growth factor, **EGFR**; Epidermal growth factor receptor, **TGF- α** ; Transforming growth factor alpha, **FGR**; Tyrosine-protein kinase FGR, (+); positive, (-); Negative

Table 3-1: Breast cell lines attributes and iTRAQ labelling.

3.2.3.3 Ultraflex II analysis

Each lyophilized OffGel fraction was resuspended in 30 μ L of 10% acetonitrile and 0.05% TFA, 5 μ L injected on to an LC Packings UltiMate 3000 nano HPLC system, washed on a C18 pre-column before transfer to analytical column. A total of 384, fractions were co-deposited with 0.4 μ L of a saturated α -cyano-4-hydroxy cinnamic acid (CHCA) matrix solution onto a MTP AnchorChip 800/384 target plate using a Proteineer FC fraction collector and allowed to air-dry. Peptide Calibration Standard II was applied between each group of four fractions. Mass spectrometric analysis was carried out using a

MALDI–TOF/TOF Ultraflex II instrument. A fully automated workflow was performed which encompassed data acquisition, data-processing, compilation of a non-redundant list of peptides from the 384 HPLC fractions. Duplicate LC–MALDI analyses were performed for each OffGel fraction (for full details, please see chapter 2, section 1.9.1).

3.2.3.4 Fusion Orbitrap analysis

Lyophilized peptide fractions were individually reconstituted in 30 μ l of loading mobile phase and 2 μ l loaded onto a Dionex Ultimate 3000 nanoLC for 4 minutes on a C18 trapping cartridge (300 μ m \times 5 mm, 5 μ m diameter, 100 Å PepMap pre-column), which was then brought in line with an analytical column (75 μ m \times 15 cm, 3 μ m diameter, 100 Å PepMap column) at a flow rate of 300 nl/minute. Several reverse phase elution gradient lengths were used (please see the details in chapter 2), proportionally extrapolated from the 120-minute gradient. MS characterisation of eluting peptides was conducted on an Orbitrap between 350 and 1500 m/z at 120,000 mass resolution and all MS² acquisition was performed on the Ion-trap, in top speed mode with a 3s cycle time. Automated Synchronous Precursor Selection (SPS) for MS³ setting was used for quantification of iTRAQ, a special feature of the Orbitrap Fusion which improves quantitative accuracy when using isobaric mass tags (Chapter 2, Section 2.9.2 for details of the LC MS method).

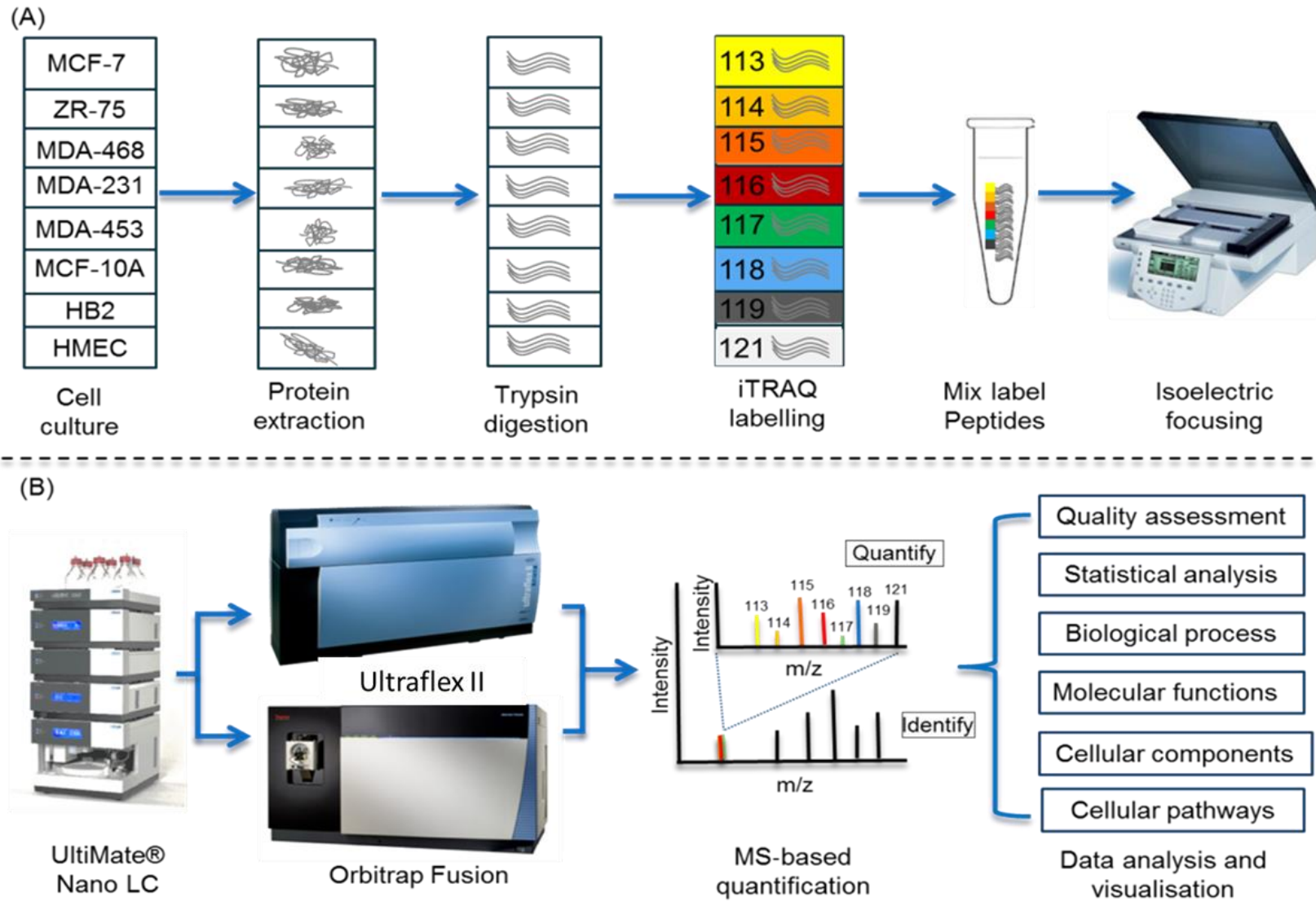


Figure 3-3: Workflow illustrating the proteomics-based approach.

3.2.3.5 MS data processing

Ultraflex II data was searched, via ProteinScape v3.0, and Orbitrap Fusion data was searched through Proteome Discoverer 2.1, by using same search method (Mascot software version 2.4 Matrix Science, U.K.) against SwissProt version 2016 containing 552,259 human protein sequences with search parameters: trypsin digestion, 2 missed cleavages, variable modification of methionine oxidation, fixed modifications of cysteine (carbamidomethylation) and iTRAQ (lysine and N-termini). For Ultraflex II; a precursor mass tolerance of 100 ppm, fragmentation mass tolerance of 0.7Da was used while on Orbitrap Fusion; a precursor mass tolerance of 10 ppm, fragmentation mass tolerance of 0.6 Da was selected. The confidence interval threshold was set to P value <0.05, which was equal to Mascot score ≥ 29 on Ultraflex II and was ≥ 22 on Orbitrap fusion. The schematic representation of the MS data processing workflow is shown in Figure 3-4. iTRAQ labelling efficiency was determined by searching MS/MS data using iTRAQ as a variable modification, and performing a survey of labelled and non-labelled peptides in the 20 highest scoring proteins from individual LC-MS experiments. Non-redundant protein profiles for each experiment were created in Proteome Discoverer/ ProteinScape by combining the corresponding LC-MS datasets. The list of protein identifications was assessed manually, and all proteins defined as Master Protein Candidates that could not be differentiated from the Master Protein (i.e. no unique peptides) were omitted.

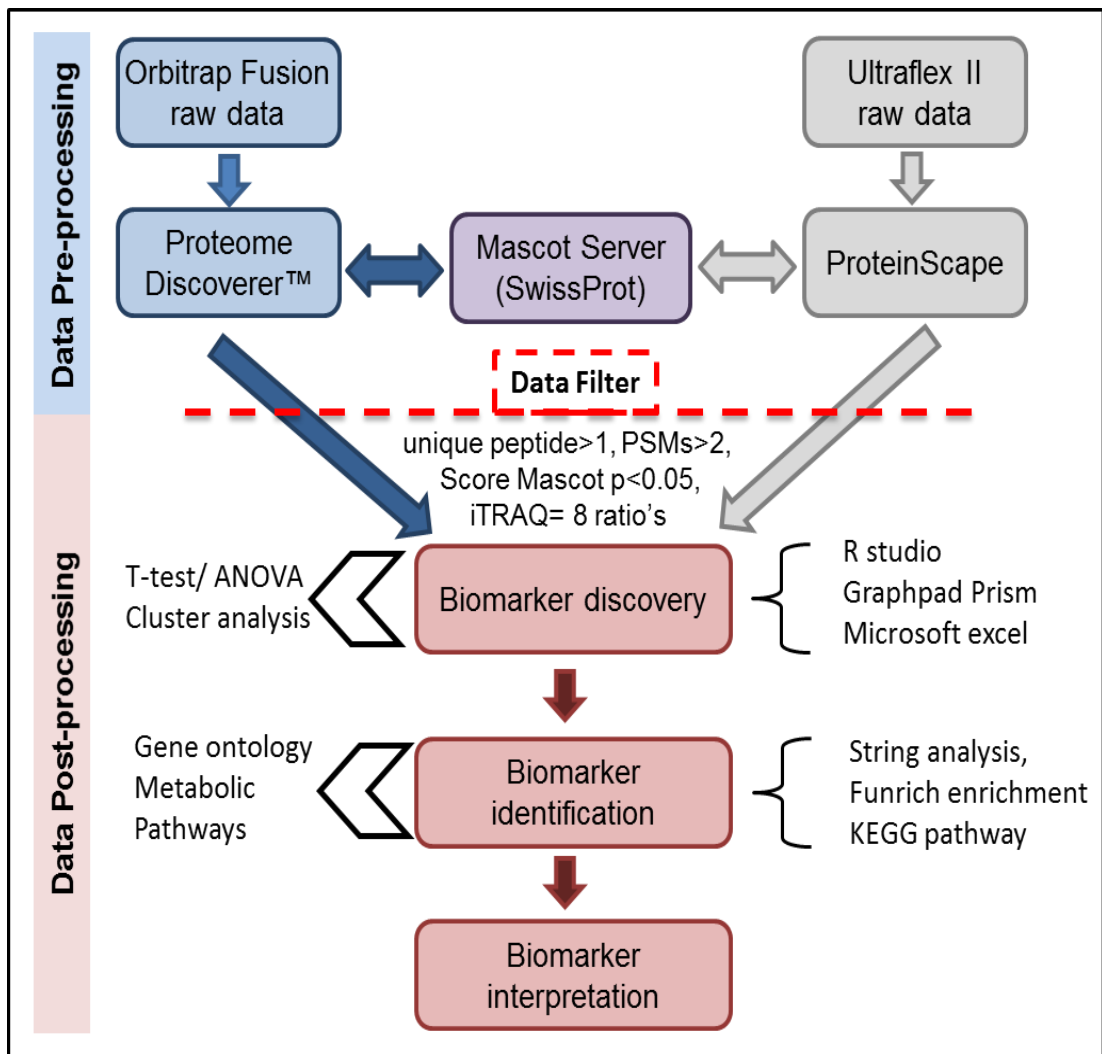


Figure 3-4: Schematic representation of the MS data processing workflow.

3.2.3.6 Statistical analysis

For each dataset (Ultraflex II and Orbitrap Fusion), the median ratio was determined for each cell line, and individual protein ratios normalised (relative to the median ratio), to allow for experimental variation and enable comparison between cell lines and instruments. The ratios were converted to log₂ and those components that were significantly changed, compared to primary human breast epithelial cells, were defined by using LIMMA statistics in R Studio (see Chapter 2, Section 2.13). PRISM 6.0 software was used for graphical presentation of data. All differentially expressed proteins were classified broadly into several groups according to the Gene Ontology (GO) annotation.

Protein–protein interactions and network associations for significantly altered components were performed using STRING.

3.2.4 Western blot analysis

Cell line extracts, equivalent to 20 µg of protein, were mixed with 5 µL of SDS reducing buffer (Laemmli buffer), heated to 60°C for 15 min, and on cooling were applied to SDS polyacrylamide gels. Proteins were transferred to a nitrocellulose membrane by electro-blotting. The blotted membrane was incubated in blocking buffer and then the appropriate primary antibody (Chapter 2, Table 2-2) in blocking buffer was incubated overnight at 4 °C. After removal of the primary antibody with three TBS-Tween buffer washes, the appropriate secondary antibodies (TABLE 2-2) and proteins detected by the addition of ECL Plus Western Blotting reagent, then exposed to X-ray film. (see Chapter 2, Section 1.7.4 for full details of western blot assay).

3.3 Results

3.3.1 Comparison of protein and peptide identification from two MS platforms

The experimental outline for data acquisition can be seen in Figure 3-3 and results of protein identifications from Orbitrap Fusion and Ultraflex II are summarized in Table 3-2. A total of 2013 unique proteins were identified by using the Orbitrap Fusion with an average 6.25 peptides per protein whereas a total of 1884 unique proteins were identified on the Ultraflex II with an average of 4.94 peptides per protein. Hence, the results showed that the Orbitrap Fusion identified more proteins and at a greater depth of information content. There was a significance (p value <0.0001), on the performance of both instruments, for total number of PSMs, peptides, Mascot score and sequence coverage of identified proteins. The raw iTRAQ ratios (non-normalised), were also different, with p value of <0.0001 , for all quantifiable proteins on both platforms (Table 3-2). The MS acquisition time for each OffGel fraction was 2 hours on Orbitrap Fusion (online), while on Ultraflex II was 13 hours; 2 hours for fraction collection on MTP AnchorChip target (offline) and 11 hours for acquisition of average 3,000 MS^2 spectra from each OffGel fraction (Table 3-2).

The average fold change range (\log_2) on Ultraflex II, was from -3.38 to 2.12, while on Orbitrap Fusion was from -4.76 to 6.42 (Figure 3-6). iTRAQ ratios of all common proteins ($n=1430$) in both instruments (Figure 3-5), were median normalised and changed in to \log_2 values to compare the dynamic range of fold change.

	Orbitrap Fusion	Ultraflex II	P-value
Total MS/MS	318,730	247,236	N/A
Total PSMs	29,298	16,706	***
Total Peptides	12,575	9,313	***
Total Proteins	2,013	1,884	N/A
MS/MS count per PSMs*	10.88	14.80	N/A
PSMs count per peptide*	2.33	1.79	N/A
Peptides count per protein*	6.25	4.94	N/A
Sequence coverage*	18.10	13.61	***
Mascot score*	352.49	202.23	***
Raw data 113/121**	0.81	1.46	***
Raw data 114/121**	0.74	1.48	***
Raw data 115/121**	0.65	0.89	***
Raw data 116/121**	0.56	1.39	***
Raw data 117/121**	0.66	1.84	***
Raw data 118/121**	0.41	2.16	***
Raw data 119/121**	0.77	1.62	***
MS acquisition time (hrs)[§]	2	13	***

The table presents the number of MS/MS, **PSMs** (peptide spectrum matches), peptides and number of proteins identified with the different instruments using Mascot search engine v2.4 at p<0.05. (*); Average, (**); non-normalised median ratios ([§]); Average per fraction (***) p value < 0.0001, **N/A**; not applicable.

Table 3-2: Proteome data comparison of instruments.

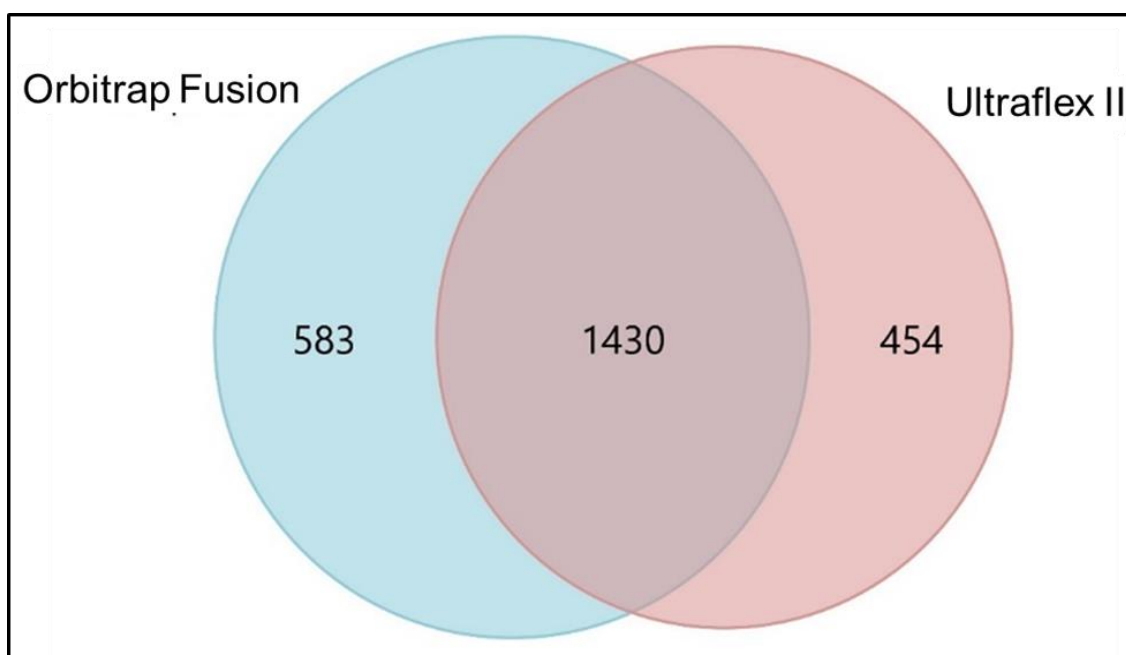


Figure 3-5: Comparison of Ultraflex II and Orbitrap Fusion proteome.

After median normalisation of iTRAQ ratios for each sample; paired t-test statistics was applied to compare fold change of each protein (common data only) on orbitrap Fusion and Ultraflex II. Dynamic range compression of iTRAQ ratios (113/121; p value 0.00242, 114/121; p value 0.00453, 115/121; p value <0.0001, 116/121; p value 0.00306, 117/121; p value <0.0001, 118/121; p value 0.0131 and 119/121; p value <0.0001), was observed in MS² data acquired on the Ultraflex II (Figure 3-6) compared to Orbitrap Fusion. The synchronous precursor selection (SPS) function of the Orbitrap Fusion for MS³ analysis of the reporter ions reduced ratio compression (increases dynamic range) (McAlister et al., 2014, Williamson et al., 2016) (Figure 3-6) and provided a more accurate picture of the protein expression changes compared to the Ultraflex II.

Pearson correlation coefficient statistics was applied on both MS data sets (n=1430) to compare the slope of the linear regressions. Where SPS method is compared with MS² method, good iTRAQ log₂ ratio correlation was observed between the two instruments for all iTRAQ ratios except 118/121; (113/121 R² = 0.603, 114/121 R² = 0.595, 115/121 R² = 0.600, 116/121 R² = 0.500, 117/121 R² = 0.297, 118/121 R² = 0.086, 119/121 R² = 0.315) (Figure 3-7).

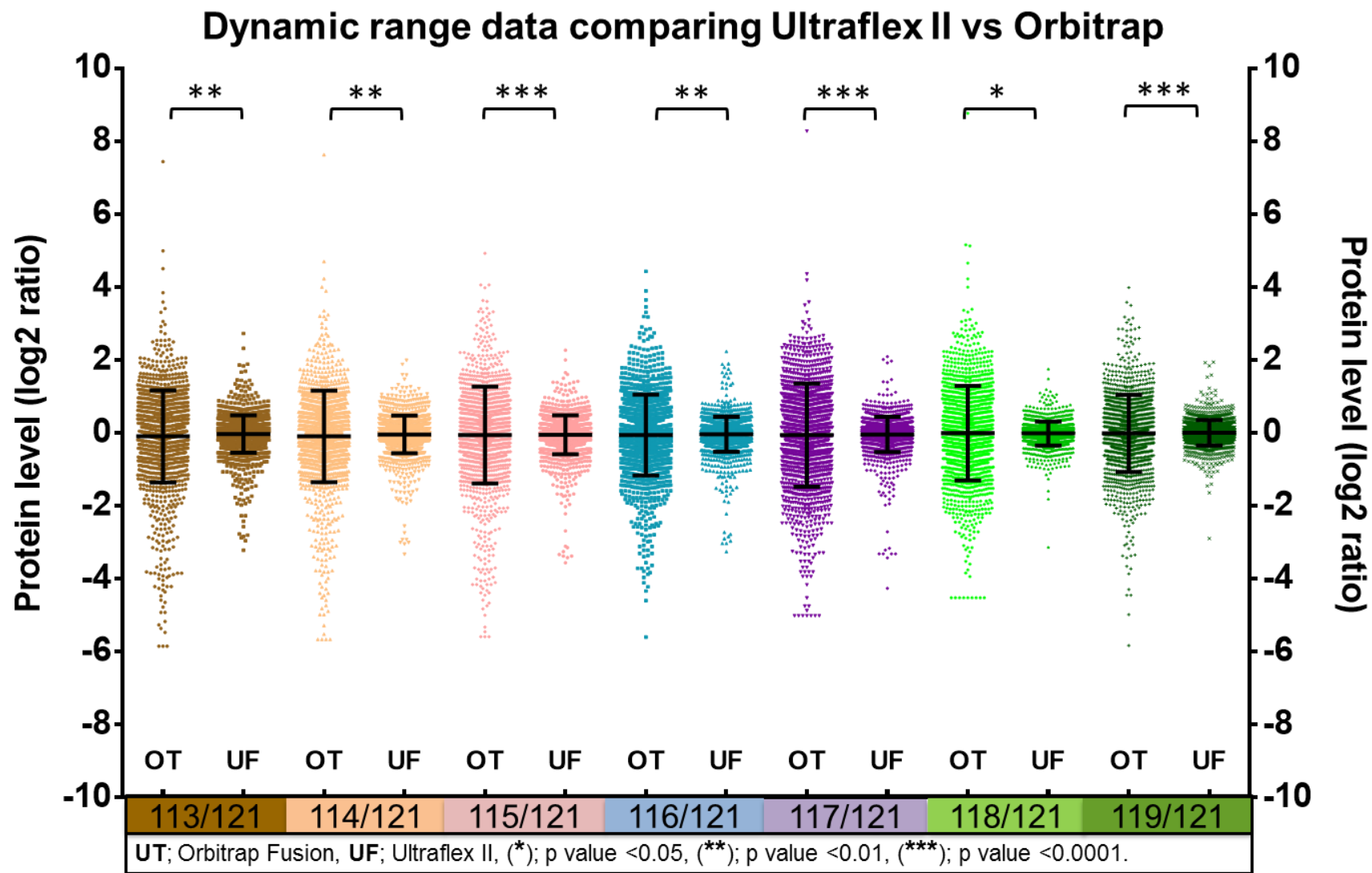


Figure 3-6: Dynamic range of proteins identified with Ultraflex II vs Orbitrap Fusion.

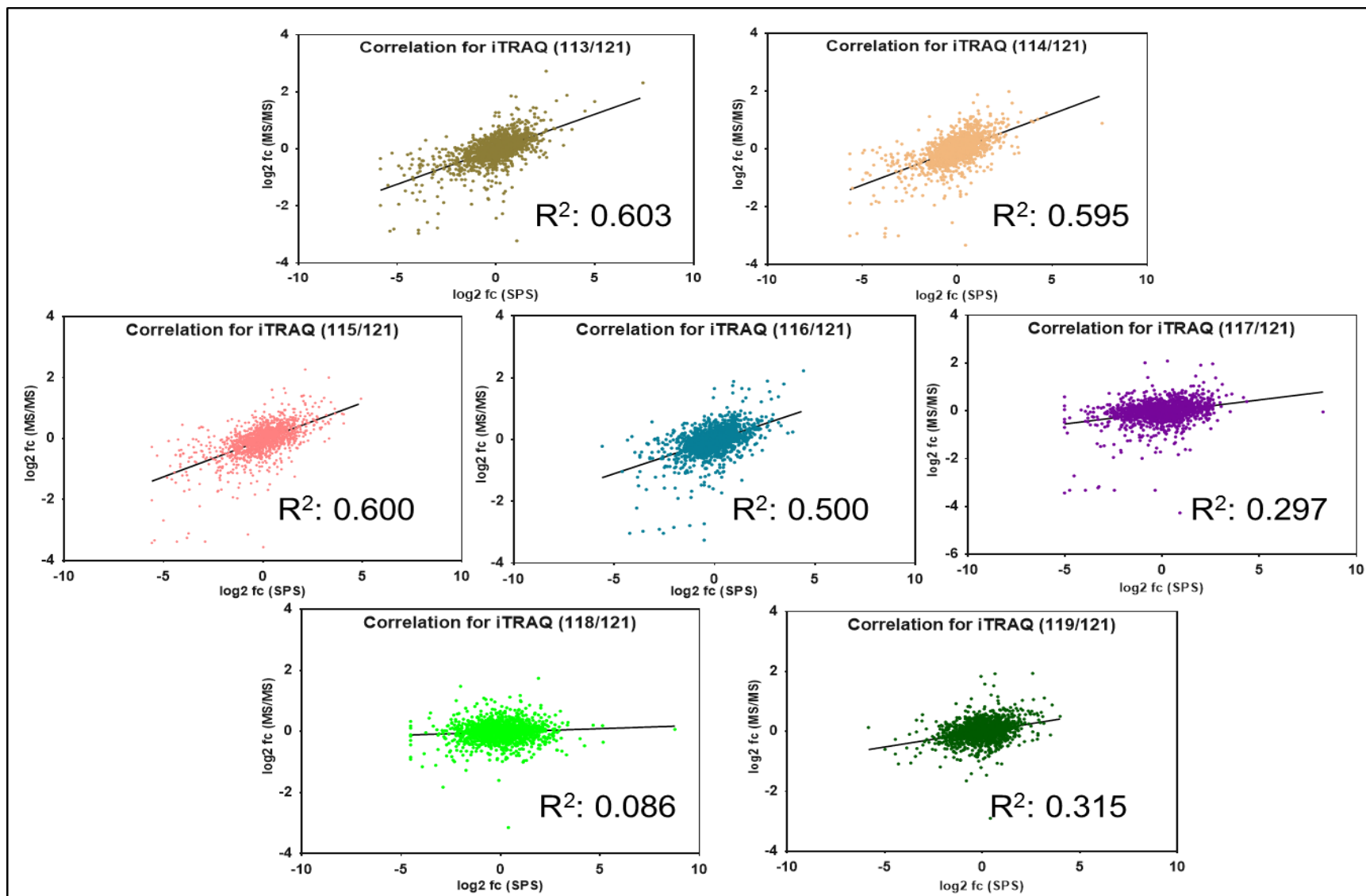


Figure 3-7: Pearson correlation coefficient between \log_2 ratios of common proteins in both MS datasets – (MS/MS; Ultraflex ratio and SPS; Orbitrap Fusion ratios).

3.3.2 Quantitative comparison of breast cell lines proteome

A panel of breast cell lines was used to characterise the proteomes of Ultraflex II and Orbitrap Fusion, with the aim of identifying the dynamic range compression of MS² data. Cell lines for each breast cancer subtype; luminal A represented by MCF-7, luminal B by ZR-75, HER2 by MDA-MB-453, basal-Like by MDA-MB-468, claudin-low by MDA-MB-231, were included, and premalignant cell lines were exemplified by MCF-10a (basal) and HB2 (luminal-like). Primary human mammary epithelial cells (HMEC) were used to exemplify normal, healthy cells and as a reference for comparison of protein expression changes in cell lines in an iTRAQ-based proteomic analysis.

To understand this effect, some pre-knowledge about the heat shock response in human cell lines was presumed i.e. the expression of certain heat shock proteins (HSPs), is changed under stressful conditions such as exposure to cold or heat, UV light and wound healing or tissue remodelling (Morimoto, 1993, Jago et al., 2013, Lianos et al., 2015). The family of HSP consists of 29 members, and many members of the family play a role as chaperone; folding or unfolding of proteins that were damaged by cell stress (Uniprot.org, 2015). Therefore, heat shock proteins are useful control proteins to compare SPS and MS² quantitative method. A total of 15 heat shock proteins in all 7 cell lines and of the total ratios calculated by MS² and SPS, 93% showed the same trend (Table 9-3). However, SPS generated 14 measurements that correlated with significant changes (median \pm standard deviation) whereas only 4 were changed by MS², indicating that greater dynamic range of the former compared to the latter.

No.	Gene name	Accession	Description	Method	MCF-7	ZR75	MDA-MB-453	MDA-MB-468	MDA-MD-231	MCF-10A	HB2
1	HSPE1	P61604	10 kDa heat shock protein, mitochondrial	SPS	1.08	1.13	1.00	0.61	2.25	1.92	1.23
				MS2	0.46	0.44	0.34	0.22	0.12	0.19	-0.11
2	HSPD1	P10809	60 kDa heat shock protein, mitochondrial	SPS	1.27	1.02	1.28	1.03	1.30	1.31	0.91
				MS2	0.21	0.00	0.16	-0.03	-0.13	0.08	-0.10
3	AHSA1	Q95433	Activator of 90 kDa heat shock protein ATPase homolog 1	SPS	0.82	0.17	0.73	0.39	0.43	0.31	0.29
				MS2	0.09	0.05	0.21	0.02	0.07	-0.34	-0.20
4	HSPA4	P34932	Heat shock 70 kDa protein 4	SPS	-0.38	-0.42	-0.53	0.15	-0.66	-0.73	-0.15
				MS2	0.04	0.00	-0.04	0.16	0.06	-0.22	0.12
5	HSPA8	P11142	Heat shock cognate 71 kDa protein	SPS	0.39	0.22	0.35	0.46	0.96	0.36	0.73
				MS2	-0.11	-0.34	-0.27	-0.25	-0.16	-0.02	-0.06
6	HSPH1	Q92598	Heat shock protein 105 kDa	SPS	0.33	0.06	0.84	0.49	-0.67	-0.42	0.47
				MS2	0.19	0.22	0.36	0.25	0.13	-0.25	-0.15
7	TRAP1	Q12931	Heat shock protein 75 kDa	SPS	1.20	1.39	1.30	1.31	0.83	1.68	1.82
				MS2	0.23	0.19	0.18	0.32	0.05	0.21	0.05
8	HSPB1	P04792	Heat shock protein beta-1	SPS	1.38	0.68	-2.25	-0.40	-0.71	2.22	-1.23
				MS2	0.12	-0.49	-1.71	-0.93	-1.43	-0.41	-0.43
9	HSP90AA1	P07900	Heat shock protein HSP 90-alpha	SPS	1.25	0.95	1.19	0.51	0.58	0.53	0.68
				MS2	0.46	0.13	0.38	0.29	0.19	0.06	0.32
10	HSP90AB1	P08238	Heat shock protein HSP 90-beta	SPS	0.94	-0.32	0.95	1.74	1.22	0.28	0.42
				MS2	0.31	-0.17	0.23	0.37	0.14	-0.40	-0.43
11	HSP90AB4P	Q58FF6	Putative heat shock protein HSP	SPS	1.27	0.70	2.04	0.61	1.55	0.46	0.69
				MS2	0.19	-0.36	0.19	0.31	0.07	-0.03	0.07
12	HSPA5	P11021	78 kDa glucose-regulated protein	SPS	-0.79	-1.24	-0.88	0.75	-1.18	-0.35	-0.43
				MS2	0.02	-0.25	-0.18	0.02	-0.24	0.28	0.65
13	HSPA9	P38646	Stress-70 protein, mitochondrial	SPS	0.44	0.21	0.34	0.27	0.30	0.29	0.69
				MS2	0.32	0.25	0.14	0.11	0.20	0.05	0.06
14	HSP90B1	P14625	Endoplasmic	SPS	-0.51	-0.59	-0.46	0.64	-0.50	-0.39	0.10
				MS2	-0.11	-0.30	-0.16	-0.20	-0.20	0.01	-0.09
15	HYOU1	Q9Y4L1	Hypoxia up-regulated protein 1	SPS	0.58	-0.10	0.24	1.33	0.36	0.65	0.48
				MS2	0.19	-0.04	0.02	0.09	-0.11	-0.20	0.11

The iTRAQ ratios highlighted in green indicate upregulated (Median + S.D) and in red indicate down-regulated (Median - S.D) proteins in one or both MS data sets. **SPS**; Orbitrap Fusion, **MS2**; Ultraflex II

Table 3-3: Median normalised iTRAQ ratios of heat shock proteins in breast cell lines, acquired on Orbitrap Fusion (SPS) and Ultraflex II (MS²).

3.3.3 Evaluation of Orbitrap Fusion data

Overall there is a strong association between the two datasets, but SPS mode data with large dynamic range and protein IDs, was further evaluated for proteome profiling. In this study, breast cancer cell lines; MCF-7, ZR-75, MDA-MB-453, MDA-MB-468, MDA-MB-231, were included, and premalignant cell lines were exemplified by MCF-10a and HB2. Primary human mammary epithelial cells (HMEC) were used to exemplify normal, healthy cells and as a reference for comparison of protein expression changes in cell lines in an iTRAQ-based proteomic analysis.

A total of 2013 proteins from unique genes (excluding SNP and splicing variants) were identified with full iTRAQ ratios, of which 76% did not show a significant difference (within the Median range \pm 1 standard deviation) in expression in any cell line compared to HMEC. An average of 480 proteins were significantly increased or decreased in at least one of the cell lines compared to HMEC. Those proteins that were changed in all cell lines compared to HMEC were identified; all increased - 18 proteins (Table 3-4) or all decreased - 56 proteins (Table 3-5). Detailed GO-annotations for all up and down-regulated proteins are presented in supplementary data "All up and down compared to HMEC-Go-Analysis" as CD format.

No	Accession	Gene ID	Description	MW [kDa]	PSMs	Peptides	Unique Peptides	Coverage	Score Mascot	MCF-7 (LOG2)	ZR-75 (LOG2)	MDA-MB-468 (LOG2)	MDA-MB-231 (LOG2)	MDA-MB-453 (LOG2)	MCF-10A (LOG2)	HB2 (LOG2)
1	P06493	CDK1	Cyclin-dependent kinase 1	34	13	5	4	17	84	1.64	2.20	3.19	4.19	3.32	1.71	2.97
2	Q9H773	DCTPP1	dCTP pyrophosphatase 1	19	12	3	3	19	49	2.50	2.56	1.88	1.83	3.40	2.65	2.82
3	Q13268	DHRS2	Dehydrogenase	30	5	1	1	6	176	7.44	7.63	8.27	6.65	4.23	8.76	6.58
4	P49736	MCM2	DNA replication licensing factor MCM2	102	29	10	10	15	540	2.17	1.66	2.28	1.53	3.02	1.83	3.10
5	P25205	MCM3	DNA replication licensing factor MCM3	91	16	8	8	9	155	2.30	1.67	2.74	2.03	3.27	2.00	3.15
6	P33992	MCM5	DNA replication licensing factor MCM5	82	11	6	6	10	134	2.42	1.72	2.35	1.76	3.61	1.71	3.58
7	P36954	POLR2I	DNA-directed RNA polymerase II subunit RPB9	15	4	2	2	14	33	3.62	2.09	1.97	1.70	2.24	5.38	2.44
8	Q9BUP0	EFHD1	EF-hand domain-containing protein D1	27	4	3	2	13	33	2.09	5.09	4.41	2.50	2.51	2.78	1.74
9	P22626	HNRNPA2B1	Heterogeneous nuclear ribonucleoproteins A2	37	248	24	20	69	4914	1.73	1.53	1.90	2.35	1.63	2.54	1.38
10	P26583	HMGB2	High mobility group protein B2	24	18	4	2	17	313	1.95	1.68	1.48	3.31	2.38	3.01	2.05
11	P0C0S5	H2AFZ	Histone H2A.Z	14	61	3	1	23	336	1.93	1.59	1.83	2.26	2.03	2.77	2.67
12	O60814	HIST1H2BK	Histone H2B type 1-K	14	154	8	4	53	2614	2.27	1.97	2.11	2.39	2.50	3.36	2.48
13	Q8N257	HIST3H2BB	Histone H2B type 3-B	14	132	7	3	52	2470	2.46	2.76	2.02	2.98	2.89	3.54	2.33
14	P20700	LMNB1	Lamin-B1	66	76	17	15	32	662	1.81	1.72	1.81	2.30	2.03	2.85	1.58
15	Q13247	SRSF6	Serine	40	33	8	8	22	202	1.70	1.51	1.91	2.59	2.13	2.83	1.89
16	P16949	STMN1	Stathmin	17	36	8	8	46	751	1.54	2.44	3.31	4.35	2.87	2.63	2.38
17	P55327	TPD52	Tumor protein D52	24	6	4	4	21	56	1.98	2.17	2.01	2.05	1.84	2.56	1.22
18	Q9Y2W2	WBP11	WW domain-binding protein 11	70	9	2	2	5	246	1.55	2.33	2.36	2.29	1.84	2.87	1.49

Table 3-4: Significantly up-regulated proteins in cancer and premalignant cell lines compared to primary human mammary epithelial cells.

No	Accession	Gene ID	Description	MW [kDa]	PSMs	Peptides	Unique Peptides	Coverage	Score Mascot	MCF-7 (LOG2)	ZR-75 (LOG2)	MDA-MB-468 (LOG2)	MDA-MB-231 (LOG2)	MDA-MB-453 (LOG2)	MCF-10A (LOG2)	HB2 (LOG2)
1	Q04446	GBE1	1,4-alpha-glucan-branching enzyme	80	8	5	5	8	144	-2.83	-2.87	-2.97	-2.25	-2.61	-3.04	-2.33
2	O15143	ARPC1B	Actin-related protein 2/3 complex subunit 1B	41	10	4	4	12	74	-1.37	-1.88	-1.46	-2.25	-3.70	-2.49	-2.22
3	Q9Y6K8	AK5	Adenylate kinase isoenzyme 5	63	8	2	1	4	59	-5.85	-4.52	-2.15	-5.01	-5.01	-4.52	-5.83
4	P47895	ALDH1A3	Aldehyde dehydrogenase family 1 member A3	56	40	13	12	31	267	-3.85	-3.90	-1.51	-3.17	-3.70	-3.15	-1.56
5	Q04828	AKR1C1	Aldo-keto reductase family 1 member C1	37	25	7	7	28	203	-2.94	-4.08	-1.27	-2.53	-4.27	-3.20	-4.02
6	P12814	ACTN1	Alpha-actinin-1	103	190	36	22	49	2301	-2.30	-2.75	-1.65	-1.69	-2.28	-2.20	-1.80
7	P07738	BPGM	Bisphosphoglycerate mutase	30	2	1	1	4	59	-4.42	-3.70	-1.37	-3.29	-4.74	-3.76	-3.57
8	P00918	CA2	Carbonic anhydrase 2	29	8	3	3	17	26	-1.59	-3.03	-2.53	-4.88	-4.59	-1.74	-2.15
9	P07858	CTSB	Cathepsin B	38	46	5	5	22	849	-2.62	-2.37	-2.60	-2.72	-4.16	-2.09	-3.07
10	P12109	COL6A1	Collagen alpha-1(VI) chain	108	47	14	14	19	687	-4.71	-5.28	-4.34	-4.75	-5.33	-4.26	-5.07
11	P12110	COL6A2	Collagen alpha-2(VI) chain	109	22	8	8	8	286	-4.42	-4.59	-4.34	-3.53	-4.83	-4.26	-4.15
12	P12111	COL6A3	Collagen alpha-3(VI) chain	343	117	31	31	11	1373	-5.08	-5.28	-5.01	-5.01	-5.01	-4.52	-5.07
13	Q96CG8	CTHRC1	Collagen triple helix repeat-containing protein 1	26	4	1	1	3	23	-5.71	-5.66	-5.60	-5.01	-5.59	-4.52	-5.83
14	Q9ULV4	CORO1C	Coronin-1C	53	33	8	8	18	467	-2.03	-2.37	-1.86	-2.38	-2.17	-2.39	-1.46
15	Q14195	DPYSL3	Dihydropyrimidinase-related protein 3	62	22	7	5	13	106	-2.96	-2.57	-2.43	-2.34	-2.08	-2.99	-1.89
16	Q9UDY4	DNAJB4	DnaJ homolog subfamily B member 4	38	2	1	1	3	29	-5.27	-5.66	-4.22	-2.84	-5.59	-4.52	-4.15
17	Q16658	FSCN1	Fascin	54	46	13	13	29	476	-3.85	-4.59	-3.22	-3.69	-4.66	-3.20	-1.32
18	P05413	FABP3	Fatty acid-binding protein, heart	15	21	6	6	52	272	-4.08	-3.49	-4.11	-2.91	-4.01	-3.45	-4.02
19	P02751	FN1	Fibronectin	262	169	43	43	24	2284	-4.31	-4.46	-4.34	-3.94	-4.27	-2.23	-2.87
20	O95479	H6PD	GDH/6PGL endoplasmic bifunctional protein	89	6	4	4	7	37	-3.37	-4.18	-4.40	-3.43	-5.59	-4.26	-3.32
21	Q9Y2G5	POFUT2	GDP-fucose protein O-fucosyltransferase 2	50	3	1	1	3	60	-1.63	-2.56	-2.51	-1.77	-4.45	-2.89	-2.48
22	P06396	GSN	Gelsolin	86	42	12	12	19	261	-2.75	-2.40	-2.94	-1.91	-4.39	-2.09	-2.55
23	P07093	SERPINE2	Glia-derived nexin	44	8	2	2	6	157	-5.85	-5.66	-5.60	-3.58	-5.59	-4.52	-5.07
24	O94925	GLS	Glutaminase kidney isoform, mitochondrial	73	4	2	2	4	107	-3.68	-2.94	-1.92	-2.09	-4.16	-2.72	-1.25
25	P11216	PYGB	Glycogen phosphorylase, brain form	97	55	23	20	32	713	-2.43	-2.19	-2.28	-2.78	-2.27	-2.01	-2.23
26	P02100	HBE1	Hemoglobin subunit epsilon	16	24	2	2	13	279	-5.08	-5.66	-5.34	-4.43	-4.59	-3.26	-4.35
27	P05556	ITGB1	Integrin beta-1	88	20	8	8	13	182	-2.37	-2.18	-2.66	-1.71	-2.66	-1.89	-1.24
28	Q13418	ILK	Integrin-linked protein kinase	51	7	3	3	8	61	-2.25	-2.31	-1.59	-1.88	-2.13	-2.45	-2.30

Table 3-5, a: Significantly down-regulated proteins in cancer and premalignant cell lines compared to primary human mammary epithelial cells.

No	Accession	Gene ID	Description	MW [kDa]	PSMs	Peptides	Unique Peptides	Coverage	Score Mascot	MCF-7 (LOG2)	ZR-75 (LOG2)	MDA-MB-468 (LOG2)	MDA-MB-231 (LOG2)	MDA-MB-453 (LOG2)	MCF-10A (LOG2)	HB2 (LOG2)
29	P19012	KRT15	Keratin, type I cytoskeletal 15	49	166	19	11	38	2143	-4.92	-4.52	-3.88	-4.53	-5.01	-3.94	-4.30
30	P08779	KRT16	Keratin, type I cytoskeletal 16	51	200	28	15	60	2779	-3.92	-3.81	-2.75	-3.29	-3.83	-2.32	-3.18
31	Q04695	KRT17	Keratin, type I cytoskeletal 17	48	280	33	18	70	3566	-5.37	-5.28	-2.03	-4.75	-5.45	-3.26	-3.66
32	P04259	KRT6B	Keratin, type II cytoskeletal 6B	60	235	31	3	50	2798	-5.17	-5.66	-4.22	-5.01	-5.59	-4.52	-4.98
33	Q08431	MFGE8	Lactadherin	43	3	2	2	7	34	-3.56	-4.08	-3.11	-2.25	-4.05	-2.01	-2.36
34	Q14764	MVP	Major vault protein	99	42	15	15	22	467	-3.20	-3.01	-2.35	-1.61	-3.87	-2.76	-3.05
35	Q3SY69	ALDH1L2	Mitochondrial 10-formyltetrahydrofolate dehydrogenase	102	5	3	3	5	45	-5.08	-5.40	-2.50	-5.01	-5.21	-4.52	-3.73
36	P35579	MYH9	Myosin-9	226	330	79	65	43	4463	-3.20	-2.49	-2.31	-2.61	-1.90	-3.09	-1.86
37	P08473	MME	Nephrilysin	85	9	6	6	9	222	-5.47	-5.66	-2.14	-5.01	-4.52	-4.04	-3.57
38	P80188	LCN2	Neutrophil gelatinase-associated lipocalin	23	3	2	2	14	37	-4.85	-5.66	-2.15	-3.21	-5.59	-4.52	-4.40
39	P40261	NNMT	Nicotinamide N-methyltransferase	30	6	2	2	7	81	-5.85	-5.66	-5.60	-3.09	-5.33	-3.52	-1.99
40	Q96AY3	FKBP10	Peptidyl-prolyl cis-trans isomerase FKBP10	64	32	8	8	16	384	-3.96	-3.34	-3.71	-3.13	-3.59	-2.85	-1.95
41	O95302	FKBP9	Peptidyl-prolyl cis-trans isomerase FKBP9	63	8	4	4	7	67	-2.31	-2.54	-1.85	-2.15	-2.72	-2.20	-1.94
42	Q9H307	PNN	Pinin	82	4	1	1	2	25	-2.83	-3.81	-2.11	-1.77	-3.70	-1.58	-2.98
43	O00469	PLOD2	Procollagen-lysine,2-oxoglutarate 5-dioxygenase 2	85	34	13	13	19	314	-2.87	-4.40	-3.71	-3.81	-4.01	-2.45	-1.39
44	Q07954	LRP1	Prolow-density lipoprotein receptor-related protein 1	504	5	4	4	1	22	-3.81	-3.98	-4.11	-4.75	-3.74	-4.52	-3.43
45	Q32P28	P3H1	Prolyl 3-hydroxylase 1	83	6	4	4	6	74	-3.89	-4.66	-3.83	-1.81	-3.33	-3.94	-3.48
46	Q8IVL6	P3H3	Prolyl 3-hydroxylase 3	82	4	2	2	5	106	-4.47	-2.04	-3.71	-5.01	-3.11	-4.52	-3.37
47	Q6P1N9	TATDN1	Putative deoxyribonuclease TATDN1	34	2	1	1	3	36	-1.42	-2.10	-1.38	-4.75	-2.09	-2.23	-2.09
48	Q96D15	RCN3	Reticulocalbin-3	37	7	4	4	13	48	-4.65	-4.74	-4.17	-3.75	-4.59	-4.15	-4.57
49	P09486	SPARC	SPARC	35	2	1	1	4	40	-2.71	-2.85	-2.37	-2.88	-2.49	-3.15	-2.28
50	P04179	SOD2	Superoxide dismutase [Mn], mitochondrial	25	35	9	9	51	285	-3.75	-3.77	-3.64	-1.84	-4.39	-2.80	-2.85
51	Q9Y490	TLN1	Talin-1	270	168	61	61	34	2138	-2.40	-2.48	-1.59	-1.96	-1.78	-2.26	-1.97
52	P07996	THBS1	Thrombospondin-1	129	33	11	11	11	288	-3.24	-3.66	-2.81	-2.91	-1.97	-2.94	-2.33
53	P04216	THY1	Thy-1 membrane glycoprotein	18	16	3	3	24	215	-3.89	-3.10	-3.83	-3.81	-3.70	-3.60	-1.62
54	Q9UKE5	TNIK	TRAF2 and NCK-interacting protein kinase	155	5	1	1	1	62	-3.24	-3.98	-2.04	-1.52	-5.45	-4.52	-1.79
55	Q01995	TAGLN	Transgelin	23	27	8	8	45	243	-4.37	-4.40	-3.83	-4.34	-4.27	-4.04	-3.83
56	P09936	UCHL1	Ubiquitin carboxyl-terminal hydrolase isozyme L1	25	28	8	8	48	221	-5.00	-4.98	-5.46	-4.88	-5.59	-3.76	-2.39

Table 3-5, b: Significantly down-regulated proteins in cancer and premalignant cell lines compared to primary human mammary epithelial cells.

3.3.4 Hierarchical Clustering of breast cell lines

Because of the heterogeneity of breast cancer, it is challenging to differentiate between each subtype based on a single gene or protein. Therefore, a group of markers are required that can serve as a signature for diagnosing different types of breast cancer. Gene expression profiling has played an important role in understanding the heterogeneity of breast cancer at a molecular level and refining the taxonomy based on the presence of productive markers like ER, PR and HER2 to more sophisticated grouping comprising luminal A, luminal B, basal-like, claudin-Low and HER2-positive phenotypes (Holliday and Speirs, 2011). Hierarchical clustering of breast cell lines proteome (SPS data), was performed by R Studio (method section for full details of R script), to compare the proteome of each cell lines. Luminal and basal cancer cell lines clustered together as expected, however the normal luminal (HB-2) and basal (MCF-10A) cell lines exhibited less similarity to their prospective lineages.

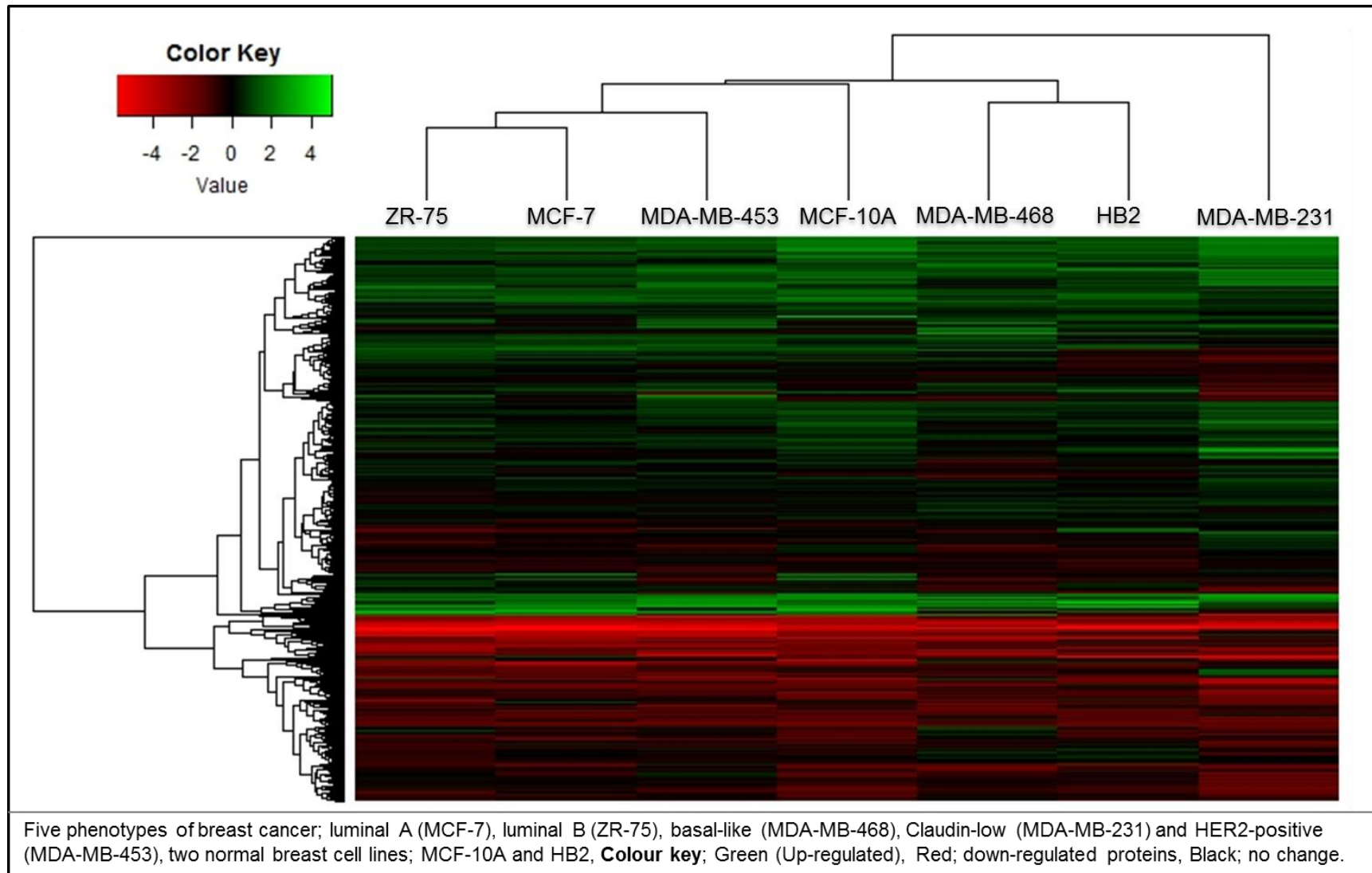


Figure 3-8: Hierarchical clustering of breast cell lines (cancer and normal-like) proteome compared to primary human mammary epithelial cells.

3.3.5 Breast cancer stratification using cancer cell line proteome profiles

Statistical analysis was performed through R Studio software package; Bioconductor 3.4-Limma (section 2.12.3 above) to determine those proteins that were most significantly changed in specific cell lines (and therefore potentially, phenotype-specific) was used to determine unique protein signatures. The significant difference (both up- and down-regulated) of the iTRAQ ratio for one cell line relative to the ratios of the other cell lines ($p < 0.05$) provides proteins that are unique protein classifiers for each phenotype (Figure 3-9). Interestingly, there number of proteins that were significantly different relative to HMEC increased relative to the stage of normal mammary development, with the smallest group of proteins defining differentiated Luminal A (MCF-7, (34 proteins) and the largest signature (495 proteins) associated with the stem cell-derived cancer cell line, MDA-MB-231, whose lineage is furthest removed from the mature differentiated healthy primary HMECs used as a control in this quantitative study.

The expression of 57 proteins was significantly ($p < 0.05$) changed in normal-like cell lines (MCF-10A and HB2) compared to Breast cancer cell lines (Detailed annotations for all signature proteins are presented in supplementary data, “Breast cancer specific-Go-Analysis” as CD format). Among these, 17 proteins were down regulated and 40 proteins were up regulated in HB2 and MCF-10A breast cell lines. Cellular localisation (p value 0.004) and viral transcription process ($p < 0.0001$) was up-regulated in normal-like cell lines compared to cancer cell lines. Four microRNAs in the Cancer pathway (p value

0.03); SERPINB5, CDKN2A, EGFR and FSCN1, were also up-regulated in HB2 and MCF-10A.

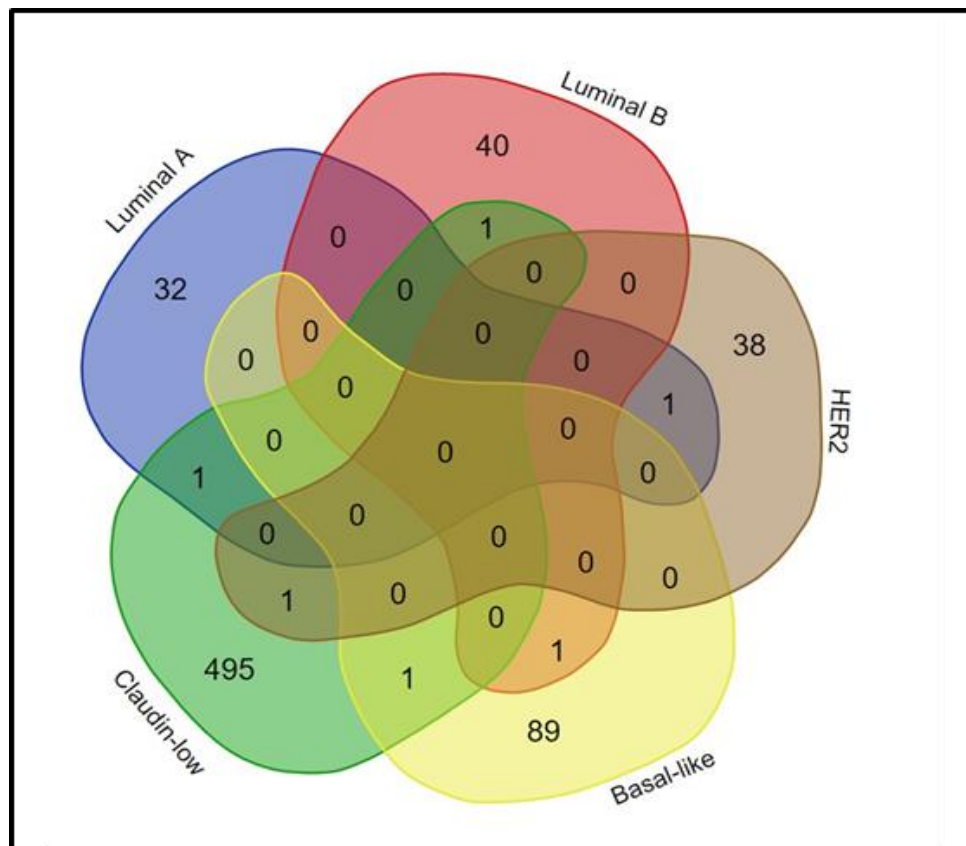


Figure 3-9: Venn diagram comparison of signature proteomes for breast cancer cell lines.

3.3.5.1 Luminal vs Basal

A group of 275 proteins ($p < 0.05$) was significantly altered between luminal (MCF-7, ZR-75 and MDA-MB-453) vs basal (MDA-MB-468 and MDA-MB-231) (Figure 3-10, Supplementary Table 3.1). One of the most significantly increased proteins in luminal cancer cell lines was anterior gradient protein 2 homolog (AGR2), a proto-oncogene and protein disulphide isomerase, which plays a role post-transcriptional synthesis and secretion of mucins (Park et al., 2009), not only in breast cells (ER positive) (Fritzsche et al., 2006), but many normal tissues (most strongly in the digestive tract, respiratory tissues, and male

and female reproductive organs) (Park et al., 2009). Using an iTRAQ approach, Chung et al, demonstrated elevated levels of AGR2 in lung carcinoma compared to adjacent normal tissues, which was verified by immunohistochemistry/tissue microarray of 268 cases, and proposed as a possible biomarker (CITY, 2012).

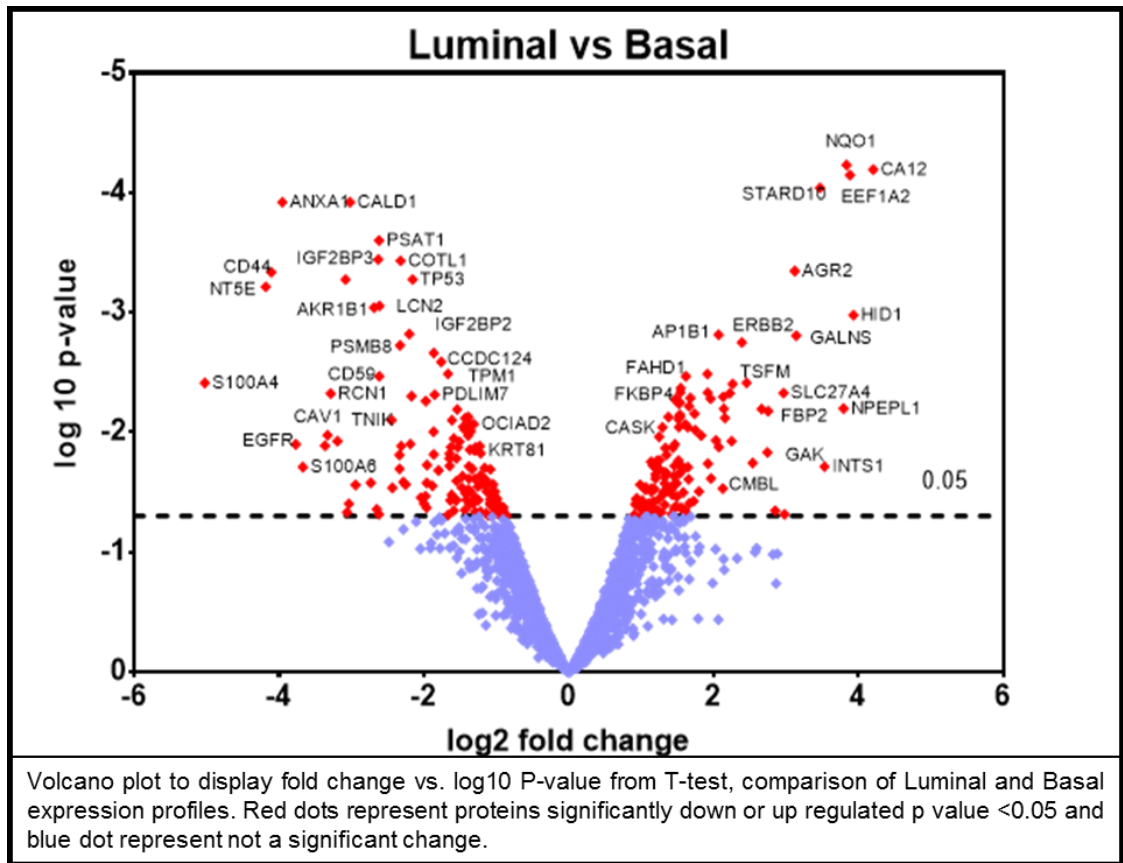


Figure 3-10: Volcano plot analysis of differentially expressed proteins in Luminal vs Basal breast cancer cell lines.

3.3.5.2 Luminal A (MCF-7)

For luminal A phenotype (MCF-7), a group of 34 proteins ($p < 0.05$) were identified (Figure 3-11, Table 3-6), STRING analysis indicated that members were associated with fructose metabolism ($1.640e-2$, 3 components, all higher than in other cell lines relative to HMEC) and AMPK signalling pathway ($4.80e-2$, 3 components all increased). The expression of kynureninase (KYNU) was significantly ($p < 0.001$) higher in MCF-7 compared to all other cancer cell lines. Detailed annotations for all signature proteins are presented in supplementary data, “Breast cancer specific-Go-Analysis” as CD format.

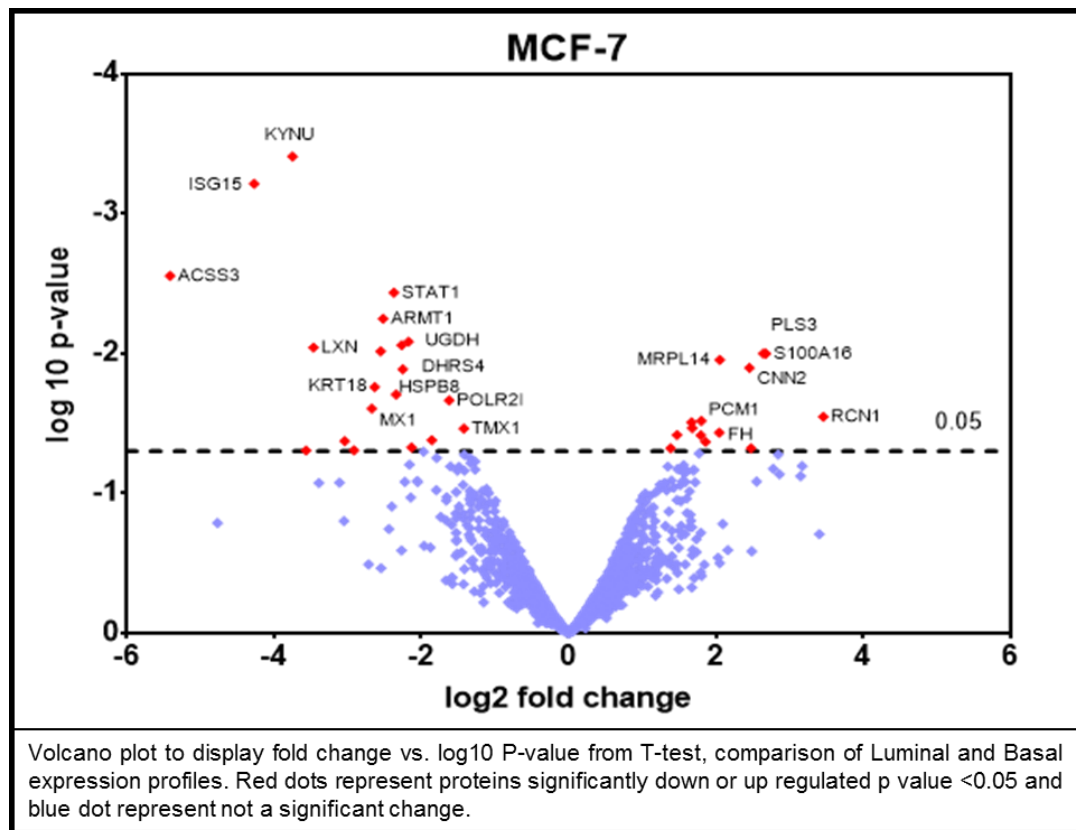


Figure 3-11: Volcano plot analysis of differentially expressed proteins in MCF-7 (Luminal A).

Sr. no	Accession	Gene ID	Description	PSMs	Peptides	Coverage	Unique Peptides	Score Mascot	MCF-7 (logFC)	MCF-7 (P.Value)
1	Q16719	KYNU	Kynureninase	4	2	6	2	30	-3.75	0.0004
2	P05161	ISG15	Ubiquitin-like protein ISG15	25	5	27	5	235	-4.27	0.0006
3	Q9H6R3	ACSS3	Acyl-CoA synthetase short-chain family member 3, mitochondrial	2	2	5	2	39	-5.41	0.0028
4	P42224	STAT1	Signal transducer and activator of transcription 1-alpha	49	14	23	14	559	-2.37	0.0037
5	Q9H993	ARMT1	Protein-glutamate O-methyltransferase	3	3	5	3	51	-2.52	0.0056
6	P09972	ALDOC	Fructose-bisphosphate aldolase C	28	6	25	3	190	-2.17	0.0082
7	Q9BTZ2	DHRS4	Dehydrogenase	8	5	28	5	137	-2.27	0.0087
8	Q9BS40	LXN	Latexin	4	1	8	1	208	-3.46	0.0090
9	O60701	UGDH	UDP-glucose 6-dehydrogenase	37	13	35	13	575	-2.55	0.0096
10	P13797	PLS3	Plastin-3	11	6	11	3	70	2.65	0.0100
11	Q96FQ6	S100A16	Protein S100-A16	18	4	45	4	391	2.68	0.0100
12	Q6P1L8	MRPL14	39S ribosomal protein L14, mitochondrial	4	1	6	1	59	2.05	0.0111
13	Q99439	CNN2	Calponin-2	18	6	23	5	143	2.46	0.0126
14	P23443	RPS6KB1	Ribosomal protein S6 kinase beta-1	3	2	5	2	102	-2.25	0.0129
15	Q9UJY1	HSPB8	Heat shock protein beta-8	5	2	9	2	43	-2.63	0.0173
16	P05783	KRT18	Keratin, type I cytoskeletal 18	617	33	78	31	11934	-2.34	0.0196
17	P36954	POLR2I	DNA-directed RNA polymerase II subunit RPB9	4	2	14	2	33	-1.62	0.0216
18	P20591	MX1	Interferon-induced GTP-binding protein Mx1	12	4	7	4	122	-2.67	0.0247
19	Q15293	RCN1	Reticulocalbin-1	27	6	18	6	211	3.46	0.0282
20	Q6IBS0	TWF2	Twinfilin-2	12	6	21	3	122	1.80	0.0303
21	Q9C0C2	TNKS1BP1	182 kDa tankyrase-1-binding protein	12	6	5	6	124	1.67	0.0310
22	P30536	TSPO	Translocator protein	7	2	18	2	88	1.68	0.0340
23	Q9H3N1	TMX1	Thioredoxin-related transmembrane protein 1	7	3	13	3	83	-1.42	0.0343
24	Q6UB35	MTHFD1L	Monofunctional C1-tetrahydrofolate synthase, mitochondrial	2	1	2	1	23	2.05	0.0367
25	P07954	FH	Fumarate hydratase, mitochondrial	36	11	28	11	770	1.48	0.0382
26	Q15154	PCM1	Pericentriolar material 1 protein	3	2	1	2	31	1.80	0.0383
27	P06280	GLA	Alpha-galactosidase A	2	1	3	1	103	-1.86	0.0416
28	P21266	GSTM3	Glutathione S-transferase Mu 3	10	4	17	4	157	-3.04	0.0421
29	P17655	CAPN2	Calpain-2 catalytic subunit	36	8	15	8	733	1.86	0.0427
30	P05787	KRT8	Keratin, type II cytoskeletal 8	672	46	83	38	7529	-2.13	0.0468
31	P51159	RAB27A	Ras-related protein Rab-27A	3	1	4	1	27	2.48	0.0475
32	Q9NP61	ARFGAP3	ADP-ribosylation factor GTPase-activating protein 3	3	2	4	1	49	1.39	0.0475
33	O00757	FBP2	Fructose-1,6-bisphosphatase isozyme 2	4	2	6	1	100	-2.91	0.0491
34	P09467	FBP1	Fructose-1,6-bisphosphatase 1	12	6	23	5	178	-3.57	0.0493

The table represents Uniprot Accession number, Gene ID, protein name, number of PSMs, total Peptides, unique peptides, protein coverage and Mascot score. LogFC: log fold change of MCF-7 vs all cancer cell lines.

Table 3-6: Signature proteins for luminal A type breast cancer.

3.3.5.3 Luminal B (ZR-75)

A group of 42 proteins ($p < 0.05$) were observed to be uniquely increased or decreased in the luminal B cell line ZR-75 (Figure 3-12, Table 3-7). Glutathione metabolism pathway (3.71×10^{-2} , 3 components), was significantly altered, using Gene Ontology analysis. Detailed GO-annotations for all signature proteins are presented in supplementary data, “Breast cancer specific-Go-Analysis” as CD format. Programmed cell death protein 4 (PDCD4), a tumour suppressor gene and plays a role in apoptosis (Yang et al., 2006), was significantly decreased (p value 0.005) in luminal type breast cancer cell line ZR-75. The expression of proteasome assembly chaperone 2 (PSMG2), a negative regulation of apoptotic process (Uniprot.org, 2015), was increased (p value 0.01) in ZR-75 compared to all other breast cancer cell lines.

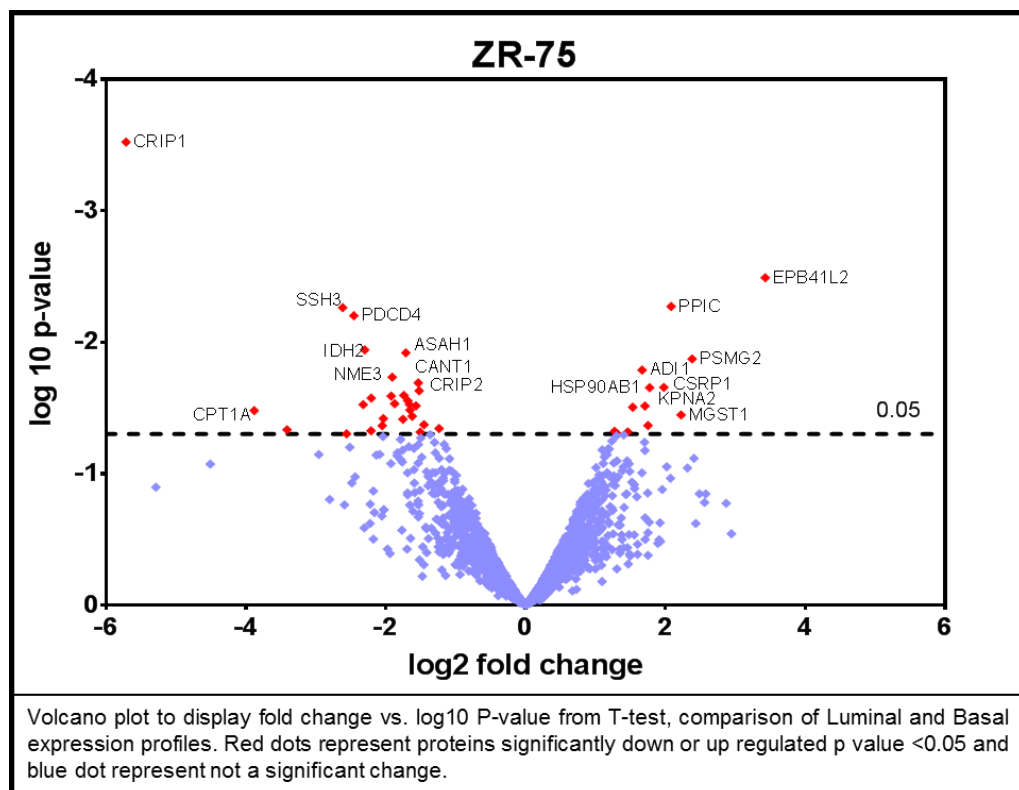


Figure 3-12: Volcano plot analysis of differentially expressed proteins in ZR-75 (Luminal B).

Table 3-7: Signature proteins for luminal B type breast cancer.

The table represents Uniprot Accession number, Gene ID, protein name, number of PSMs, total Peptides, unique peptides, protein coverage and Mascot score. LogFC: log fold change of ZR-75 vs all cancer cell lines

Sr. no	Accession	Gene ID	Description	PSMs	Peptides	Coverage	Unique Peptides	Score Mascot	ZR-75 (logFC)	ZR-75 (P.Value)
1	P50238	CRIP1	Cysteine-rich protein 1	2	6	21	2	43	-5.72	0.0003
2	O43491	EPB41L2	Band 4.1-like protein 2	1	2	1	1	27	3.43	0.0032
3	P45877	PPIC	Peptidyl-prolyl cis-trans isomerase C	3	7	13	2	99	2.08	0.0053
4	Q8TE77	SSH3	Protein phosphatase Slingshot homolog 3	6	13	9	6	156	-2.62	0.0055
5	Q53EL6	PDCD4	Programmed cell death protein 4	4	8	10	4	95	-2.46	0.0063
6	Q5H9R7	PPP6R3	Serine/threonine-protein phosphatase 6 regulatory subunit 3	2	3	3	2	61	-2.30	0.0114
7	Q13510	ASAH1	Acid ceramidase	7	15	13	7	68	-1.71	0.0120
8	Q969U7	PSMG2	Proteasome assembly chaperone 2	1	8	5	1	131	2.38	0.0134
9	Q9BV57	ADI1	1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase	3	7	18	3	70	1.67	0.0162
10	Q13232	NME3	Nucleoside diphosphate kinase 3	2	2	12	2	27	-1.91	0.0184
11	Q8WVQ1	CANT1	Soluble calcium-activated nucleotidase 1	2	2	5	2	46	-1.53	0.0203
12	P21291	CSRP1	Cysteine and glycine-rich protein 1	3	11	28	3	286	1.98	0.0220
13	P52292	KPNA2	Importin subunit alpha-1	5	24	12	5	345	1.77	0.0221
14	O94905	ERLIN2	Erlin-2	2	9	6	2	45	-1.52	0.0234
15	P50897	PPT1	Palmitoyl-protein thioesterase 1	6	18	31	6	389	-1.74	0.0253
16	P32322	PYCR1	Pyrroline-5-carboxylate reductase 1, mitochondrial	7	21	31	6	372	-1.92	0.0256
17	Q9NSE4	IARS2	Isoleucine--tRNA ligase, mitochondrial	9	15	10	9	69	-2.21	0.0265
18	Q9NUI1	DECR2	Peroxisomal 2,4-dienoyl-CoA reductase	3	6	12	3	102	-1.68	0.0278
19	O75223	GGCT	Gamma-glutamylcyclotransferase	4	10	24	4	135	-1.87	0.0293
20	Q9Y277	VDAC3	Voltage-dependent anion-selective channel protein 3	9	45	34	8	406	-1.66	0.0297
21	Q9NRF8	CTPS2	CTP synthase 2	2	3	4	1	76	-2.32	0.0298
22	P52943	CRIP2	Cysteine-rich protein 2	4	10	36	4	156	-1.57	0.0305
23	P01130	LDLR	Low-density lipoprotein receptor	1	2	2	1	28	1.71	0.0305
24	P08238	HSP90AB1	Heat shock protein HSP 90-beta	39	344	59	17	5782	1.53	0.0311
25	Q12765	SCRN1	Secernin-1	3	7	10	3	67	-1.65	0.0327
26	P50416	CPT1A	Carnitine O-palmitoyltransferase 1, liver isoform	4	12	5	4	185	-3.88	0.0330
27	P10620	MGST1	Microsomal glutathione S-transferase 1	2	4	17	2	38	2.23	0.0357
28	Q8NFU3	TSTD1	Thiosulfate sulfurtransferase	2	8	14	2	50	-1.62	0.0365
29	Q8IVL6	LEPREL2	Prolyl 3-hydroxylase 3	2	4	5	2	106	-2.04	0.0380
30	Q9BR76	CORO1B	Coronin-1B	10	25	16	10	170	-1.75	0.0385
31	P11117	ACP2	Lysosomal acid phosphatase	2	5	5	2	30	-1.45	0.0423
32	Q14155	ARHGEF7	Rho guanine nucleotide exchange factor 7	2	2	3	2	33	-1.46	0.0424
33	O95801	TTC4	Tetratricopeptide repeat protein 4	2	3	7	2	73	1.75	0.0429
34	O94919	ENDOD1	Endonuclease domain-containing 1 protein	2	5	4	2	45	-2.05	0.0430
35	O00170	AIP	AH receptor-interacting protein	3	4	10	3	69	-1.24	0.0452

Sr. no	Accession	Gene ID	Description	PSMs	Peptides	Coverage	Unique Peptides	Score Mascot	ZR-75 (logFC)	ZR-75 (P.Value)
37	Q9BUP0	EFHD1	EF-hand domain-containing protein D1	3	4	13	2	33	-2.21	0.0470
38	P02765	AHSG	Alpha-2-HS-glycoprotein	2	5	5	2	84	1.27	0.0476
39	O43447	PPIH	Peptidyl-prolyl cis-trans isomerase H	2	7	8	1	80	1.46	0.0480
40	Q9BXW7	CECR5	Cat eye syndrome critical region protein 5	3	5	12	3	78	-1.51	0.0481
41	P48735	IDH2	Isocitrate dehydrogenase [NADP], mitochondrial	16	64	36	14	470	-2.56	0.0495
42	P26641	EEF1G	Elongation factor 1-gamma	12	96	27	12	911	1.31	0.0497

3.3.5.4 HER2 positive (MDA-MB-453)

A HER2+ve signature was defined, comprising 40 proteins ($p < 0.05$, Figure 3-13, Table 3-8), of which 9 were associated with metabolic pathways, including fatty acid metabolism and 10 proteins were regulator of oxidation and reduction process. Detailed GO-annotations for all signature proteins are presented in supplementary data, “Breast cancer specific-Go-Analysis” as CD format. The expression of podocalyxin (PODXL), a regulator of both adhesion and cell morphology as well as cancer progression (Somasiri et al., 2004), was 4.7 fold higher (p value 0.01) in HER2 positive breast cancer cell lines MDA-MB-453. The expression of cathepsin B (CTSB), which plays a role in tumour invasion and metastasis (Aggarwal and Sloane, 2014), was also higher (p value 0.02) in MDA-MB-453 compared to all breast cancer cell lines.

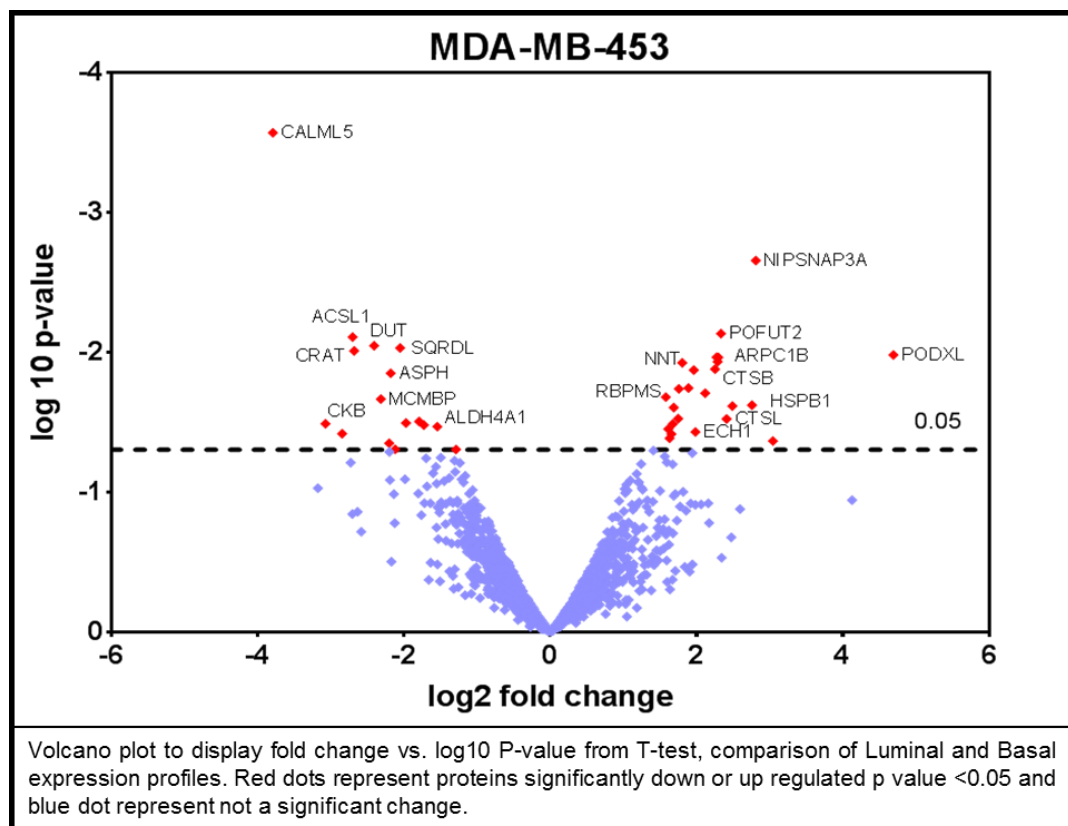


Figure 3-13: Volcano plot analysis of differentially expressed proteins in MDA-MB-453 (HER2 positive).

Sr. no	Accession	Gene ID	Description	PSMs	Peptides	Coverage	Unique Peptides	Score Mascot	MDA-MB-453 (logFC)	MDA-MB-453 (P.Value)
1	Q9NZT1	CALML5	Calmodulin-like protein 5	3	2	16	2	31	-3.79	0.0003
2	Q9UFN0	NIPSNAP3A	Protein NipSnap homolog 3A	2	1	6	1	56	2.81	0.0022
3	Q9Y2G5	POFUT2	GDP-fucose protein O-fucosyltransferase 2	3	1	3	1	60	2.34	0.0074
4	P33121	ACSL1	Long-chain-fatty-acid--CoA ligase 1	19	5	8	5	270	-2.70	0.0078
5	P33316	DUT	Deoxyuridine 5'-triphosphate nucleotidohydrolase, mitochondrial	39	7	33	7	466	-2.41	0.0090
6	Q9Y6N5	SQRDL	Sulfide:quinone oxidoreductase, mitochondrial	37	15	45	15	719	-2.05	0.0093
7	P43155	CRAT	Carnitine O-acetyltransferase	7	5	8	5	36	-2.68	0.0098
8	O00592	PODXL	Podocalyxin	5	1	1	1	34	4.69	0.0105
9	Q13641	TPBG	Trophoblast glycoprotein	11	3	10	3	143	2.27	0.0108
10	Q13509	TUBB3	Tubulin beta-3 chain	296	18	49	4	4106	2.30	0.0109
11	P42765	ACAA2	3-ketoacyl-CoA thiolase, mitochondrial	7	4	12	4	145	2.29	0.0117
12	Q13423	NNT	NAD(P) transhydrogenase, mitochondrial	13	9	10	9	100	1.81	0.0119
13	Q08380	LGALS3BP	Galectin-3-binding protein	13	7	16	7	189	2.25	0.0132
14	O15143	ARPC1B	Actin-related protein 2	10	4	12	4	74	1.96	0.0134
15	Q12797	ASPH	Aspartyl	23	9	12	9	295	-2.18	0.0141
16	P06396	GSN	Gelsolin	42	12	19	12	261	1.89	0.0180
17	P51812	RPS6KA3	Ribosomal protein S6 kinase alpha-3	4	1	1	1	45	1.76	0.0183
18	P50995	ANXA11	Annexin A11	24	8	17	8	238	2.12	0.0196
19	P07858	CTSB	Cathepsin B	46	5	22	5	849	1.58	0.0209
20	Q9BTE3	MCMBP	Mini-chromosome maintenance complex-binding protein	6	2	3	2	27	-2.32	0.0216
21	Q9UKE5	TNIK	TRAF2 and NCK-interacting protein kinase	5	1	1	1	62	2.76	0.0239
22	P04792	HSPB1	Heat shock protein beta-1	135	14	83	14	2315	2.49	0.0243
23	Q9Y696	CLIC4	Chloride intracellular channel protein 4	15	5	25	5	90	1.69	0.0249
24	Q95479	H6PD	GDH/6PGL endoplasmic bifunctional protein	6	4	7	4	37	1.75	0.0299
25	P07711	CTSL	Cathepsin L1	3	1	4	1	85	2.41	0.0300
26	Q99797	MIPEP	Mitochondrial intermediate peptidase	2	2	2	2	56	-1.79	0.0313
27	P27105	STOM	Erythrocyte band 7 integral membrane protein	18	7	29	7	132	-1.97	0.0320
28	P12277	CKB	Creatine kinase B-type	24	9	32	9	237	-3.07	0.0325
29	O00151	PDLIM1	PDZ and LIM domain protein 1	15	6	23	6	224	1.68	0.0326
30	Q14166	TLL12	Tubulin--tyrosine ligase-like protein 12	19	9	20	9	205	-1.73	0.0332
31	P33992	MCM5	DNA replication licensing factor MCM5	11	6	10	6	134	-1.54	0.0341
32	P49748	ACADVL	Very long-chain specific acyl-CoA dehydrogenase, mitochondrial	25	11	19	11	284	1.62	0.0354
33	Q13011	ECH1	Delta(3,5)-Delta(2,4)-dienoyl-CoA isomerase, mitochondrial	40	9	34	9	448	1.98	0.0371
34	Q8TD19	NEK9	Serine/threonine-protein kinase Nek9	2	1	1	1	47	-2.84	0.0383
35	Q93062	RBPMS	RNA-binding protein with multiple splicing	2	1	7	1	36	1.66	0.0385
36	P07951	TPM2	Tropomyosin beta chain	50	11	23	1	725	1.63	0.0412
37	P21266	GSTM3	Glutathione S-transferase Mu 3	10	4	17	4	157	3.05	0.0432
38	P30038	ALDH4A1	Delta-1-pyrroline-5-carboxylate dehydrogenase, mitochondrial	14	6	12	6	140	-2.20	0.0447
39	Q00796	SORD	Sorbitol dehydrogenase	20	7	28	7	455	-2.12	0.0494
40	P12268	IMPDH2	Inosine-5'-monophosphate dehydrogenase 2	40	14	37	12	485	-1.29	0.0496

The table represents Uniprot Accession number, Gene ID, protein name, number of PSMs, total Peptides, unique peptides, protein coverage and Mascot score. LogFC: log fold change of MDA-MB-453 vs all cancer cell lines.

Table 3-8: Signature proteins for HER2 positive breast cancer.

3.3.5.5 Basal-like specific (MDA-MB-468)

A group of 91 proteins ($p < 0.05$) were observed to be uniquely expressed in the basal-like cell line MDA-MB-468 (Figure 3-14, Table 3-9), including organonitrogen compound biosynthetic process (4.29×10^{-5} , 13 components) and cellular amino acid metabolic process (0.0004 , 11 components). The metabolic pathway (7.99×10^{-6} , 21 components) and synthesis of amino acids (0.002 , 5 components) were two altered KEGG pathways. Detailed GO annotations for all signature proteins are presented in supplementary data, Breast cancer specific-Go-Analysis as CD format. The expression of Epidermal growth factor receptor (EGFR), a key regulator of cell growth, migration and proliferation, also plays role in DNA repair and replication by EGFR signalling pathway (Nicholson et al., 2001, Engelman and Cantley, 2008), was down-regulated in basal-like breast cancer cell line MDA-MB-468.

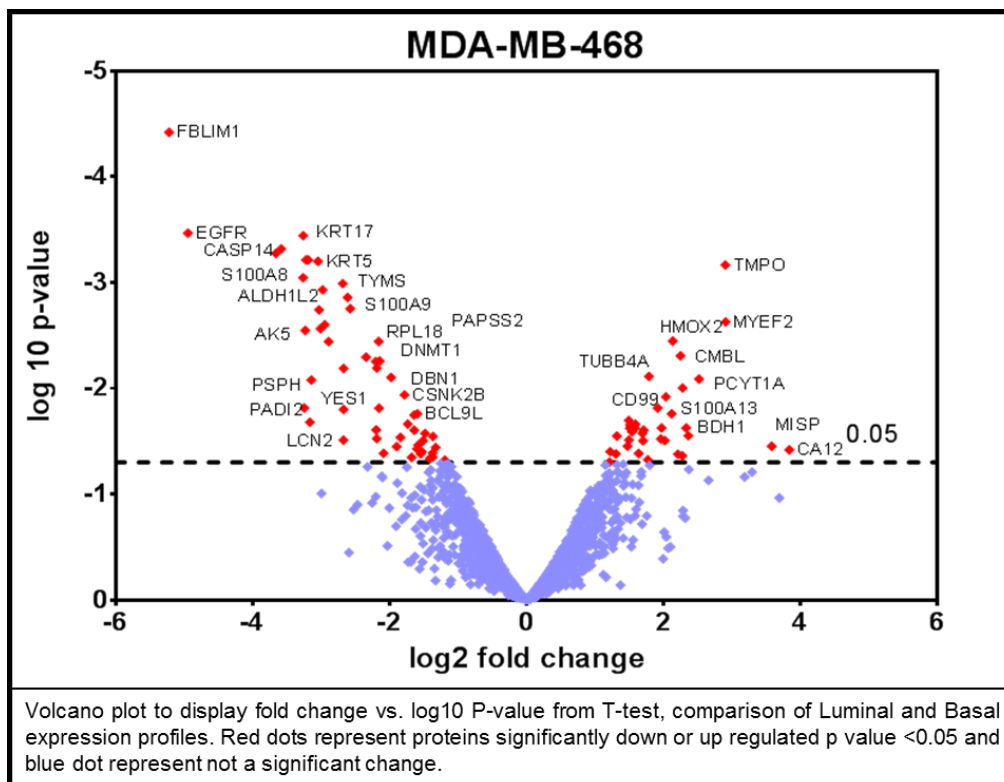


Figure 3-14: Volcano plot analysis of differentially expressed proteins in MDA-MB-468 (Basal-Like).

Table 3-9: Signature proteins for Basal-like breast cancer.

The table represents Uniprot Accession number, Gene ID, protein name, number of PSMs, total Peptides, unique peptides, protein coverage and Mascot score. LogFC: log fold change of MDA-MB-468 vs all cancer cell lines.

Sr. no	Accession	Gene ID	Description	PSMs	Peptides	Coverage	Unique Peptides	Score Mascot	MDA-MB-468 (logFC)	MDA-MB-468 (P.Value)
1	Q8WUP2	FBLIM1	Filamin-binding LIM protein 1	4	1	3	1	88	-5.21	<0.0001
2	P00533	EGFR	Epidermal growth factor receptor	17	11	12	10	266	-4.94	0.0003
3	Q9NS86	LANCL2	LanC-like protein 2	3	2	5	2	33	-3.26	0.0004
4	P31944	CASP14	Caspase-14	5	4	18	4	24	-3.58	0.0005
5	Q9UHW9	SLC12A6	Solute carrier family 12 member 6	3	2	2	2	26	-3.66	0.0005
6	Q04695	KRT17	Keratin, type I cytoskeletal 17	280	33	70	18	3566	-3.19	0.0006
7	P05109	S100A8	Protein S100-A8	6	3	33	3	91	-3.22	0.0006
8	P13647	KRT5	Keratin, type II cytoskeletal 5	218	30	49	18	2285	-3.04	0.0006
9	P42166	TMPO	Lamina-associated polypeptide 2, isoform alpha	19	5	10	1	376	2.90	0.0007
10	Q14956	GNMB	Transmembrane glycoprotein NMB	2	1	2	1	40	-3.26	0.0009
11	Q3SY69	ALDH1L2	Mitochondrial 10-formyltetrahydrofolate dehydrogenase	5	3	5	3	45	-2.68	0.0010
12	Q6UWP2	DHRS11	Dehydrogenase	5	2	11	2	96	-2.97	0.0012
13	P06702	S100A9	Protein S100-A9	6	3	31	3	22	-2.61	0.0014
14	P02511	CRYAB	Alpha-crystallin B chain	7	3	16	3	71	-2.57	0.0018
15	P08473	MME	Neprilysin	9	6	9	6	222	-3.03	0.0018
16	Q9P2K5	MYEF2	Myelin expression factor 2	2	2	4	2	26	2.91	0.0024
17	Q9Y6K8	AK5	Adenylate kinase isoenzyme 5	8	2	4	1	59	-2.94	0.0025
18	P04818	TYMS	Thymidylate synthase	13	5	21	5	126	-3.01	0.0027
19	O95340	PAPSS2	Bifunctional 3'-phosphoadenosine 5'-phosphosulfate synthase 2	11	6	13	4	85	-3.23	0.0028
20	Q71DI3	HIST2H3A	Histone H3.2	36	4	40	1	258	2.13	0.0036
21	Q07020	RPL18	60S ribosomal protein L18	8	2	13	2	144	-2.16	0.0036
22	P54578	USP14	Ubiquitin carboxyl-terminal hydrolase 14	42	8	17	8	657	-2.89	0.0036
23	P30519	HMOX2	Heme oxygenase 2	3	2	12	2	119	2.25	0.0049
24	Q8IWA5	SLC44A2	Choline transporter-like protein 2	4	2	3	2	25	-2.34	0.0051
25	P47895	ALDH1A3	Aldehyde dehydrogenase family 1 member A3	40	13	31	12	267	-2.14	0.0055
26	Q9NQG5	RPRD1B	Regulation of nuclear pre-mRNA domain-containing protein 1B	5	1	2	1	31	-2.20	0.0056
27	P26358	DNMT1	DNA (cytosine-5)-methyltransferase 1	7	4	3	4	78	-2.18	0.0064
28	P07738	BPGM	Bisphosphoglycerate mutase	2	1	4	1	59	-2.67	0.0065
29	P04350	TUBB4A	Tubulin beta-4A chain	454	21	70	1	5780	1.79	0.0077
30	Q16643	DBN1	Drebrin	4	4	8	4	39	-1.97	0.0079
31	Q96DG6	CMBL	Carboxymethylenebutenolidase homolog	5	2	7	2	23	2.52	0.0082
32	P78330	PSPH	Phosphoserine phosphatase	18	4	20	4	335	-3.14	0.0083
33	Q9H1E3	NUCKS1	Nuclear ubiquitous casein and cyclin-dependent kinase substrate 1	3	1	4	1	35	2.28	0.0099
34	P67870	CSNK2B	Casein kinase II subunit beta	8	2	12	2	163	-1.78	0.0115
35	P49585	PCYT1A	Choline-phosphate cytidyltransferase A	5	2	7	2	130	2.04	0.0120
36	O00425	IGF2BP3	Insulin-like growth factor 2 mRNA-binding protein 3	13	5	10	3	154	-3.24	0.0153
37	P07686	HEXB	Beta-hexosaminidase subunit beta	5	4	10	4	61	1.91	0.0153
38	O43175	PHGDH	D-3-phosphoglycerate dehydrogenase	67	15	34	15	945	-2.15	0.0154

Sr. no	Accession	Gene ID	Description	PSMs	Peptides	Coverage	Unique Peptides	Score Mascot	MDA-MB-468 (logFC)	MDA-MB-468 (P.Value)
40	Q86UU0	BCL9L	B-cell CLL	2	1	1	1	33	-1.59	0.0174
41	Q86WC4	OSTM1	Osteopetrosis-associated transmembrane protein 1	2	1	4	1	50	2.12	0.0174
42	P15924	DSP	Desmoplakin	34	14	5	14	380	-1.64	0.0178
43	P68371	TUBB4B	Tubulin beta-4B chain	510	25	82	1	6906	1.49	0.0200
44	Q9Y2J8	PADI2	Protein-arginine deiminase type-2	13	4	6	4	144	-3.16	0.0208
45	P36915	GNL1	Guanine nucleotide-binding protein-like 1	5	2	4	2	162	-1.73	0.0217
46	P61970	NUTF2	Nuclear transport factor 2	5	3	45	3	74	1.59	0.0217
47	O60936	NOL3	Nucleolar protein 3	6	3	18	3	77	1.51	0.0222
48	P20962	PTMS	Parathymosin	5	2	22	2	101	2.33	0.0236
49	Q99584	S100A13	Protein S100-A13	21	5	49	5	142	1.97	0.0236
50	Q9NVS9	PNPO	Pyridoxine-5'-phosphate oxidase	3	2	9	2	57	1.58	0.0236
51	O15382	BCAT2	Branched-chain-amino-acid aminotransferase, mitochondrial	6	4	13	4	163	1.51	0.0237
52	Q6ZVX7	NCCRP1	F-box only protein 50	2	2	6	2	29	-2.19	0.0246
53	P04264	KRT1	Keratin, type II cytoskeletal 1	52	13	21	8	554	1.71	0.0248
54	P04040	CAT	Catalase	14	6	13	6	66	-1.64	0.0248
55	Q9ULZ3	PYCARD	Apoptosis-associated speck-like protein containing a CARD	9	3	13	3	58	1.54	0.0256
56	P07305	H1FO	Histone H1.0	7	3	15	3	58	1.69	0.0266
57	Q9GZP4	PITHD1	PITH domain-containing protein 1	3	2	8	2	65	-1.48	0.0266
58	Q02338	BDH1	D-beta-hydroxybutyrate dehydrogenase, mitochondrial	12	3	11	3	80	2.36	0.0280
59	P04062	GBA	Glucosylceramidase	2	2	4	2	50	1.32	0.0282
60	Q16543	CDC37	Hsp90 co-chaperone Cdc37	17	2	5	2	122	-1.36	0.0285
61	P22102	GART	Trifunctional purine biosynthetic protein adenosine-3	31	11	14	11	378	-1.84	0.0290
62	Q04828	AKR1C1	Aldo-keto reductase family 1 member C1	25	7	28	7	203	-2.18	0.0298
63	Q14344	GNA13	Guanine nucleotide-binding protein subunit alpha-13	13	2	6	1	235	1.96	0.0299
64	P07108	DBI	Acyl-CoA-binding protein	16	5	63	5	117	1.50	0.0307
65	P80188	LCN2	Neutrophil gelatinase-associated lipocalin	3	2	14	2	37	-2.67	0.0308
66	P62330	ARF6	ADP-ribosylation factor 6	12	3	23	3	139	2.02	0.0311
67	Q6P5R6	RPL22L1	60S ribosomal protein L22-like 1	2	1	10	1	26	1.70	0.0311
68	Q9UKS6	PACSN3	Protein kinase C and casein kinase substrate in neurons protein 3	2	1	2	1	89	-1.51	0.0312
69	Q9BZL1	UBL5	Ubiquitin-like protein 5	5	1	12	1	32	-1.58	0.0343
70	P14209	CD99	CD99 antigen	5	2	13	2	53	1.47	0.0350
71	Q81VT2	MISP	Mitotic interactor and substrate of PLK1	2	2	3	2	38	3.58	0.0351
72	P05187	ALPP	Alkaline phosphatase, placental type	5	2	4	2	35	-1.89	0.0354
73	P08240	SRPR	Signal recognition particle receptor subunit alpha	6	4	8	4	62	-1.33	0.0361
74	P48507	GCLM	Glutamate--cysteine ligase regulatory subunit	2	1	5	1	70	-1.60	0.0373
75	O43570	CA12	Carbonic anhydrase 12	2	2	9	2	118	3.83	0.0380
76	Q15046	KARS	Lysine--tRNA ligase	10	4	7	4	60	-1.53	0.0395
77	Q86SX6	GLRX5	Glutaredoxin-related protein 5, mitochondrial	12	4	32	4	212	1.22	0.0396
78	Q99426	TBCB	Tubulin-folding cofactor B	14	5	21	5	152	-1.37	0.0407
79	P35520	CBS	Cystathionine beta-synthase	4	3	6	3	36	-2.09	0.0410
80	Q9H8Y8	GORASP2	Golgi reassembly-stacking protein 2	7	2	8	2	159	1.64	0.0412
81	Q95571	ETHE1	Persulfide dioxygenase ETHE1, mitochondrial	14	5	31	5	213	-1.54	0.0414
82	Q8IUE6	HIST2H2AB	Histone H2A type 2-B	53	4	35	1	405	1.31	0.0416

Sr. no	Accession	Gene ID	Description	PSMs	Peptides	Coverage	Unique Peptides	Score Mascot	MDA-MB-468 (logFC)	MDA-MB-468 (P.Value)
84	P45954	ACADSB	Short	2	2	7	2	53	2.27	0.0433
85	Q8NE86	MCU	Calcium uniporter protein, mitochondrial	5	2	3	2	123	-1.37	0.0444
86	P26640	VAR5	Valine--tRNA ligase	45	17	16	17	566	-1.68	0.0448
87	Q9HD26	GOPC	Golgi-associated PDZ and coiled-coil motif-containing protein	5	2	5	2	110	-1.41	0.0460
88	P07477	PRSS1	Trypsin-1	15	1	4	1	115	1.77	0.0472
89	P31153	MAT2A	S-adenosylmethionine synthase isoform type-2	38	9	29	9	610	-1.19	0.0477
90	P07237	P4HB	Protein disulfide-isomerase	161	33	70	33	1376	1.21	0.0495
91	P09758	TACSTD2	Tumor-associated calcium signal transducer 2	4	3	11	3	59	-1.21	0.0496

3.3.5.6 Stem cell characteristics of Claudin-low cell line MDA-MB-231

Vimentin and CD44, which are markers of mesenchymal and stem cell-like properties in claudin-low tumours had the highest levels in MDA-MB-231 compared to the other cancer cell lines. CD44, a marker for breast cancer stem cells, plays an essential role in cell migration, tumour growth and progression by formation of invadopodia (Vikesaa et al., 2006). The expression of CD44 is reported to be higher in myoepithelial cells from basal tumours compared to those from luminal-derived malignancies (Charafe-Jauffret et al., 2005). The proteomics data correlated with these findings, exhibiting low expression in Luminal cells and higher expression in basal-like and Claudin-low subtypes. A signature, comprising 499 proteins ($p < 0.05$), was defined for claudin-low phenotype of breast cancer (Figure 3-15, Table 3-10 **Error! Reference source not found.**, Supplementary Table 3.2). The String analysis of these signature proteins reveal 468 biological process including intercellular transporters (140 components, 1.28×10^{-11}), cellular localization (103 components, 1.7×10^{-12}), Cell surface receptor signalling pathway ((68 components, 0.01), regulation of cell death (64 components, 2.1×10^{-5}) and apoptotic process (80 components, 0.0001). Detailed GO annotations for all signature proteins are presented in supplementary data, “Breast cancer specific-Go-Analysis” as CD format.

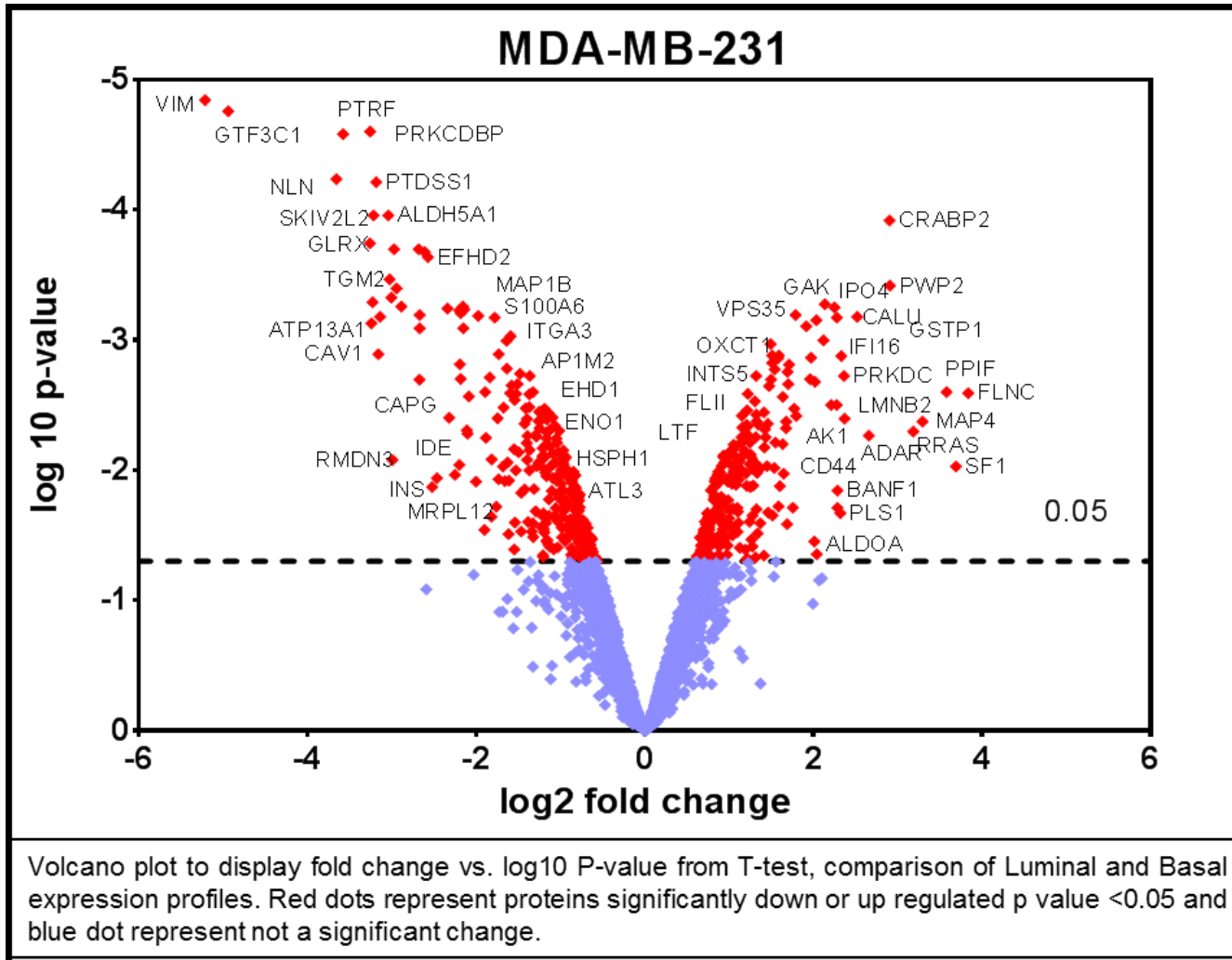


Figure 3-15: Volcano plot analysis of differentially expressed proteins in MDA-MB-231 cell line (Claudin-Low).

Table 3-10: Signature proteins for Claudin-Low breast cancer.

The table represents Uniprot Accession number, Gene ID, protein name, number of PSMs, total Peptides, unique peptides, protein coverage and Mascot score. LogFC: log fold change of MDA-MB-231 vs all cancer cell lines.

Sr. no	Accession	Gene ID	Description	PSMs	Peptides	Coverage	Unique Peptides	Score Mascot	MDA-MB-231 (logFC)	MDA-MB-231 (P.Value)
1	P08670	VIM	Vimentin	448	47	88	44	5307	-4.85	<0.0001
2	P29373	CRABP2	Cellular retinoic acid-binding protein 2	43	7	61	7	498	3.2	0.0001
3	Q9BYT8	NLN	Neurolysin, mitochondrial	5	1	1	1	66	4.74	0.0001
4	P42285	SKIV2L2	Superkiller viralicidic activity 2-like 2	4	2	2	2	38	3.41	0.0001
5	P35754	GLRX	Glutaredoxin-1	3	2	11	2	41	-2.83	0.0002
6	Q8N201	INTS1	Integrator complex subunit 1	2	2	1	2	45	5.45	0.0002
7	P21980	TGM2	Protein-glutamine gamma-glutamyltransferase 2	14	8	14	8	131	-3.29	0.0003
8	Q15269	PWP2	Periodic tryptophan protein 2 homolog	8	3	3	3	119	3.72	0.0004
9	P46821	MAP1B	Microtubule-associated protein 1B	4	3	2	3	31	-3.54	0.0005
10	Q09666	AHNAK	Neuroblast differentiation-associated protein AHNAK	227	82	36	82	1961	-2.58	0.0005
11	P40261	NNMT	Nicotinamide N-methyltransferase	6	2	7	2	81	-2.52	0.0005
12	P08758	ANXA5	Annexin A5	110	16	48	15	1162	-2.33	0.0006
13	Q9HD20	ATP13A1	Manganese-transporting ATPase 13A1	6	3	3	3	81	2.9	0.0006
14	Q9HCU5	PREB	Prolactin regulatory element-binding protein	2	2	9	2	76	2.49	0.0006
15	Q6P1N9	TATDN1	Putative deoxyribonuclease TATDN1	2	1	3	1	36	3	0.0006
16	O43852	CALU	Calumenin	28	6	17	6	316	-2.58	0.0007
17	Q03135	CAV1	Caveolin-1	8	2	12	2	210	-4.73	0.0007
18	P09211	GSTP1	Glutathione S-transferase P	37	9	61	9	360	-4.59	0.0007
19	P07195	LDHB	L-lactate dehydrogenase B chain	74	12	43	10	653	-3.81	0.0007
20	P09382	LGALS1	Galectin-1	116	9	81	9	1437	-2.61	0.0008
21	Q13330	MTA1	Metastasis-associated protein MTA1	4	2	3	1	36	-2.34	0.0008
22	Q15437	SEC23B	Protein transport protein Sec23B	22	6	9	4	325	2.6	0.0008
23	P55209	NAP1L1	Nucleosome assembly protein 1-like 1	26	6	18	5	518	-2.2	0.001
24	Q9UJZ1	STOML2	Stomatin-like protein 2, mitochondrial	7	4	16	4	112	-2.57	0.001
25	P26006	ITGA3	Integrin alpha-3	12	4	5	4	158	-2.1	0.0013
26	Q6UW68	TMEM205	Transmembrane protein 205	10	2	16	2	92	3.08	0.0013
27	Q96QK1	VPS35	Vacuolar protein sorting-associated protein 35	25	7	9	7	469	2.26	0.0013
28	Q16531	DDB1	DNA damage-binding protein 1	13	9	10	9	106	2.54	0.0014
29	Q8N6R0	METTL13	Methyltransferase-like protein 13	4	3	3	3	38	2.22	0.0014
30	O14976	GAK	Cyclin-G-associated kinase	3	2	2	2	41	3.69	0.0015
31	P67936	TPM4	Tropomyosin alpha-4 chain	76	19	54	8	855	-2.67	0.0016
32	Q6P9B9	INTS5	Integrator complex subunit 5	3	2	2	2	46	1.89	0.0017
33	Q15691	MAPRE1	Microtubule-associated protein RP	3	2	10	2	42	-2.36	0.0017
34	P78527	PRKDC	DNA-dependent protein kinase catalytic subunit	93	35	9	35	940	1.92	0.0019
35	P43490	NAMPT	Nicotinamide phosphoribosyltransferase	31	11	27	11	228	-2.91	0.0019
36	Q92973	TNPO1	Transporthin-1	15	9	12	9	193	1.95	0.0019
37	P62158	CALM3	Calmodulin	37	7	49	7	328	-2.05	0.002
38	P40121	CAPG	Macrophage-capping protein	15	5	18	5	151	-3.21	0.002
39	O14950	MYL12B	Myosin regulatory light chain 12B	60	7	52	3	1098	-2.43	0.002
40	A3KMH1	VWA8	von Willebrand factor A domain-containing protein 8	4	2	1	2	55	2.59	0.002
41	P62942	FKBP1A	Peptidyl-prolyl cis-trans isomerase FKBP1A	22	4	41	4	424	-1.99	0.0021

Sr. no	Accession	Gene ID	Description	PSMs	Peptides	Coverage	Unique Peptides	Score Mascot	MDA-MB-231 (logFC)	MDA-MB-231 (P.Value)
43	P13640	MT1G	Metallothionein-1G	22	3	34	3	174	-3.28	0.0022
45	P30405	PPIF	Peptidyl-prolyl cis-trans isomerase F, mitochondrial	6	4	26	3	24	-1.71	0.0025
46	P58107	EPPK1	Epiplakin	97	28	27	23	1634	2.45	0.0026
47	Q86Y46	KRT73	Keratin, type II cytoskeletal 73	28	5	5	1	168	5.34	0.0026
48	Q9H299	SH3BGRL3	SH3 domain-binding glutamic acid-rich-like protein 3	26	4	39	4	335	-1.8	0.0026
49	Q9BV44	THUMPD3	THUMP domain-containing protein 3	6	4	8	4	41	2.21	0.0029
50	P04179	SOD2	Superoxide dismutase [Mn], mitochondrial	35	9	51	9	285	-2.04	0.0031
51	O00186	STXBP3	Syntaxin-binding protein 3	3	2	5	2	26	2.35	0.0033
52	Q6P1M0	SLC27A4	Long-chain fatty acid transport protein 4	6	2	3	2	37	3.78	0.0034
53	Q9NRP0	OSTC	Oligosaccharyltransferase complex subunit OSTC	3	1	8	1	32	1.9	0.0035
54	Q96RP9	GFM1	Elongation factor G, mitochondrial	6	2	2	2	36	1.77	0.0036
55	Q9HD45	TM9SF3	Transmembrane 9 superfamily member 3	6	2	4	2	99	3.56	0.0038
56	O60664	PLIN3	Perilipin-3	46	10	39	10	956	-1.6	0.0039
57	Q04941	PLP2	Proteolipid protein 2	8	2	18	2	207	-2.81	0.0039
58	Q03252	LMNB2	Lamin-B2	43	15	26	13	352	-1.88	0.004
59	P27816	MAP4	Microtubule-associated protein 4	32	14	15	14	428	-1.52	0.0042
60	Q6P2Q9	PRPF8	Pre-mRNA-processing-splicing factor 8	12	9	4	9	160	1.67	0.0042
61	Q9ULA0	DNPEP	Aspartyl aminopeptidase	12	4	11	4	196	2.07	0.0045
62	Q8N766	EMC1	ER membrane protein complex subunit 1	4	3	4	3	23	1.79	0.0046
63	Q9H6S3	EPS8L2	Epidermal growth factor receptor kinase substrate 8-like protein 2	6	2	3	2	54	-2.05	0.0047
64	P42704	LRPPRC	Leucine-rich PPR motif-containing protein, mitochondrial	53	20	16	20	640	1.77	0.0047
65	Q8WWI5	SLC44A1	Choline transporter-like protein 1	3	1	2	1	61	1.99	0.0048
66	P07814	EPRS	Bifunctional glutamate	64	21	16	21	649	1.99	0.0049
67	Q6UWP7	LCLAT1	Lysocardiolipin acyltransferase 1	2	2	4	2	35	1.69	0.0049
68	Q9BVQ7	SPATA5L1	Spermatogenesis-associated protein 5-like protein 1	4	3	3	3	43	1.46	0.0053
69	P53992	SEC24C	Protein transport protein Sec24C	7	4	5	4	109	1.81	0.0054
70	Q9UGP8	SEC63	Translocation protein SEC63 homolog	4	2	3	2	140	2.81	0.0054
71	Q7Z2K6	ERMP1	Endoplasmic reticulum metalloproteinase 1	5	2	3	2	85	2.61	0.0055
72	Q15121	PEA15	Astrocytic phosphoprotein PEA-15	10	1	9	1	248	-2.55	0.0056
73	O95671	ASMTL	N-acetylserotonin O-methyltransferase-like protein	3	1	2	1	98	4.41	0.0056
74	Q9P2E9	RRBP1	Ribosome-binding protein 1	47	20	22	20	215	-1.55	0.0056
75	P62258	YWHAE	14-3-3 protein epsilon	138	16	69	13	1806	-1.57	0.0057
76	Q9BQ69	MACROD1	O-acetyl-ADP-ribose deacetylase MACROD1	7	3	14	3	187	1.6	0.006
77	P78417	GSTO1	Glutathione S-transferase omega-1	23	6	24	6	234	-2.01	0.0061
78	O94826	TOMM70A	Mitochondrial import receptor subunit TOM70	5	3	5	3	50	1.88	0.0061
79	Q8WV74	NUDT8	Nucleoside diphosphate-linked moiety X motif 8	3	1	4	1	44	2.78	0.0061
80	O43399	TPD52L2	Tumor protein D54	29	6	31	6	475	-1.77	0.0063
81-499	Full list of 499 proteins for claudin low breast cancer cell lines is given in supplementary data.									

3.3.6 Western blot analysis

CD44 antigen (CD44), Aldehyde dehydrogenase, mitochondrial (ALDH2), calreticulin (CALR), prostaglandin E synthase 3 (PTGES3) and 10kDa heat shock protein, mitochondrial (HSPE1), were analysed by western blot (Figure 3-16). All immunoblots were analysed by GelAnalyzer 2010, to determine the quantitative expression of target proteins in all breast cell lines (Table 3-11).

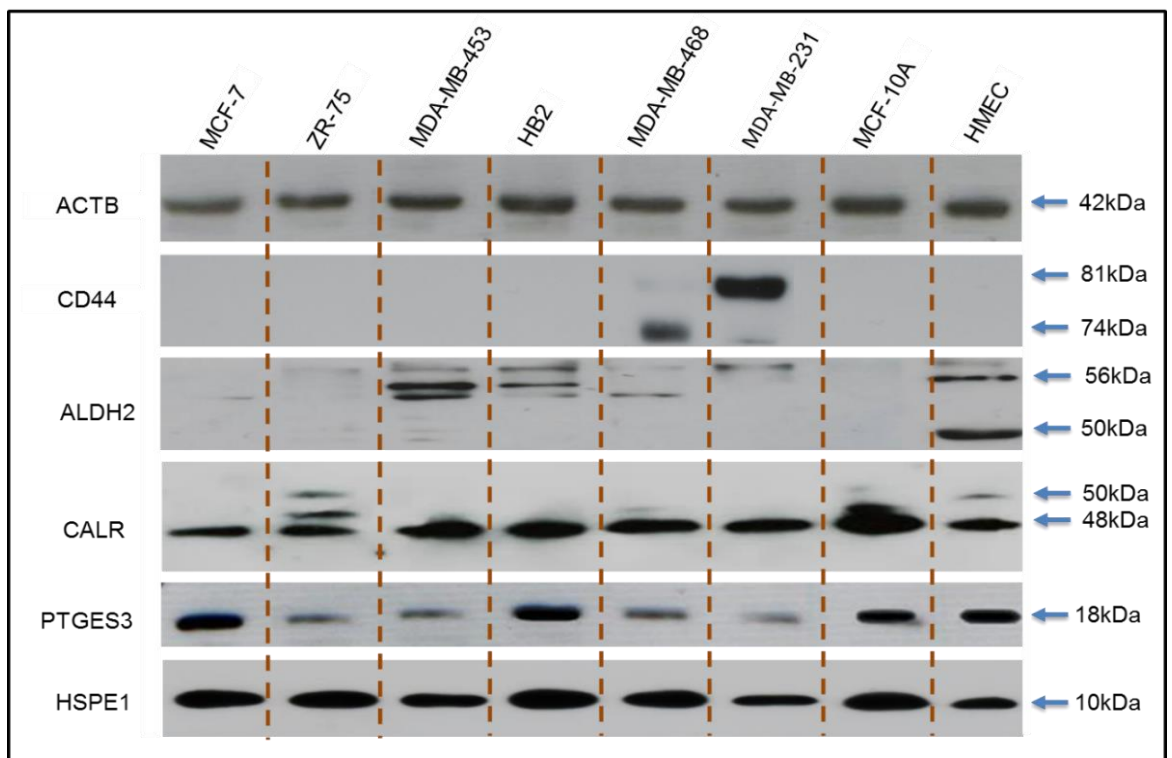


Figure 3-16: Western blot analysis of ACTB, CD44, ALDH2, CALR, PTGES3, HSPE1 in breast cell lines.

ACTB is a house-keeping protein and was used as a loading control in Western blotting. Its molecular weight is approximately 42 kDa (Uniprot.org, 2015). CD44 was only detected in basal-like (MDA-MB-468) and Claudin-low (MDA-MB-231) breast cancer cell lines, but with distinctly different molecular weights of 74 and 81kDa respectively. ALDH2 exhibited differential

expression and extensive heterogeneity with 6 bands between 50kDa and 58kDa, with protein absent completely in only MCF-10A (Figure 3-16). The expression of ALDH2 (combine quantification of Isoforms 1 and 2), was increased in HER2 positive cell lines (ZR-75 and MDA-MB-453) but decreased in basal type breast cancer (MDA-MB-468 and MDA-MB-231). Calreticulin was expressed in all cell lines tested with one major band (48kDa), but less intense bands present up to 52kDa in a cell line-specific manner (Figure 3-16). Post-translational modifications, N6-acetyllysine and glycosylation have been reported for calreticulin and may contribute to this heterogeneity (Uniprot.org, 2015). PTGES3 was observed as a band at 18kDa with stronger expression in MCF-7, normal breast cell lines (HB-2 and MCF-10A), and primary epithelial cells (Figure 3-16).

	ACTB			PTGES3			HSPE1		
Cell line	Band intensity (raw)	Band intensity (norm)	Fold Change	Band intensity (raw)	Band intensity (norm)	Fold Change	Band intensity (raw)	Band intensity (norm)	Fold Change
MCF-7	2029	2246	1.22	3082	3411	1.78	7635	8450	2.45
ZR-75	2071	2246	1.19	994	1078	0.56	7519	8153	2.37
MDA-MB-453	2308	2246	1.07	937	912	0.48	5817	5659	1.64
HB2	2601	2246	0.95	2785	2404	1.25	8520	7356	2.14
MDA-MB-468	2183	2246	1.13	1119	1151	0.60	7182	7388	2.14
MDA-MB-231	2099	2246	1.18	592	633	0.33	4339	4642	1.35
MCF-10A	2349	2246	1.05	2141	2047	1.07	7378	7053	2.05
HMEC	2469	2246	1.00	2108	1917	1.00	3788	3445	1.00
	CALR			ALDH2			CD44		
Cell line	Band intensity (raw)	Band intensity (norm)	Fold Change	Band intensity (raw)	Band intensity (norm)	Fold Change	Band intensity (raw)	Band intensity (norm)	Fold Change*
MCF-7	1607	1778	1.10	266	294	0.06	ND	ND	ND
ZR-75	2581	2798	1.73	310	336	0.07	ND	ND	ND
MDA-MB-453	3051	2968	1.83	5813	5656	1.18	ND	ND	ND
HB2	2866	2474	1.53	2758	2381	0.50	ND	ND	ND
MDA-MB-468	2476	2547	1.57	996	1025	0.21	2490	2561	2.50
MDA-MB-231	2313	2474	1.53	826	884	0.18	5726	6126	5.97
MCF-10A	3622	3462	2.14	ND	ND	ND	ND	ND	ND
HMEC	1781	1620	1.00	5267	4790	1.00	ND	ND	ND

All immunoblots were analysed by GelAnalyzer 2010, to determine the raw peak intensities. After normalisation of all target proteins with ACTB, fold change of each protein was calculated compared to primary human mammary epithelial cells (HMEC). (*); fold change was calculated by average expression of CD44 in MDA-MB-468 and MDA-MB-231. ND; not detected.

Table 3-11: Quantitative expression of ACTB, CD44, ALDH2, CALR, PTGES3 and HSPE1 in breast cell lines.

3.4 Discussion

3.4.1 MS2 and SPS quantification

Proteomics has the potential to study the expression of proteins in complex biological samples. Developments in protein fractionation and labelling techniques have also improved identification of less abundant proteins in biological samples. The enzymatic digestion of proteins prior labelling, increase the complexity of samples and result inherent drawback of reporter ions technology (Zieske, 2006). The focus of this quantitative approach was to understand the fact that co-isolation of multiple precursors in iTRAQ experiments compromises the dynamic range of reported ratios (Savitski et al., 2013, Ting et al., 2011). Different data acquisition schemes are used on Ultraflex II and Orbitrap Fusion in quantitative proteomics. Every full MS scan is followed by data-dependent CID MS² scans in the Ultraflex II analysis, both identification and quantification of peptides/ proteins are derived from MS² spectra. A key challenge in MS² based iTRAQ quantification, is underestimation of fold change because of compression of the iTRAQ ratio. The propensity for underestimation was noted in comparative biological profiling, where confirmation of differential expression of selected targets was analysed through other methods such as western blotting (Keshamouni et al., 2006a, Glen et al., 2008).

10 kDa heat shock protein, mitochondrial (HSPE1), is a chaperone that interacts in a two-step folding mechanism of protein biogenesis in the mitochondria (Corrao et al., 2010) . In SPS data, we found that HSPE1 was

up-regulated in MDA-MB-231 and MCF-10A cell lines (Table 3-3, Figure 3-17) while no significant different was observed in MS² data (Table 3-3) and WB analysis; a single band at 10kDa was detected in all cell lines (Table 3-11, Figure 3-16, Figure 3-17). The over expression of HSPE1, a cancer related gene, has been reported in glioma, breast, head and neck, liver, lungs, melanoma, prostate ovarian, cervical, renal, testis and thyroid (Shaheed et al., 2015, Cappello et al., 2007). In our previous study, Shaheed et al 2013, we found the increased expression of HSPE1 in tissue biopsies collected from invasive carcinoma patients compared to normal (Shaheed et al., 2013a).

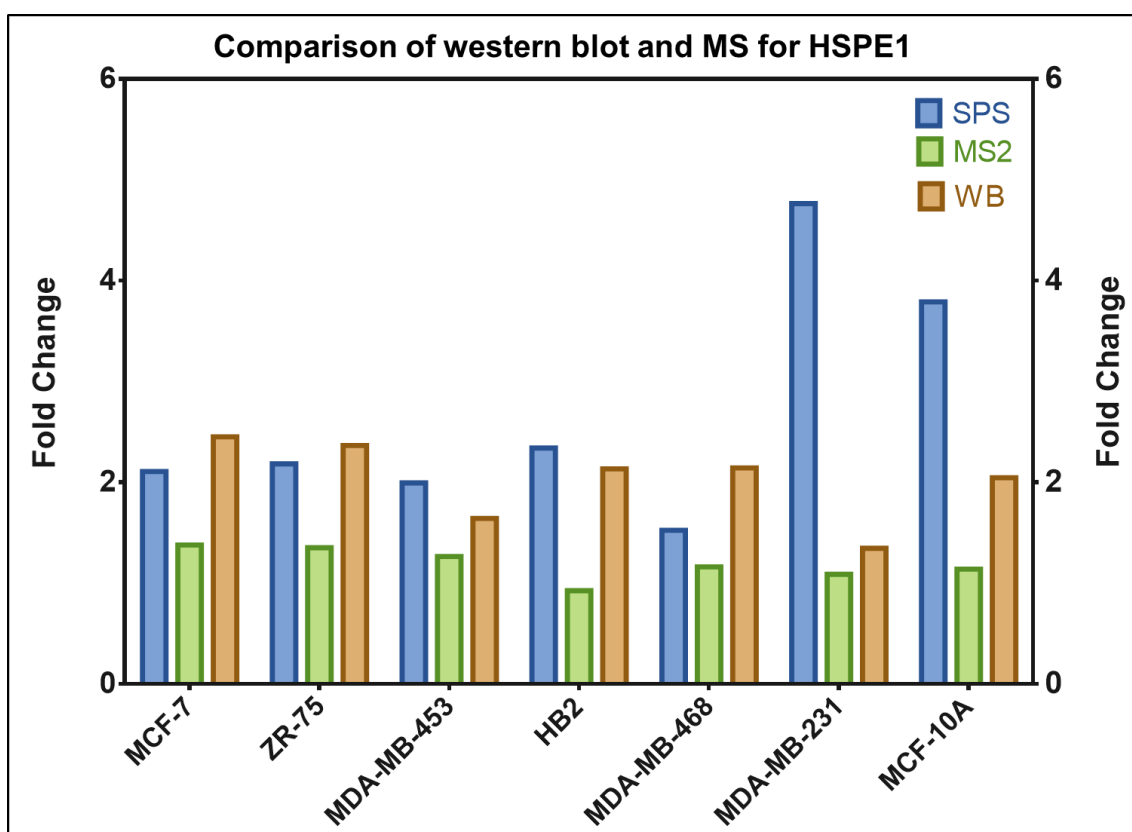


Figure 3-17: Comparison of western blot analysis and MS results for HSPE1 in breast cell lines.

When the data for highest scoring peptide of HSPE1, (VLQATVVAVGSGSK), acquired on Orbitrap Fusion (Mascot Score; 71) and

Ultraflex II (Mascot score; 64); was evaluated, similar fragmentation patterns was observed on both instruments (Figure 3-18, A, B). The low dynamic range for iTRAQ ratios were observed in MS² spectra, acquired on Ultraflex II (Figure 3-18, C). The synchronous precursor selection (SPS)-based MS³ is a unique feature of Fusion, where SPS mode alleviates dynamic range compression compared to MS²-based analysis. In SPS mode reporter ions are isolated from multiple MS² fragment ions in parallel and analysed in MS³ fragmentation (Ting et al., 2011). The reporters' ions acquired through SPS mode of MS analysis, had wider dynamic range compared to MS² quantifications were also observed in our data (Figure 3-6, Figure 3-18, D).

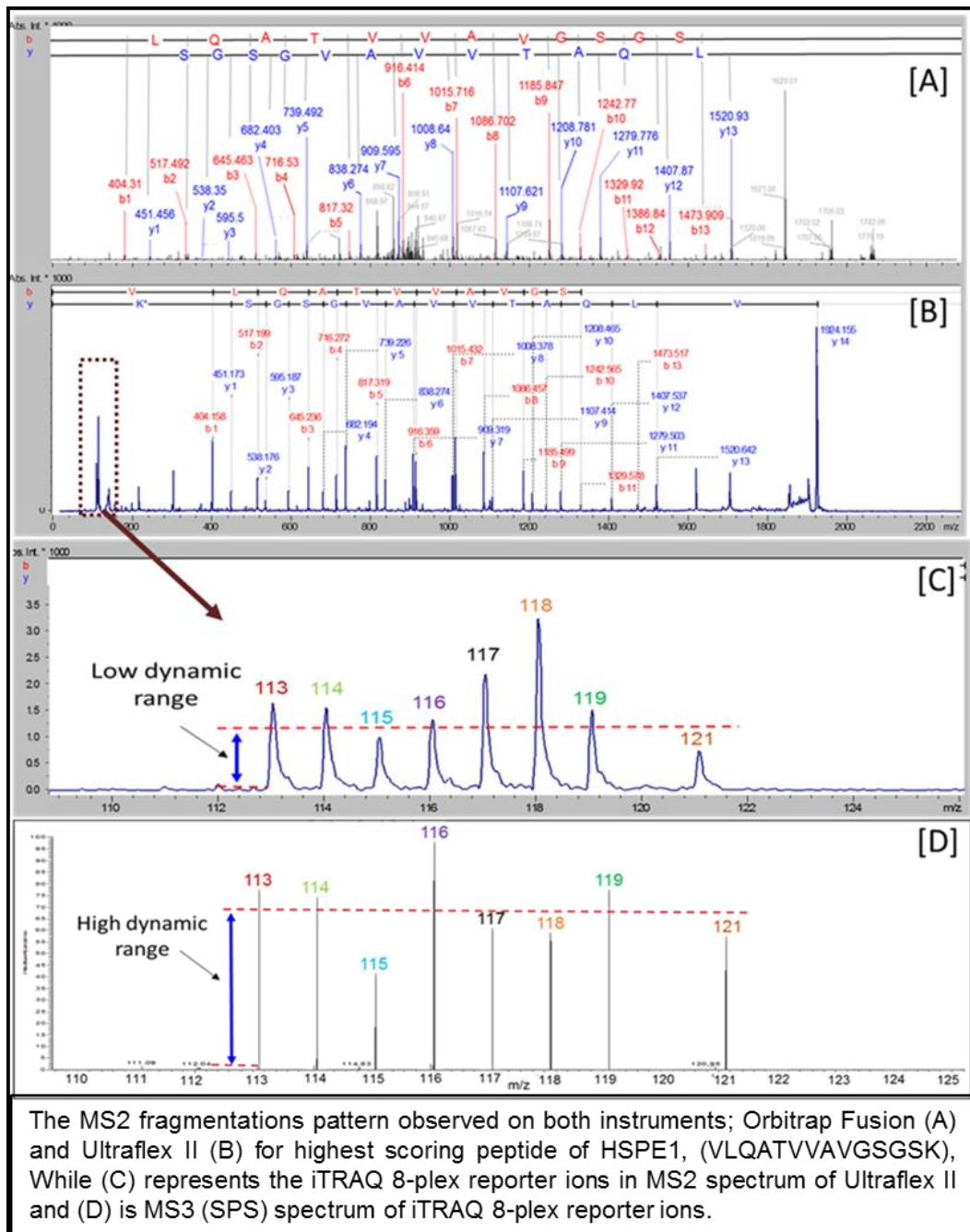


Figure 3-18 Comparison of MS² fragmentations for a peptide of HSPE1 and dynamic range of reporter ions on Orbitrap Fusion and Ultraflex II.

3.4.2 Immortalised cell lines compared to primary cells

In this study, the proteomes of five phenotypes of breast cancer; luminal A (MCF-7), luminal B (ZR-75), basal-like (MDA-MB-468), claudin-low (MDA-MB-231) and HER2-positive (MDA-MB-453), two normal breast cell lines; MCF-10A and HB2, and human mammary epithelial cells (HMEC) were

compared. Significantly up-regulated or down-regulated proteins were clustered according to types of cell lines compared to HMEC. GO enrichment analysis was performed by calculating a p-value, if an annotation term is enriched in certain set of gene products relative to all genes in the genome, providing biological significance to the data (Cox and Mann, 2012).

Gene ontology enrichment for proteins up-regulated in all cell lines compared to HMEC indicated a strong correlation (PPI enrichment p-value, $1.96e-09$, Figure 3-19), with DNA replication (KEGG Pathways, p-value <0.001 , 3 proteins). The histone family (Histone H2B type 1-K, H2AFZ, Histone H2A.Z, Histone H2B type 3-B), involved in transcription regulation, DNA repair, DNA replication and chromosomal stability, expression is increased in all cell lines compared to HMEC (Geiger et al., 2012). Three components (MCM2, MCM3 and MCM5) of the toroidal hexameric MCM (mini-chromosome complex) ring complex (p-value $5.23e-05$, 3 proteins), a helicase required for DNA replication and elongation, and associated with replication specific DNA polymerase alpha (Remus, 2016, Raynaud et al., 2014), were increased in expression across all the cell lines. Cyclin-dependent kinase 1 (CDK1), responsible for cell proliferation, DNA replication and apoptosis was also increased in both cancer and transformed normal cells compared to HMEC (Castedo et al., 2002). However, the increased level of the MCM complex and CDK1, in normal-like breast cell lines (MCF-10A and HB2), indicates the change in expression was not due to oncogenic processes but transformation-related processes.

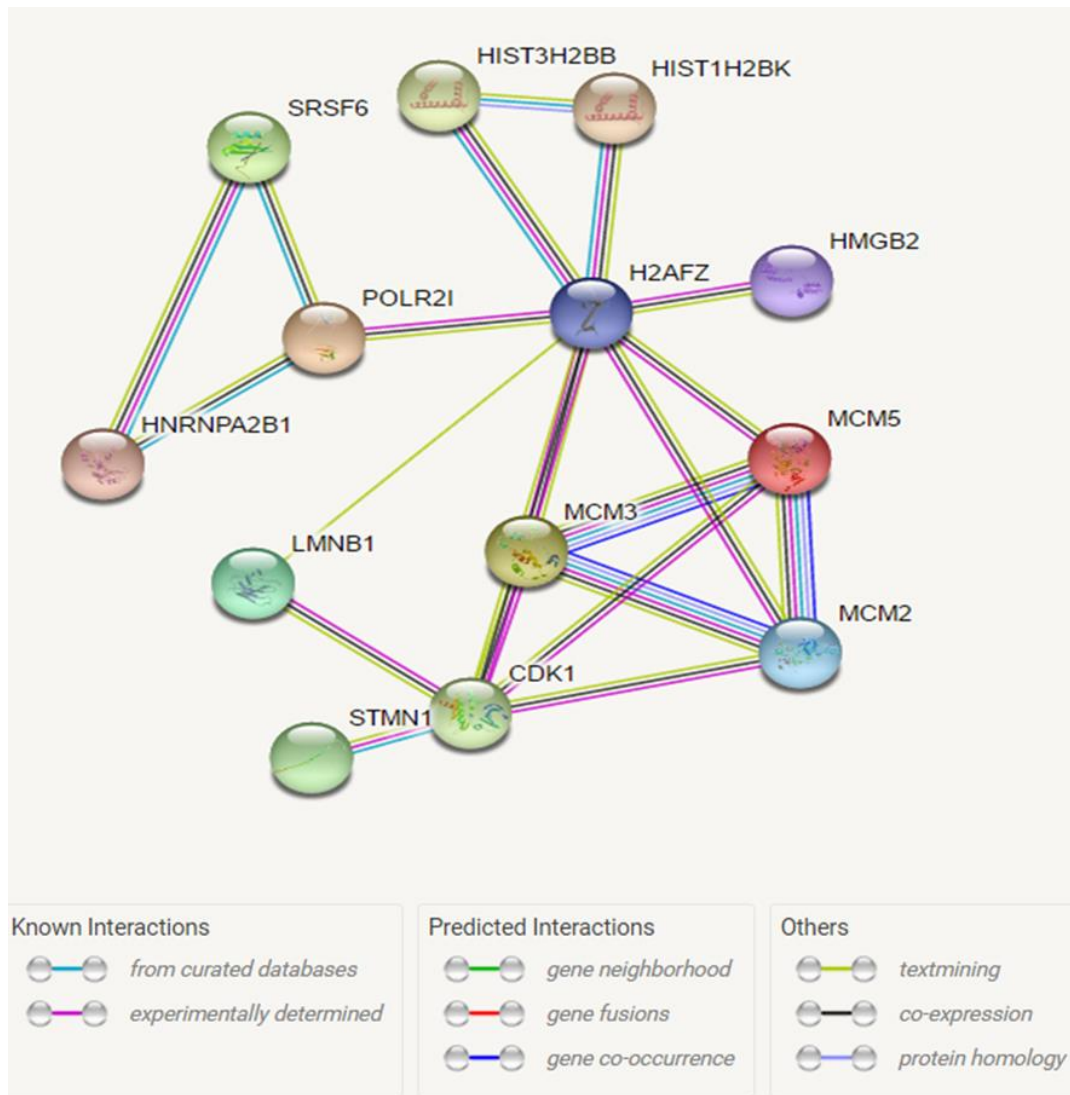


Figure 3-19: Up-regulated DNA replication protein-protein interactions ubiquitous to all cell lines compared to HMEC

Down-regulated in all cell lines indicated that protein processing in the extracellular matrix organization (p-value 3.38e-8, 13 components), focal adhesion (p-value 9.2e-7, 9 components), ECM-receptor interaction pathways (p-value 1.52e-5, 6 components) and multicellular organismal development (p-value 5.61e-6, 28 components) were the main biological functions suppressed (Figure 3-20). Normal myoepithelial cells secrete extracellular matrix (ECM) components and express high level of adhesion proteins, to provide structural and biochemical support in tissue formation (Adriance et al., 2005). Hence the expression of ECM and adhesion proteins

was generically decreased in transformed and tumour cells grown as cell cultures. The most dramatic changes occurred in cell adhesion proteins such as integrin alpha-2 (ITGA2), integrin alpha-V (ITGAV), and integrin beta-4 (ITGB4), highlighting their importance for detachment of the tumour cells from the original tissue. This correlated with previous studies that showed low expression of ITGA2 and ITGB4 in metastatic sites of breast cancers (Geiger et al., 2012). The loss of adhesion proteins in the cancer cell lines, reflected the general collapse of normal tissue architecture, which is a known feature of the development of carcinomas. Conversely, the ability of cells to grow singularly has reduced cell-cell interaction functionality and lowered expression of adhesion/ extracellular matrix interaction receptors and cytoskeletal proteins that would have provided discrete morphology within the normal tissues environment. Adenoviral and lentiviral vectors are commonly used for virus-mediated transfection because it is highly efficient and easy to achieve sustainable transgene expression. However, virus-mediated transfection has a drawback; it may cause an inflammatory reaction and insertional mutations such as activation of oncogenes by disrupting tumour suppressor genes (Kim and Eberwine, 2010). Not surprisingly, the transformation of cells to immortalisation, with dysregulation of the molecular processes which control replication, results in increased expression of proteins responsibly for DNA replication. That these changes were observed in virally-transfected normal breast cell lines as well as cancer cell lines indicates that caution is required when considering the use of the latter in the study of distinct oncologic responses, as they may not be cancer-specific at all.

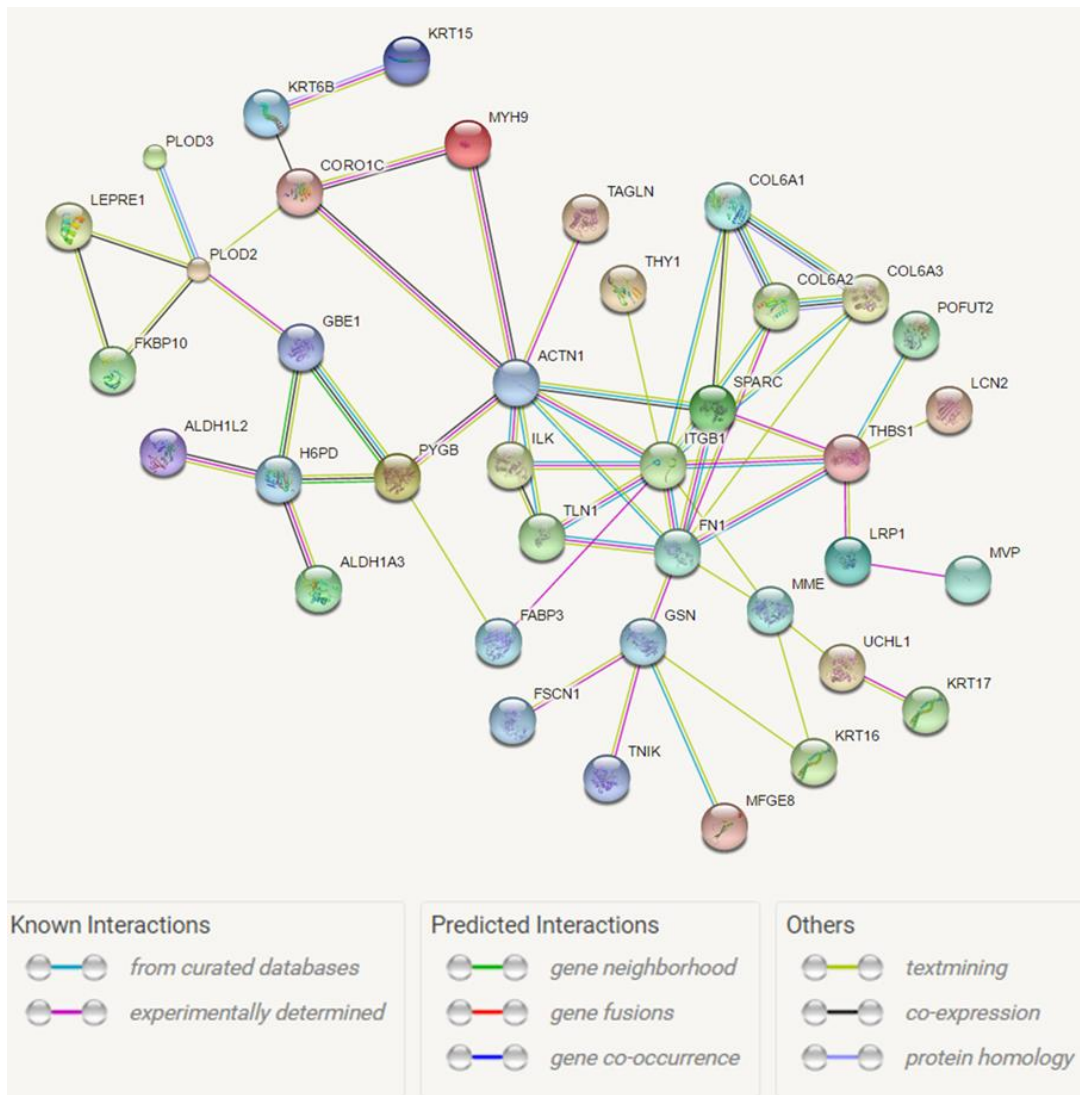


Figure 3-20: Protein-protein interaction map of down regulated proteins for all cell lines compared to HMEC.

Cell lines are widely used in cancer research as in vitro models because they are easy to handle and provide unlimited supply of a relatively homogeneous cell population. On the other hand, when these cell lines are cultured on plastic in two dimensions, they lack the complex inter-relationships that exist between cells in vivo and also continual culturing can introduce genotypic and phenotypic drift in cell lines (Holliday and Speirs, 2011). These drawbacks can be controlled by using primary cultures derived directly from tumour or normal tissues.

3.4.3 Breast cancer classifiers

Because of the heterogeneity of breast cancer, it is challenging to differentiate between each subtype on the basis of a single gene or protein. Therefore, a group of markers are required that can serve as a signature for diagnosing different types of breast cancer. Gene expression profiling has played an important role in understanding the heterogeneity of breast cancer at a molecular level and refining the taxonomy based on the presence of productive markers like ER, PR and HER2 to more sophisticated grouping comprising luminal A, luminal B, basal-like, claudin-low and HER2-positive phenotypes (Holliday and Speirs, 2011).

Two of the four main phenotype classifiers (HER2 and EGFR) were detected and interestingly exhibited diametrically opposite expression, with luminal cancer cell lines expressing HER2 (highest levels in HER2-positive MDA-MD-453) and low levels of EGFR and vice versa in basal cancer cell lines (highest levels of EGFR in MDA-MB-468, Figure 3-21) (Holliday and Speirs, 2011). In lung cancer cell lines EGFR has been shown to play an important role in the delivery of Mucin-4 to the plasma membrane from the endoplasmic reticulum, and hence a direct impact on EGFR-mediated proliferation (Li et al., 2014c). Furthermore, it has been proposed as a candidate serum biomarker in pancreatic ductal carcinoma (Makawita et al., 2013).

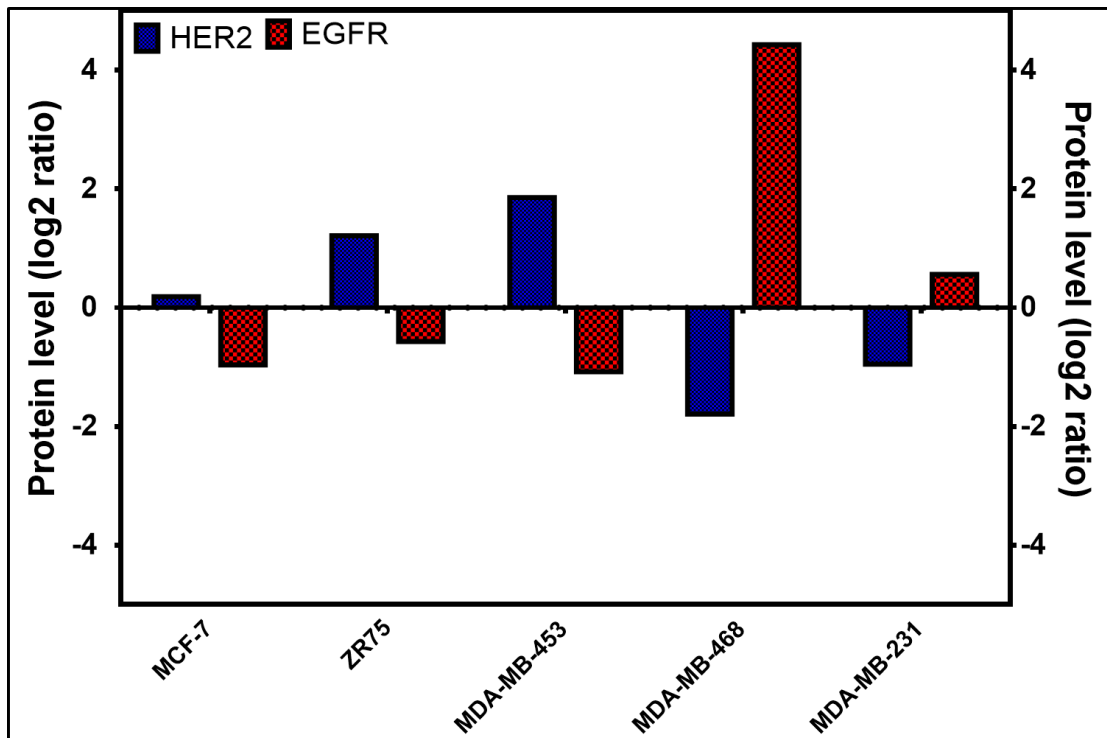


Figure 3-21: Expression of HER2 and EGFR in breast cancer cell lines.

The well-known markers of Luminal type breast cancer; ER and PR, were not detected in luminal-type cells (MCF-7, ZR-75, MDA-MB-453) which is a limitation of the iTRAQ qualification approach. By combining the proteome of 8 samples after iTRAQ labelling, the increased complexity of the sample results in identification of fewer peptides and proteins (Hultin-Rosenberg et al., 2013).

The most significantly increased protein in luminal-type (MCF-7, ZR-75, MDA-MB-453 and HB2) consisted of Anterior gradient protein 2 homolog (AGR2), which is involved in cell migration, cell differentiation (Wang et al., 2008), whereas three tumour related proteins; Integrin alpha-3 (ITGA3), involved in breast metastasis (Seguin et al., 2015), epidermal growth factor receptor (EGFR) and cellular tumour antigen p53 (TP53), were decreased in luminal compare to basal-type (MDA-MB-468 and MDA-MB-231).

Total 19 members of Keratin family were detected in Orbitrap data; 12 of these components (KRT5, KRT6A, KRT6B, KRT7, KRT10, KRT14, KRT15, KRT16, KRT17, KRT77, KRT80, KRT81) expression was decreased in all cell lines as compared to HMEC. The expression of Keratins; KRT1, KRT2, KRT8, KRT9, KRT18, KRT19, and KRT73 was higher in luminal A (MCF-7) and luminal B (ZR-75) cell lines, whereas for all other cell lines, the expression was unique; KRT1, KRT 2, KRT 9, were decreased in MDA-MB-468 and KRT73 was down-regulated in HER2 (MDA-MB-453) and claudin-low (MDA-MB-231) cell lines. Interestingly, keratin family have been considered as basal markers (Shao et al., 2012), but our results show that their expression demarcates cancer stage rather than subtype.

Prostaglandin E synthase 3 (PTGES3) is a multifunctional protein that modulates the activity of aryl hydrocarbon receptor and, in conjunction with HSP90, increases the affinity of steroid hormone receptors for their respective ligands (Sullivan et al., 1997, Oxelmark et al., 2006). It is also involved in the recruitment of steroid receptors and telomerase to the nucleus where they regulate transcriptional expression of target genes (Toogun et al., 2008); it stabilizes specific kinases and has glutathione-dependent cytoplasmic prostaglandin E synthase 3 enzyme activities (Tanioka et al., 2003). Overexpression of PTGES3 in breast cancer enhances ER-dependent transcriptional events including promoting transition from non-invasive to invasive cells through activation of metastasis-related genes, and has been established as a potential target to prevent development of secondary tumours. Proteomics analysis indicated that PTGES3 was increased in all cell lines, including normal cell lines compared to HMECs,

with highest levels in MCF-7 (luminal A) (Figure 3-16, Figure 3-22 and Table 3-11).

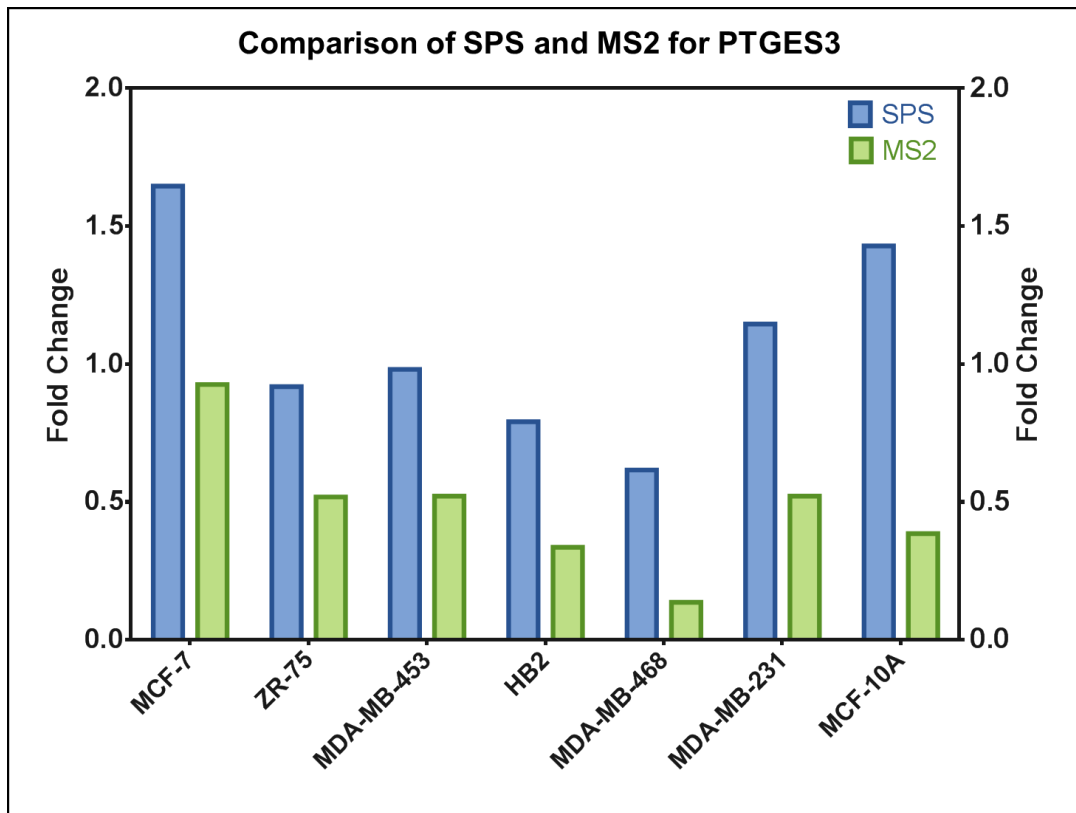


Figure 3-22: Comparison of SPS analysis and MS results for PTGES3 in breast cell lines.

Gene encoding for isocitrate dehydrogenases 2 (IDH2) was observed at higher levels in Luminal B (ZR-75) as compared to luminal A (MCF-7), with no change in other breast cancer phenotypes. IDH2 is frequently mutated in enchondroma (benign tumour of bones) and glioma cancer, resulting in abnormal histone and DNA methylation, a common feature of tumorigenesis (Yang et al., 2012). The expression of programmed cell death protein 4 (PDCD4), involved in apoptosis and negative regulator of tumour invasion by inhibiting mitogen-activated protein kinases 85 (MAPK85) (Frankel et al., 2008), was higher in ZR-75 as compare to all other cancer cell lines. A study on 420 patients with ER-positive breast cancer

found that down-regulation of PDCD4 was associated with a lower rate of disease-free survival and high histologic grade of breast tumours (Chen et al., 2015). Calmodulin-like protein 5 (CALML5) was significantly ($p < 0.0001$) up-regulated gene in HER2 positive phenotype. CALML5 binds with HER2 and stimulates neoplastic transformation and tumour progression (White et al., 2011). The increase expression of both proteins (CALML5 and HER2) in HER2+ (MDA-MB-453) breast cancer cell line was observed in our data.

The high expression of aldehyde dehydrogenase (ALDH) isoforms is associated with poor clinical outcome in human cancer; breast, colon, bone, head and neck, lung, liver, brain, thyroid, pancreas, skin (melanoma), cervix, prostate and bladder (Rodriguez-Torres and Allan, 2016). A Immunohistochemistry based analysis of tissues from 160 patients with breast cancer, found higher expression of ALDH4A1 but no significance difference for ER, PR and HER2 status (Qiu et al., 2014). Aldehyde dehydrogenase family 1 member A3 (ALDH1A3) and Mitochondrial 10-formyl tetra hydro folate dehydrogenase (ALDH1L2) decreased in all cell lines. Different studies reported the low expression of ALDH2 in oropharyngeal, laryngeal and esophageal cancers, with increased risk of alcohol-associated cancer (Seitz and Stickel, 2010). The levels of ALDH2 were significantly decreased in invasive cancer breast tissues comparing with normal tissues and adenoma breast cancer tissues (Shaheed et al., 2013a). In iTRAQ data, the expression of ALDH2 was decreased in all transformed cell lines compared to HMEC but Western blot results showed the increase expression in HER2 (MDA-MB-453) and Luminal B (ZR-75) cell types (Figure 3-16). iTRAQ quantification does not discriminate between different isoforms of a protein or heterogeneity that may be caused by post translation modification

but western blot can differentiate these factors. This may in part be due to variable expression of two isoforms; isoform-1 (expected molecular weight = 56.38kDa) and isoform-2, which lacks a sequence of 47 amino acid residues within the N-terminal region of the protein (expected molecular weight = 50.99kDa), but may also be due to processing of an N-terminal transit peptide or N6-acetyllysine PTMs. The right isoform of ALDH2 can be further validated by analysing the 56.38kDa and 50.99kDa bands by GEL-LC-MS analysis (Shevchenko et al., 1996). These observations highlight the value of Western blotting in identifying molecular weight variation of the intact protein that was not detected by MudPIT proteomics.

S100 protein family performs a wide range of intracellular and extracellular functions; (i) cell proliferation, (ii) apoptosis, (iii) cytoskeleton interactions, (iv) regulation of calcium homeostasis, (v) protein phosphorylation, (vi) cell invasion and motility, (vii) regulation of transcriptional factors, (viii) inflammation, (ix) chemotaxis and (x) autoimmunity (Chen et al., 2014). The altered expression of S100 proteins, was associated with tumour progression and prognosis (Chen et al., 2014). S100A proteins comprise 21 subunits (Chen et al., 2014), of which 9 were detected in Orbitrap Fusion analysis. Of these, 3 exhibited the opposite expression in basal-like cell line compare to other breast cancer cell lines. The expression of S100A13 was decreased but S100A8 and S100A9 was increased in MDA-MB-468. The S100A13 is involved in cancer cell motility, invasion and migration (Chen et al., 2014). The S100A8 and S100A9 work as single unit and play role in cancer cell differentiation, apoptosis, cell cycle and growth (Chen et al., 2014).

Basal type cells showed increased level of CD44 antigen (CD44) compared to other cells, and is involved in cell migration, evasion of apoptosis, tumour growth, angiogenesis and progression (Vikesaa et al., 2006, Louderbough and Schroeder, 2011). CD44 performed these functions independently or in conjunction with other cellular components. CD44 has dual functionality; either it can activate tumour growth by Rho GTPases, Ras-MAPK and PI3K/AKT pathways, or can inhibit angiogenesis and invasion by promoting apoptosis(Louderbough and Schroeder, 2011). High expression of CD44 has been reported in Head and neck cancer, melanoma, pancreatic, breast, cervical, lymphoma, colorectal and lungs cancer. CD44 is an accepted marker for breast cancer stem cells (BCSCs) and high expression of CD44 in BCSCs, promote tumorigenesis and metastasis (Smith and Cai, 2012). In this study, higher expression of CD44 in claudin low (MDA-MB-231) and basal-like (MDA-MB-468), compared to luminal-type breast cancer, was observed. Western blot analysis of same samples; showed increased expression of CD44 in basal type cell lines, verifying iTRAQ-based quantification (Figure 3-15 and Figure 3-16, Table 3-11, Figure 3-23). So, higher expression of CD44 in MDA-MB-468 and MDA-MB-231, confirmed the stem cell like origin of these cell lines.

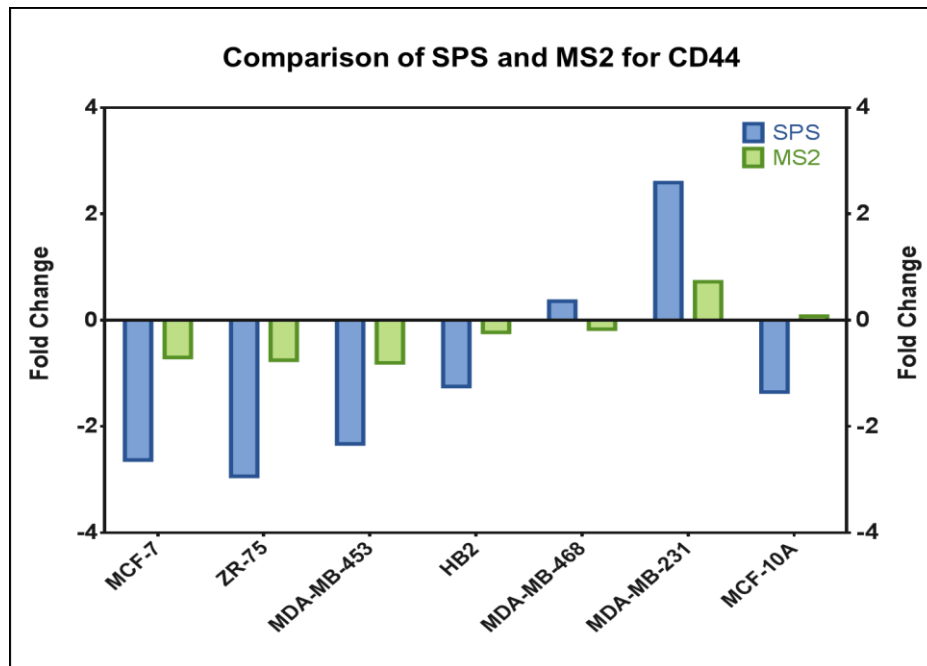


Figure 3-23: Comparison of SPS analysis and MS results for CD44 in breast cell lines.

The high expression of vimentin (VIM), a marker of epithelial-to-mesenchymal transition (EMT), is related with chemo-resistance and high tumour invasion and metastasis (Makki et al., 2015, Thompson et al., 1992). Macrophage-capping protein (CapG), has been reported as an oncogene with increased expression in breast, colorectal, pancreatic and ovarian cancer (Glaser et al., 2014, Thompson et al., 2007, Neumann et al., 2014). Caveolin-1 (CAV1), is not only involved in tumour invasion and metastasis but also plays a role in Ras-ERK pathway by promoting cell cycle (Sloan et al., 2004). The increased expression of CAV1 inhibits primary breast tumour growth and spontaneous metastasis of breast cancer (Sloan et al., 2004) and plays important role in EGR induced migration and proliferation of stem cells (Park et al., 2005). Cytoskeleton-associated membrane protein 4 (CKAP4) is an epithelial cell surface receptor for anti-proliferative factor signalling and elevated level has been reported in breast, cervical, lungs, liver, ovarian and pancreatic cancer (Li et al., 2014a, Kimura et al., 2016, Li et al., 2014b). The

expression of VIM, CAV1, CKAP4 and CapG, was high in MDA-MB-231 compared to all other breast cancer cell lines, indicate high level of tumour invasion and metastasis associated with claudin-low breast cancer.

Calreticulin (CALR) is a chaperone protein located in the endoplasmic reticulum, which induces a variety of cellular functions, including the control of calcium function as part of cell signalling, operating exclusively through Integrin alpha 3 subunit (Coppolino et al., 1997). In this iTRAQ proteomics approach, the expression of CALR was decreased in MCF-7 breast cancer cell line as compare to HMEC and was also verified by Western blotting (Figure 3-16 and Figure 3-24, Table 3-11). But high level of CALR was observed in claudin-low (MDA-MB-231) breast cancer. Over expression of CALR has been proposed as a biomarker due to the high levels observed in urine of bladder cancer patients (Kageyama et al., 2009) and increased expression was observed with the progression of breast cancer from fibroadenoma to invasive carcinoma (Shaheed et al., 2013a). An immunohistochemical study of gastric cancer also showed that CALR was found to be correlated with high micro-vessel density, serosal and perineural invasion, lymph node dissemination, and poor patient survival (Chen et al., 2009). However, in neuroblastoma, the most common malignancy in infants, positive immunohistochemical staining for CALR was correlated with improved prognosis and patient survival (Hsu et al., 2005). The low expression of CALR in lung cancer cell lines was associated with a stimulating effect on rate of proliferation (Bergner et al., 2009). Higher expression of CALR is also reported in oesophageal squamous carcinoma cells and plays important role in cell migration and metastasis (Shi et al., 2014).

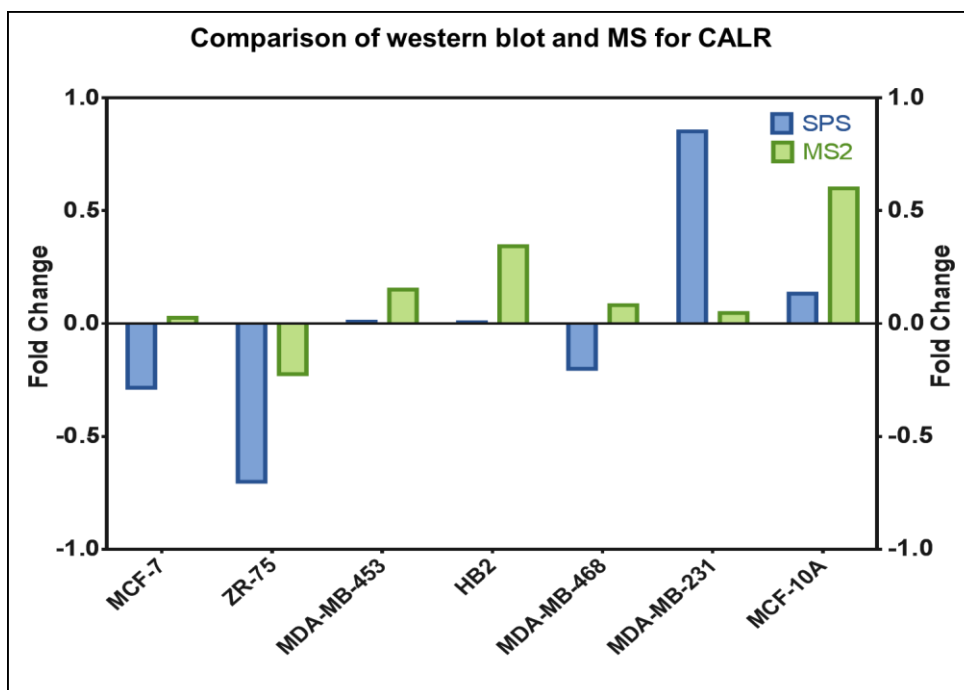


Figure 3-24: Comparison of SPS analysis and MS results for Calreticulin in breast cell lines.

In Summary, this study builds on other deep proteome characterisation of breast cancer (Geiger et al., 2012, Lawrence et al., 2015, Tyanova et al., 2016) and represent the first deep iTRAQ based-proteome characterization the breast cancer phenotypes compared to primary human mammary epithelial cells. The proteome profiling of breast cancer cell line introduced here provides the research community with an additional resource to select the most appropriate MS instrument and model for their research.

3.5 Conclusions

Current developments in proteomics techniques allow high-throughput analyses for the detection, identification, and functional investigation of proteome. Using an Orbitrap Fusion operated in CID-MS²-Ion trap and HCD-MS³ (SPS)-Orbitrap configuration, improved the number of PSMs, peptides, Mascot score, protein sequence coverage, number of unique proteins detected compared to Ultraflex II. SPS mode acquisition on Orbitrap Fusion also decreased the dynamic range compression of iTRAQ ratios, due to background proteome and noise signals

Our proteomics approach on breast cell lines, captured the general biological processes that are altered in cancer and premalignant cell lines compared primary human mammary epithelial cells. Signature proteins for five different phenotypes of breast cancer; luminal A (MCF-7, 32 proteins), luminal B (ZR-75, 40 proteins), HER2 (MDA-MB-453, 38 proteins), basal like (MDA-MB-468, 89 proteins) and largest signature (495 proteins) associated with the stem cell-derived cancer cell line, MDA-MB-231, were found. These markers can be further investigated in biofluids on larger scale, to define their diagnostics and prognostic values for human breast cancer. The quantitative proteomics data presented here, obtained from 5 breast cancer cell lines, will yield biomarkers discoveries that will help in the diagnosis and prognosis of breast cancers. The biological characterisation of biofluids also provides a unique insight into tumour microenvironment such as immune system responses, regulation of cellular interactions and tumour related proteins secretion and shedding.

CHAPTER 4. PROTEOME PROFILING OF NIPPLE ASPIRATE FLUID.

4.1 Introduction

There remains an unmet need to provide high risk premenopausal women with a regular and convenient means of breast cancer screening. Breast secretions (nipple discharge [ND] and nipple aspirate fluid [NAF]), collectively herein described as NAF, were investigated, as they are produced by epithelial cells lining breast ducts and lobules which are most commonly associated with the disease. These secretions are protein-rich and therefore ideally suited for biomarker discovery. My aim, in this study, was to investigate the protein composition of breast secretions from stage-specific patients using state-of-the-art proteomics techniques and determine if the profiles provided diagnostic value.

Although the number of women, aged 40 or less, diagnosed with breast cancer is relatively low, they experience a more aggressive forms of the disease (frequently defined by triple negative or HER2 positive phenotypes) with poorer clinical outcome (Azim and Partridge, 2014, Copson et al., 2013). They are often at higher risk due to a genetic predisposition towards the disease, of which mutations in BRCA1 and BRCA2 breast cancer susceptibility genes are the best characterised, accounting for approximately 15% of all breast cancer cases (Aloraifi et al., 2015). Awareness of these variants, by germ line genetic testing, informs the patient of the life-time risk of susceptibility to the disease compared to the general population, but it does not tell the patient when the disease will occur (Rieder et al., 2016).

Unfortunately, the discovery of a mutation can result in preventative intervention by elective surgery to remove both breasts or prophylactic administration of tamoxifen before the disease has occurred (Davies et al., 2015).

Mammography has been very successful in detecting breast cancer in post-menopausal women (98% sensitivity), but less so in younger women due to image obfuscation by breast density (30 to 48% sensitivity) (Kolb et al., 2002, Mandelson et al., 2000). An inaccurate diagnosis leads to high call-back rates, elevated costs of unnecessary biopsies, increased radiation dose exposure and patient anxiety during re-screening. For successful diagnosis by mammography, a substantial mass is required, which may have already metastasised in aggressive forms of the disease. Also, mammography cannot differentiate between benign micro-calcifications associated with low risk DCIS, which do not require surgery, and higher risk DCIS that will progress to an invasive tumour (Espina and Liotta, 2011), thereby resulting in over-diagnosis and over-treatment (Francis et al., 2015).

Therefore, new methods that can be used, safely and routinely, for the early detection of breast cancer are required to support high risk younger women. The search for diagnostic biomarkers of breast cancer has been extensive and proteomics strategies increasingly employed as part of the discovery process (Zeidan et al., 2015). Plasma is by far the most common biofluid used, but putative markers are massively diluted relative to the site of origin of the cancer, thereby reducing sensitivity (Loo et al., 2010). As an alternative, I have chosen to analyse nipple aspirate fluid (NAF), which is collected by massage or breast pump, and differentiating them from liquid

biopsies collected by lavage or using needles). NAF originates from cells associated with 85% of breast carcinomas and comprises a diverse range of endogenous substances such as micronutrients (tocopherols,, cholesterol, carotenes) (Djuric et al., 2007), hormones (estradiol, estrone, progesterone and testosterone) (Chatterton et al., 2010), carbohydrate (Thomsen Friedenreich and Tn) antigens (Deutscher et al., 2010b), microRNA (Canto et al., 2016), and microbes (Chan et al., 2016b). NAF has multiple advantages as a liquid biopsy for detection of breast cancer: (i) pre-menopausal women are more likely to produce NAF than post-menopausal women where ductal atrophy may be prevalent (Baltzell et al., 2006), (ii) NAF expression is non-invasive, causing minimal discomfort compared to other breast cancer screening procedures (De Groot et al., 2015) (iii) enables procurement of matched pairs of samples which may provide an intra-individual control for comparing disease with healthy, (iv) biomarkers remain highly concentrated for analysis (compared to blood and urine), and (v) minimal sample preparation is required, compared to tissues, therefore excluding yield-reducing protein extraction steps. NAF volumes are small, but protein concentrations are high and more than sufficient for replicate analyses with state-of-the-art mass spectrometric techniques.

A number of studies have been undertaken to characterise the NAF proteome many of which have been summarised by Pavlou et al, as part of a comparison with their own dataset of 854 proteins (Pavlou et al., 2010). In that study, 3 healthy individuals and 3 patients, was analysed on Orbitrap mass analyser coupled to a linear ion trap, after Multiple fractionation methods (size-exclusion and anion-exchange chromatography). Another deep proteome study of NAF was conducted by Brunoro et al 2015, which

again based on different fractionation methods such as strong cation-exchange (SCX) and pI-based OFFGEL fractionation (Brunoro et al., 2015). The fractionated samples were analysed on an LTQ-Orbitrap XL and identified 4466 peptides corresponded to a total of 557 proteins. A more recent study by Kurono et al 2016 identified 372 proteins in breast cancer patients and healthy volunteers (n = 19 and 12, respectively) (Kurono et al., 2016). In this study, the samples were fractionated; first at high pH, then on low pH before MS analysis. All these studies were optimisation of fractionation techniques, to increase protein IDs and sequences coverage in NAF samples.

My objective in this study, was to deal with the fundamental definitions of NAF composition and determine if the proteins present constitute biologically and physiologically relevant information for diagnosing breast health. In so doing I will identify if NAF has the potential to be a suitable liquid biopsy for biomarker discovery. Following preliminary characterisation by protein determination and SDS PAGE, I used a semi-quantitative (label-free) comparison of proteomic profiles of matched pairs from four cases, (1) a healthy volunteer (HV), and patients with (2) benign phyllodes (PB), (3) DCIS (PD) and (4) invasive carcinoma (PI). The proteomic profiles were subject to statistical and gene ontological analysis to glean significant expression changes relating to disease state.

4.2 Materials and methods

4.2.1 Patients and sample collection

NAF samples were obtained from healthy volunteers and breast cancer patients, who presented to Bradford Teaching Hospitals NHS Trust, between 2013 and 2016. All participants gave written informed consent to undergo bilateral nipple aspiration. The study protocol was approved by University of Bradford's Independent Scientific Advisory Committee (reference: application/13/051). Ethical approval was given by Leeds (East) Research Ethics Committee, reference 07/H1306/98+5. Before aspiration was attempted, the nipple was initially cleansed with an alcohol pad. NAF collection from cancer patients was performed under general anaesthetic by the clinical team, prior to surgery, assisted by massaging the breast and the liquid collected from the nipple using a sterile pipette. After collection, the samples were transferred to chilled, pre-labelled tubes containing a freeze-dried protease inhibitor cocktail mixture (Roche Diagnostics), and frozen within 30 minutes of collection. Where possible, NAF samples were collected separately from both breasts. NAF from healthy volunteers was collected in a similar manner by the individuals whilst conscious. From a bank of 100 patient samples, 15 NAF pairs were selected on the basis of a utilisable volume of sample, right and left breast expression and disease stage, in order to optimise LC-Ms assay.(Table 4-1).

Case	Pathology	Disease Breast	Grade	TNM Staging	Phenotype	Age	Pre/post menopausal	Ethnicity	Breast	NAF - Appearance	Conc. mg/ml	Total protein (mg)
1	IC (ductal)	Right	3	G3 pT1c pN0	ER-ve/HER2-ve	55	post	caucasian	Left Right	Brown-cloudy Brown-cloudy	3.43 8.49	103 255
2	IC (ductal)	Right	2	G2 pT1b N0	ER+ve/HER2-ve	76	post	caucasian	Left Right	light yellow light yellow	53.64 19.01	13409 1901
3	IC (ductal)	Left	1	ND	ER+ve/HER2-ve	59	post	caucasian	Left Right	colourless olive colour	3.28 15.82	522 1725
4	Normal	N/A	N/A	N/A	N/A	48	pre	caucasian	Left Right	colourless, white cloudy layer on top colourless, white cloudy layer on top	15.86 18.20	793 910
5	IC (ductal)	Left	1	G1 pT1c, N1a, Mx	ER+ve/HER2-ve	66	post	ND	Left Right	green-cloudy white-cloudy	12.68 9.54	634 477
6	IC (ductal)	Left	3	G3 pT1b N0	ER+ve/HER2+ve	49	pre	ND	Left Right	brown green brown	10.11 11.80	505 590
7	IC (lobular)	Right	3	G2 T3 N1	ER+ve/HER2-ve	49	pre	ND	Left Right	light green light green	3.08 9.25	62 462
8	Benign (fibrocystic)	Left	N/A	N/A	N/A	39	pre	caucasian	Left Right	brown white-cloudy	1.59 12.52	3 626
9	Benign (phyllodes)	Right	N/A	N/A	N/A	43	pre	caucasian	Left Right	white-cloudy light brown-cloudy	16.53 21.50	793 1892
10	DCIS	Left	N/A	N/A	N/A	62	post	caucasian	Left Right	yellow-cloudy yellow-cloudy	26.80 27.52	750 770
11	IC (ductal)	Right	3	G3 pT2 pN0	ER+ve/HER2-ve	50	post	caucasian	Left Right	light brown-cloudy light brown-cloudy	35.14 22.74	1687 637

Case	Pathology	Disease Breast	Grade	TNM Staging	Phenotype	Age	Pre/post menopausal	Ethnicity	Breast	NAF - Appearance	Conc. mg/ml	Total protein (mg)
12	IC (ductal)	Right	3	G3 T1c	ER+ve/HER2 -ve	86	post	caucasian	Left Right	light brown-cloudy light brown-cloudy	58.81 69.48	11643 8893
13	IC (ductal)	Left	2	G2 pT1b pN0	ND	56	post	caucasian	Left Right	white-cloudy white-cloudy	4.84 8.73	121 218
14	IC (lobular)	Left	2	G2 pT3 N2	ER+ve/HER2 -ve	68	post	caucasian	Left Right	gray-cloudy white brown-cloudy	26.35 10.14	527 203
15	Normal	N/A	N/A	N/A	N/A	22	pre	caucasian	Left Right	greenish-clear white-cloudy	21.32 32.53	597 1561

ND; not determined or not declared, **N/A**; not applicable, **G1**; Well differentiated (low grade) breast cancer, **G2**; Moderately differentiated (intermediate grade) breast cancer, **G3**; Poorly differentiated (high grade) breast cancer, **pT1b**; tumour is more than 0.5 cm but not more than 1 cm, **T1c**; tumour is more than 1 cm but not more than 2 cm, **pT2**; tumour is more than 2 centimetres but no more than 5 centimetres across, **pT3**; tumour is bigger than 5 centimetres across, **pN0**; there are no cancer cells in any nearby nodes, **N1**; cancer cells are in the lymph nodes in the ampit but the nodes are not stuck to surrounding tissues, **pN1a**; cancer cells have spread (metastasised) into 1 to 3 lymph nodes and at least one is larger than 2mm, **N2a**; cancer cells in the lymph nodes in the ampit, which are stuck to each other and to other structures, **MX**; Metastasis can't be assessed , **ER**; Estrogen receptors, **HER2**; human epidermal growth factor receptor 2.

Table 4-1: Healthy volunteer and patient parameters and NAF characteristics.

4.2.2 Sample preparation

The samples were centrifuged for 1 minute to remove particulate matter and the supernatants collected. The protein concentration was measured using the Bradford assay (Bio-Rad Laboratories) and paired samples (20µg) analysed by SDS PAGE as described previously chapter 2; materials and methods. Gels were stained with PhastGel Blue R (in 5% acetic acid, 50% methanol) for protein detection.

4.2.3 Proteomic analysis

An aliquot of each sample (200 µg) was reduced with 50 mM dithiothreitol (DTT) for 15 minutes at 60°C, alkylated with 100mM iodoacetamide (IAA) at ambient temperature for 15 minutes, and digested by modified sequencing grade trypsin (Fisher Scientific); protease-to-protein mass ratio of 1:20 (w/w) at 37°C for 20 hrs. After digestion, each sample was desalted on an Isolute C₁₈ RP LC column (Kinesis Ltd) and lyophilized.

4.2.3.1 SCX Peptide fractionation

Trypsin-digested NAF samples were re-suspended in SCX loading buffer (10mM KH₂PO₄ in 25% v/v acetonitrile [ACN], 0.01% w/v sodium azide, adjusted to pH3) and added to an Isolute SCX column (Kinesis), equilibrated with SCX loading buffer. Peptides were eluted in 12 fractions with stepwise increasing potassium chloride concentration from 0 to 1000mM (see chapter 2 for full details). Eluted fractions were diluted with 2% v/v ACN, 0.05% v/v formic acid (FA), desalted on an Isolute C₁₈ RP LC column (Kinesis) and lyophilised.

4.2.3.2 Fusion Orbitrap analysis

The lyophilised SCX fractions were re-suspended in 10µl of 0.1% of FA and analysed in triplicate (3µL/injection) on a nano-LC UltiMate 3000 capillary HPLC system coupled to an Orbitrap Fusion™ Tribrid™ Mass Spectrometer (Chapter 2, section 2.10.2 for full details). Samples were applied at 25 µL/minute and washed on a C₁₈, 300 µm × 5 mm, 5 µm diameter, 100 Å PepMap pre-column (ThermoFisher) before transfer to a C₁₈, 75 µm × 50 cm, 2 µm diameter, 100 Å PepMap column, (ThermoFisher) (please see Table 2-3 for full details). The Orbitrap Fusion parameters were as follows: for full MS spectra, the scan range was m/z 350–1500 with a resolution of 120,000 at m/z 200. All MS/MS acquisition was performed on Ion-trap, in top speed mode with 3 second cycle time, a dynamic exclusion (±5 ppm) of 60 seconds, intensity threshold 5000, with ions of charge states 2+ to 7+ sequentially fragmented by collision-induced dissociation (CID) with a normalized collision energy (NCE) of 35%. A maximum of 200 ms ion injection time was allowed.

4.2.3.3 Data Analysis

MS/MS fragment mass lists were searched, via Proteome Discoverer version 2.1 (ThermoFisher) using Mascot software version 2.4 (Matrix Science) with a percolator (strict FDR of 0.01 and a relaxed FDR of 0.05) against SwissProt version 2016 containing 552,259 human protein sequences with search parameters: trypsin digestion, 2 missed cleavages, variable modification of methionine oxidation, fixed modifications of cysteine (carbamidomethylation), precursor mass tolerance of 10 ppm, MS/MS fragmentation mass tolerance of 0.5 Da and a 95% confidence interval

threshold ($p < 0.05$, Mascot score ≥ 23). Non-redundant protein profiles for each NAF sample were created by combining the corresponding LC–Fusion datasets. The list of protein identifications, was assessed manually, and only Master Proteins (i.e. contain unique peptides) were accepted.

Protein quantitation was defined as the sum of the peak areas of the three strongest parent signals. To allow comparison of sample protein profiles, quantitation of each protein was normalised relative to the median peak area sum of the whole protein complement in the sample. Pearson correlation coefficient was calculated using PRISM 6.0 software (GraphPad Software), to determine gross similarities of paired samples. To identify cases-specific (healthy vs disease [HV vs PB, PD, PI], benign vs healthy and cancer [PB vs HV, PD, PI], non-cancer vs cancer [[HV, PB vs PD, PI]) protein expression, averaged data of paired (left and right breast) samples was used. Student t-tests were undertaken using Excel 2010 to identify significantly ($p < 0.05$) expressed proteins. A functional Enrichment analysis tool, FunRich 2.1.2 (<http://www.funrich.org/>) was used to compare proteomes of NAF samples. Proteins which met the required thresholds for identification were submitted to Database for Annotation, Visualization and Integrated Discovery (DAVID, version 6.8) analysis. Default settings were used for functional annotation with Benjamini-corrected p-values of < 0.05 , deemed significant. Protein-protein interaction analysis was performed using STRING version 10.0 (<http://string-db.org/>). Proteins were also subject to analysis in TMHMM Server v. 2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>) (Krogh et al., 2001), to determine those with transmembrane helical regions.

4.2.4 Western blot analysis

NAF samples, equivalent to 20µg of protein, were analysed by Western blotting blotting (see Chapter 2 section 2.8) using primary antibodies; anti-beta-actin mouse monoclonal, anti-EGFR mouse monoclonal, anti-vimentin mouse monoclonal, anti-CYP3A4 rabbit monoclonal, and anti-CD44 rabbit monoclonal, then followed by appropriate secondary antibodies (rabbit anti-mouse IgG-conjugated with horse radish peroxidase, or goat anti-rabbit IgG-conjugated with horse radish peroxidase (Table 2-1 for full details of antibodies). All immuno-blots were analysed by GelAnalyzer 2010a software (<http://www.gelanalyzer.com>).

4.3 RESULTS

4.3.1 NAF sample characterisation

My preliminary objective was to characterise matched pairs of NAF samples using basic biochemical procedures, measuring total protein amount, concentration and matched pairs visualisation by SDS PAGE, which has not been reported previously. NAF samples were collected from 100 breast cancer patients and healthy volunteers and grouped into 4 clinical stages - invasive carcinoma (IC), ductal carcinoma in situ (DCIS), benign lesions and healthy. From these, 15 pairs (13 with cancer and 2 non-cancer) were characterised for volume (varying from 4 to 500 μ L) and protein concentration (3 to 70mg/ml (Table 4-1). Samples were analysed by SDS PAGE demonstrating that, in the majority of cases, pairs from the same person had similar profiles (Figure 4-1, A to D). Some cases exhibited a dominant serum albumin band, suggesting a high plasma content (Figure 4-1, A, Case 2), whilst others had a relatively low albumin presence (Figure 1, A Cases 1, 3 and 4). Based on sample colour there was no indication of blood in the latter group, no apparent correlation with disease compared to healthy, and hence was not indicative of tissue damage or tumour invasiveness. Four matched pairs, a healthy volunteer (Case 4, HV), a patient with benign phyllodes tumour (Case 9, PB), a patient with DCIS (Case 10, PD) and a patient with invasive carcinoma (Case 12, PI) (Table 4-1), were selected for proteomic analysis, based on similar protein concentration (to minimise samples preparation variation) and protein quantity (providing sufficient material for validations studies).

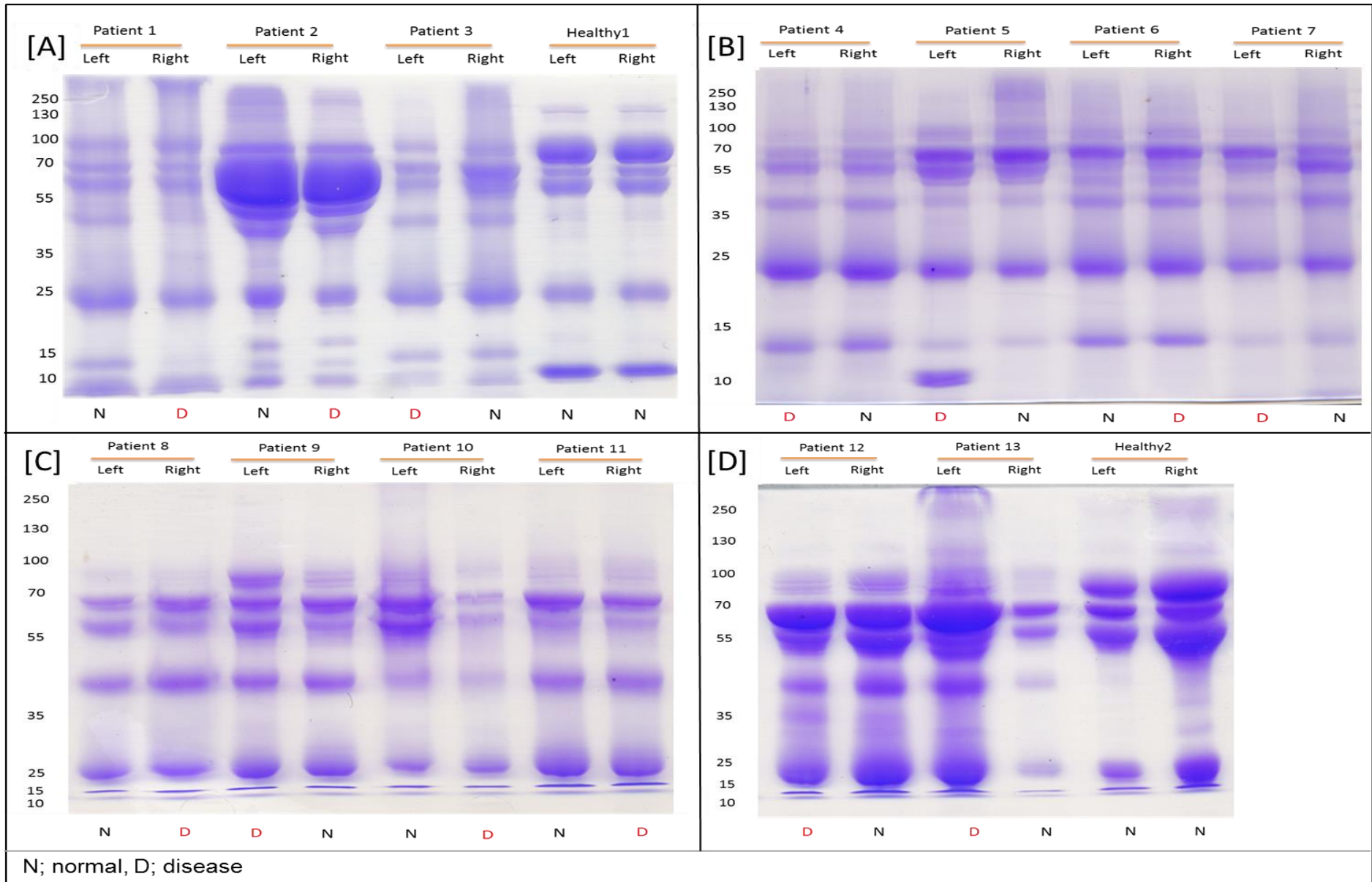


Figure 4-1: SDS PAGE analysis of matched pairs of NAF samples.

4.3.2 NAF proteomic analysis

For NAF to be a useful clinical sample it was important to establish the quality and complexity of proteomic data that can be achieved using a 2D LC separation approach as this had implications on sensitivity and analysis time. From 1D LC analyses of the 8 NAF samples, a total of 661 different proteins were identified ($p < 0.05$). An average of 389 proteins was detected per sample, with 134 common to all the samples (Figure 4-2, A). From 2D LC MS analysis, a total of 1990 proteins were identified ($p < 0.05$), with an average of 1265 proteins (Figure 2A) and 567 found in all the samples (Figure 4-2, B). At a qualitative level, 644 of the proteins were identified in both 1D and 2D approaches, with 17 proteins uniquely detected by 1D LC MS and 1346 additional proteins were observed in 2D LC MS (Figure 4-2, D). Of the proteins identified by both 1D and 2D analysis, 117 were found in all samples. Comparison of this subset of common proteins indicated that 2D LC analysis provided more confident identification, with an average 19 unique peptides per protein (compare to 14 with 1D LC), 25% increase in sequence coverage and 4-fold more spectra associated with each protein to improve label-free quantification. The Coefficient of variation (CV%) of Mascot score, peptides, and PSMs, was less in 2D proteome as compare to 1D proteome for 117 common proteins (Figure 4-2, C). Therefore, despite the longer analysis time of 2D LC MS (48 hrs per NAF sample, compared to 9 hours for 1D), it provided a much richer source of information for NAF characterisation, hence only the proteomic content of the 2D LC dataset will be described further.

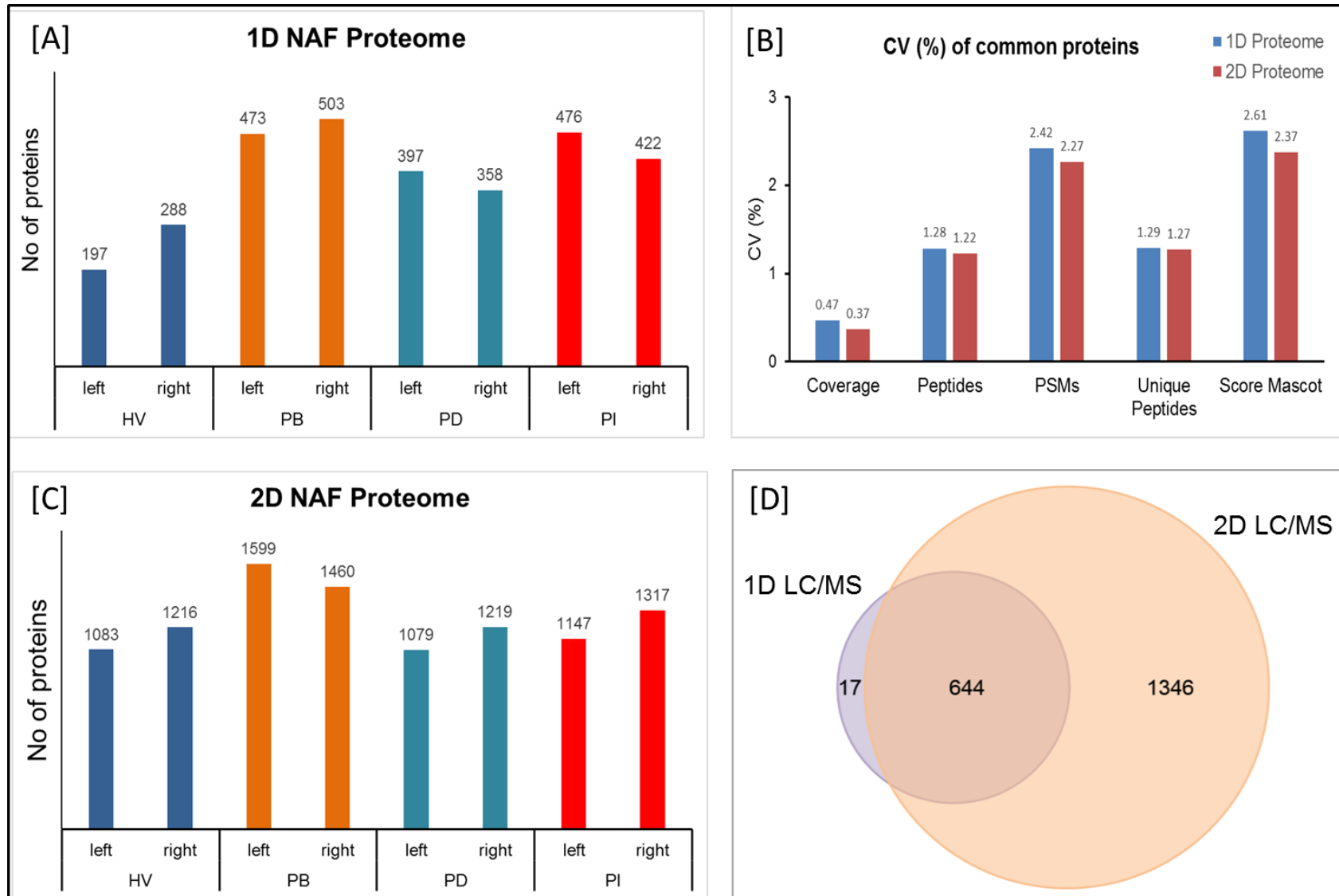


Figure 4-2: Comparison of 1D and 2D NAF proteome profiling.

4.3.3 Comparison of identified 2D NAF proteome

Comparison of pairs of proteomics profiles from 2D LC analysis showed greater than 50% similarity in protein composition (Figure 4-3). The matched pair of NAF have higher correlation (average 77%) while healthy volunteer has less similarity (average 52%) compared to NAF collected from patients with benign, DCIS and IC.

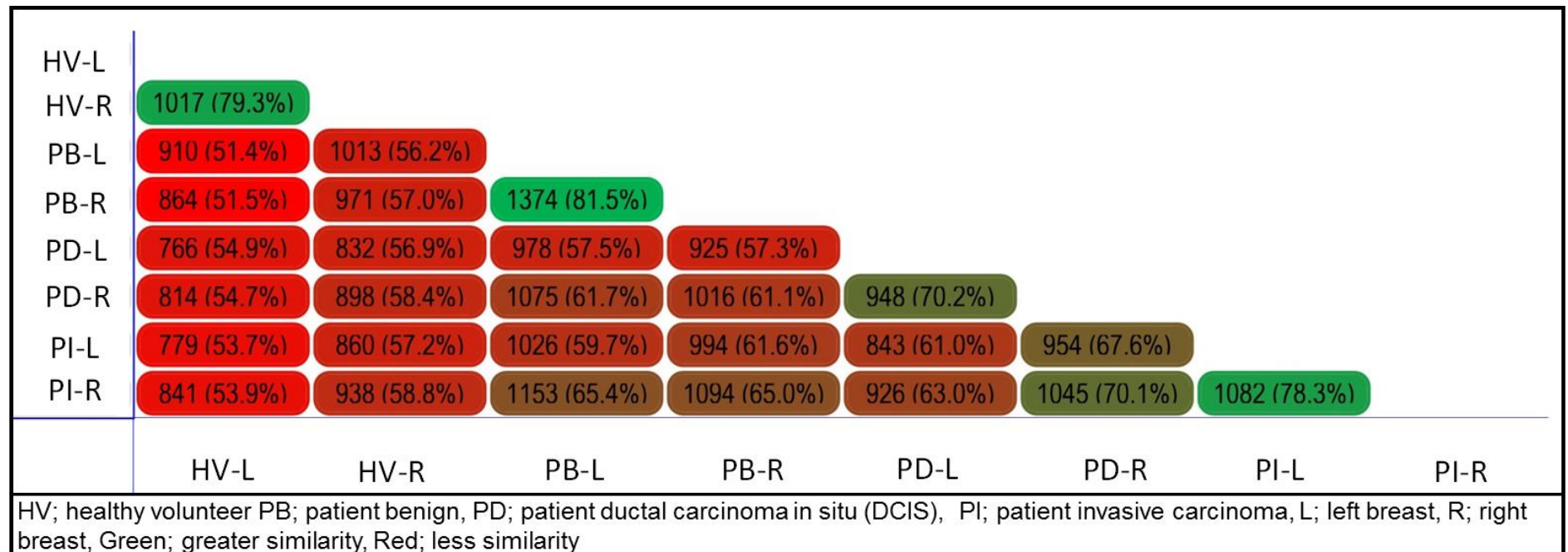


Figure 4-3: Comparison of NAF proteome acquired by 2D-LC/MS analysis.

NAF profiles for matched pairs from each individual, showed the greatest similarity with 1017 out of a total of 1282 proteins, 1374/1685, 948/1350 and 1082/1382 common for HV, PB, PD and PI pairs, respectively (Figure 4-4, A, B, C, D).

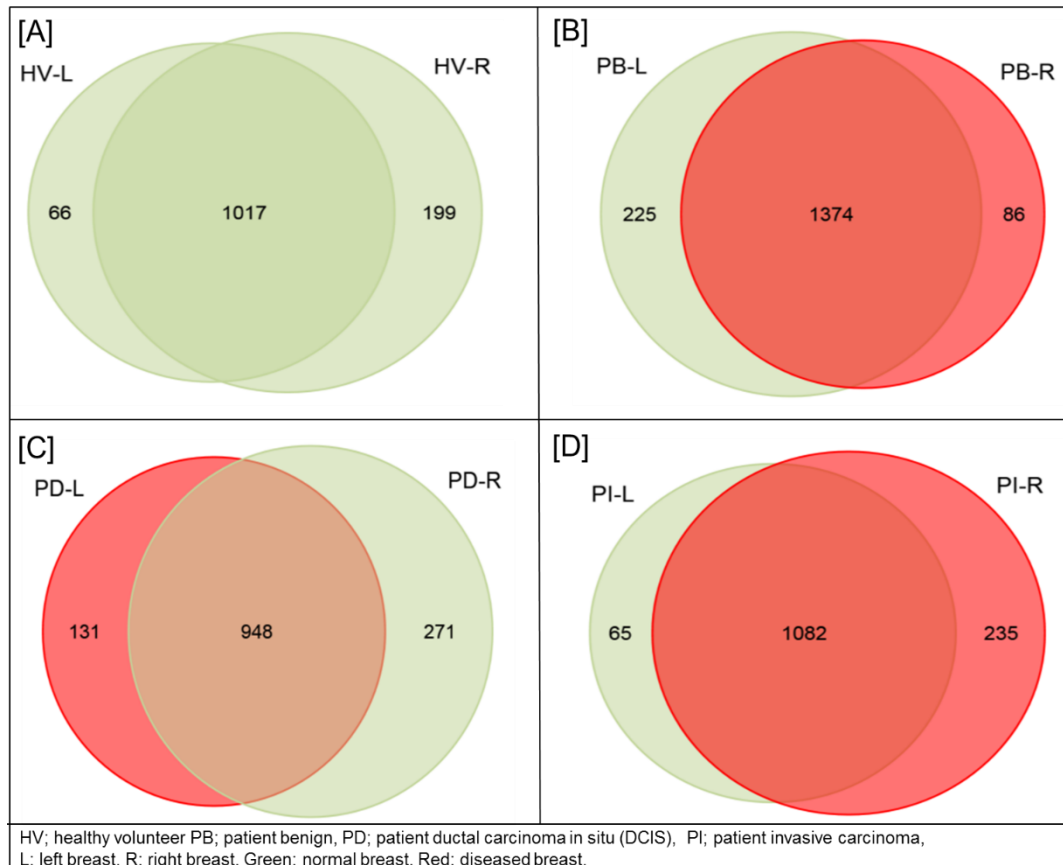


Figure 4-4: Comparison of the matched 2D NAF protein profiles.

When taking into account the quantitative data in terms of normalised peak areas for those proteins common to all the profiles (567), the natural pairs again showed greatest positive correlation (Pearson correlation coefficient values of 0.90 to 0.99) (Figure 4-5).

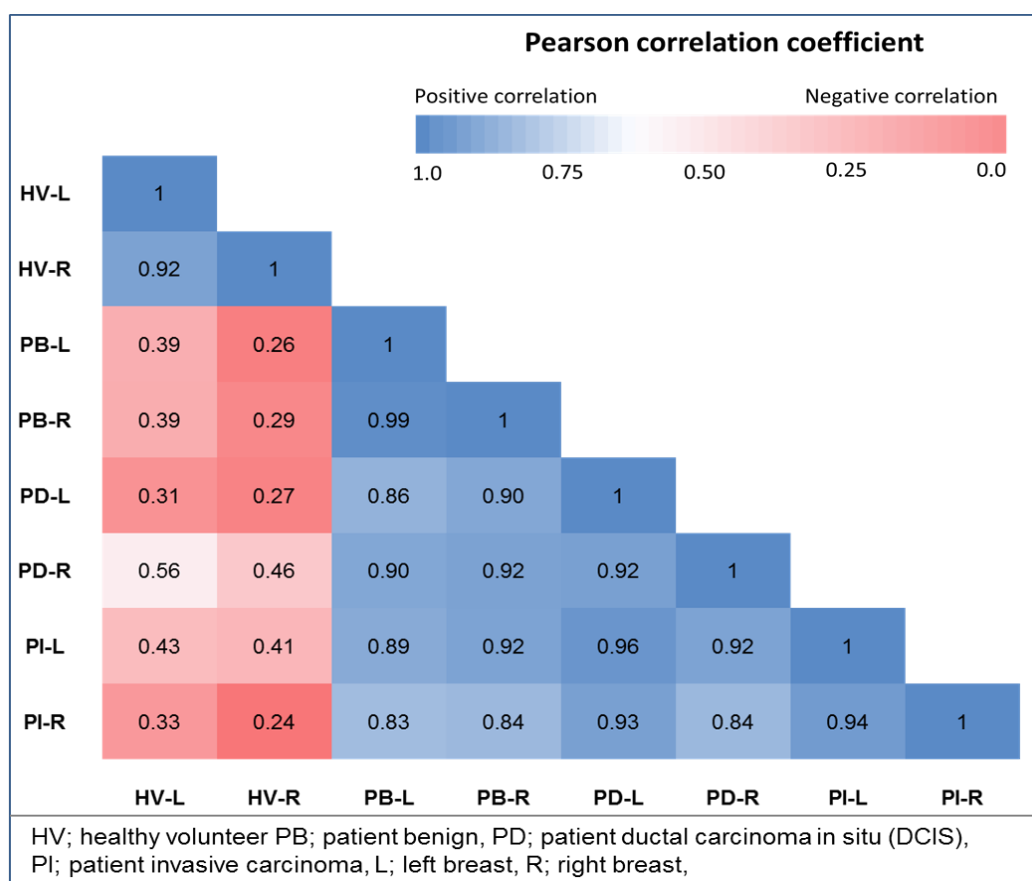


Figure 4-5: Pearson correlation coefficient for each NAF profile compared to each other, based on median normalised peak area of each protein.

4.3.4 Cellular localisation of NAF Proteome

The proteins common to all eight 2D LC analyses were manually categorised based on their normal cellular location using Uniprot (release 1026_07), into (i) intracellular, (ii) extracellular stroma (surrounding the breast cells) or (iii) plasma (Supplementary Figure 3). Intracellular proteins were further sub-divided into cytoplasmic, membrane-linked (i.e. containing transmembrane domain, GPI or lipid anchors), or organelle-specific location. Of the common proteins, 25% are normally found in plasma, 14% function within the extracellular space and 61% are from cell components. Of the 346 cellular proteins, 45% are normally found in the cytoplasm, 36% are membrane-associated, 7% lysosome, 8% endoplasmic reticulum/Golgi

apparatus, 2% mitochondrion and 2% in the nucleus. The nuclear and mitochondrial proteins were particularly under-represented, which have been estimated to represent 14% and 6% of the total human proteome respectively (Fink et al., 2008, Calvo et al., 2016). Submission of the total proteome profile to TMHMM transmembrane protein search engine identified 415 proteins with transmembrane regions (Figure 4-6) and a further 147 proteins with GPI-anchors, N-terminal or cys-modified lipid attachments were identified from UniProt.

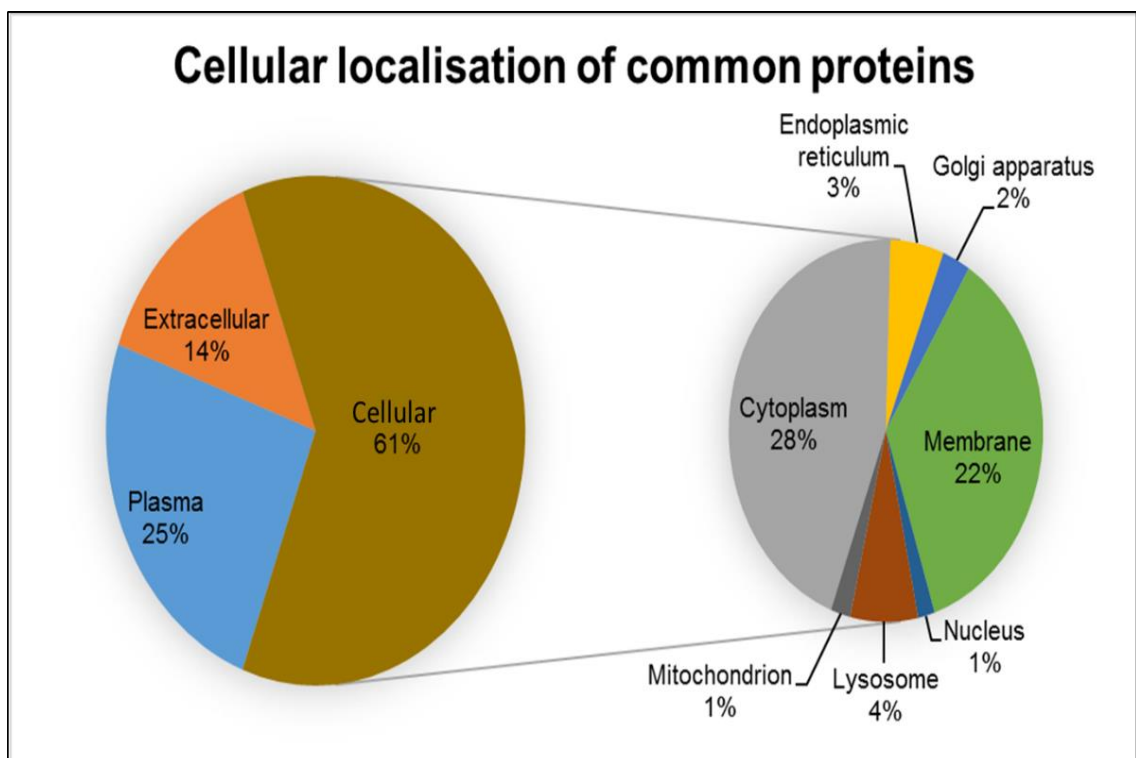


Figure 4-6: Normal cellular localisation of the 567 proteins common to all the NAF samples.

Therefore, 27% of the NAF proteins identified are normally found permanently anchored in membranes and provide a valuable source of information regarding cellular integrity. Using STRING analysis, 74 of the membrane proteins were linked to cell adhesion (FDR 2.88e-24), 50 proteins

involved in glycoprotein metabolism (FDR 2.88e-24) and 86 proteins have receptor activity (FDR 1.04e-21).

4.3.5 Functional annotation of NAF Proteome

Functional annotation of the proteins identified in the total NAF complement was performed using the DAVID bioinformatics resource. The application of the functional annotation clustering tool, which reduces redundancy in annotation of closely related biological functions across different ontological fields, identified secreted glycoproteins (enrichment score 101.14), cell-cell adhesion (41.97), antigen binding and immune response (24.46), lysosome (23.48), ribosomal protein interactions (14.93) and peptidase/protease activity (13.64).

A diverse range of cell adhesion proteins (193 components, Bonferroni p value 8.55E⁻²⁵) were also detected in the NAF samples including CEACAMs 1, 6 and 8, NCAM2, BCAM, ALCAM, ECAM, MCAM and ICAM1, 14 cadherin and protocadherin proteins and 5 integrin subunits (Table 4-2). Furthermore, 24 protein kinases (in addition to those categorised as receptors) and 18 protein phosphatases were identified which play a role in signal transduction. Stem cell (CD44) and epithelial-mesenchymal transition markers (cadherin 1, fibronectin, vimentin, cytokeratin 8 and cytokeratin 18), myoepithelial markers (cytokeratin 14 and cytokeratin 17) were also detected in all NAF samples.

Table 4-2: List of cell adhesion (GO:0007155; GO Biological Process) proteins identified in the NAF samples.

Table represents Gene ID, Protein name, UniProt accession number, Mascot score, Sequence coverage of protein, number of spectra identified in MS analysis, number of unique peptides, Frequency of detection in samples.

Gene ID	Protein Name	UniProt Accession	Spectra	Coverage	Unique Peptides	Score Mascot	Frequency
ACTB	Actin, cytoplasmic 1	P60709	3351	58.9	10	48916	8
ACTN1	Alpha-actinin-1	P12814	1957	60.8	31	32234	8
ADAM15	Disintegrin and metalloproteinase domain-containing protein 15	Q13444	55	11.9	6	630	7
ADAM9	Disintegrin and metalloproteinase domain-containing protein 9	Q13443	17	9.9	5	152	5
ALCAM	CD166 antigen	Q13740	391	41.5	17	4716	8
AMBP	Protein AMBP	P02760	449	40.6	10	6588	8
ANTXR1	Anthrax toxin receptor 1	Q9H6X2	9	1.6	1	107	4
ANXA1	Annexin A1	P04083	362	49.4	16	9026	8
ANXA2	Annexin A2	P07355	1250	64.3	22	21847	8
APOA4	Apolipoprotein A-IV	P06727	578	50.3	18	6813	8
APP	Amyloid beta A4 protein	P05067	419	25.6	12	2776	8
ARF6	ADP-ribosylation factor 6	P62330	107	41.7	6	1506	8
ATP1B1	Sodium/potassium-transporting ATPase subunit beta-1	P05026	3	3.6	1	38	0
AZGP1	Zinc-alpha-2-glycoprotein	P25311	42340	74.2	44	746399	8
B4GALT1	Beta-1,4-galactosyltransferase 1	P15291	1478	43	10	20579	8
BAIAP2	Brain-specific angiogenesis inhibitor 1-associated protein 2	Q9UQB8	258	31	12	1968	8
BAIAP2L1	Brain-specific angiogenesis inhibitor 1-associated protein 2-like protein 1	Q9UHR4	26	11	3	150	3
BCAM	Basal cell adhesion molecule	P50895	67	18.3	8	926	7
CADM4	Cell adhesion molecule 4	Q8NFZ8	395	40.2	10	4722	8
CD22	B-cell receptor CD22	P20273	45	1.4	1	87	6
CD2AP	CD2-associated protein	Q9Y5K6	10	2.8	1	39	1
CD36	Platelet glycoprotein 4	P16671	752	27.5	11	15576	8
CD44	CD44 antigen	P16070	161	4.6	4	3385	8
CD47	Leukocyte surface antigen CD47	Q08722	15	2.8	1	51	5
CD58	Lymphocyte function-associated antigen 3	P19256	23	10.8	3	258	4
CD63	CD63 antigen	P08962	35	13.4	2	441	7
CD9	CD9 antigen	P21926	707	20.6	4	8242	8
CDC42	Cell division control protein 42 homolog	P60953	306	51.3	7	2652	8
CDH1	Cadherin-1	P12830	590	22.6	14	10468	8
CEACAM1	Carcinoembryonic antigen-related cell adhesion molecule 1	P13688	271	15.8	4	3613	8
CEACAM5	Carcinoembryonic antigen-related cell adhesion molecule 5	P06731	104	10	2	1328	0
CELSR2	Cadherin EGF LAG seven-pass G-type receptor 2	Q9HCU4	69	2.4	4	310	7
CIB1	Calcium and integrin-binding protein 1	Q99828	314	51.3	9	4534	8
CLCA2	Calcium-activated chloride channel regulator 2	Q9UQC9	212	19.3	12	2570	6
CLIC1	Chloride intracellular channel protein 1	O00299	320	42.3	7	4249	8
CLSTN1	Calsyntenin-1	O94985	29	8.1	5	356	2
CLSTN3	Calsyntenin-3	Q9BQT9	17	3.1	2	157	4
CNTN1	Contactin-1	Q12860	49	9.2	6	409	6
CNTNAP3	Contactin-associated protein-like 3	Q9BZ76	13	2.8	3	97	2
COL12A1	Collagen alpha-1(XII) chain	Q99715	27	4.5	7	219	4
COL14A1	Collagen alpha-1(XIV) chain	Q05707	13	4.8	5	211	3
COL18A1	Collagen alpha-1(XVIII) chain	P39060	38	4.5	6	225	4
COL6A1	Collagen alpha-1(VI) chain	P12109	1251	32.7	22	15932	8
COL6A2	Collagen alpha-2(VI) chain	P12110	341	18.4	14	3170	8
COL6A3	Collagen alpha-3(VI) chain	P12111	11	0.4	1	164	3
CORO1A	Coronin-1A	P31146	14	13.2	3	130	3
CRISP2	Cysteine-rich secretory protein 2	P16562	8	7.8	1	75	1
CSRP1	Cysteine and glycine-rich protein 1	P21291	39	25.9	4	474	7
CSTA	Cystatin-A	P01040	45	52	3	476	7
CTNNA1	Catenin alpha-1	P35221	39	12.9	7	466	5

Gene ID	Protein Name	UniProt Accession	Spectra	Coverage	Unique Peptides	Score Mascot	Frequency
CTTN	Src substrate cortactin	Q14247	7	5.1	2	73	2
CX3CL1	Fractalkine	P78423	72	4.5	1	439	4
DAG1	Dystroglycan	Q14118	480	23.5	12	3883	7
DCHS1	Protocadherin-16	Q96JQ0	5	0.4	1	61	2
DCHS2	Protocadherin-23	Q6V1P9	38	2.9	5	230	2
DDR1	Epithelial discoidin domain-containing receptor 1	Q08345	267	13.1	9	3444	8
DPP4	Dipeptidyl peptidase 4	P27487	21	5.1	4	124	4
DSC2	Desmocollin-2	Q02487	406	10.3	9	4006	8
DSC3	Desmocollin-3	Q14574	18	7	4	57	5
DSG1	Desmoglein-1	Q02413	5	1.5	1	82	1
DSG2	Desmoglein-2	Q14126	38	9.3	7	227	6
DSP	Desmoplakin	P15924	11	2.2	4	144	2
DST	Dystonin	Q03001	12	0.1	1	130	7
EFNA1	Ephrin-A1	P20827	140	8.8	1	1131	8
EGFR	Epidermal growth factor receptor	P00533	18	3.1	3	132	4
ENG	Endoglin	P17813	196	18.2	9	2738	8
ENTPD1	Ectonucleoside triphosphate diphosphohydrolase 1	P49961	39	12.2	5	328	6
EPCAM	Epithelial cell adhesion molecule	P16422	8	22	4	83	2
EPHB3	Ephrin type-B receptor 3	P54753	21	3.6	4	313	4
EPHB4	Ephrin type-B receptor 4	P54760	38	5.5	4	188	5
EZR	Ezrin	P15311	1588	38.7	15	19419	8
F11R	Junctional adhesion molecule A	Q9Y624	58	22.7	4	885	8
FAT2	Protocadherin Fat 2	Q9NYQ8	549	22	57	4655	7
FBLN5	Fibulin-5	Q9UBX5	114	14.1	5	1130	8
FBLN7	Fibulin-7	Q53RD9	14	5.2	2	95	4
FERMT3	Fermitin family homolog 3	Q86UX7	20	7.6	3	173	4
FGA	Fibrinogen alpha chain	P02671	97	13.4	9	811	8
FGB	Fibrinogen beta chain	P02675	628	46.6	15	6466	8
FGG	Fibrinogen gamma chain	P02679	503	53	16	5667	8
FLNA	Filamin-A	P21333	269	14.3	22	3296	8
FLOT2	Flotillin-2	Q14254	72	28	9	930	8
FN1	Fibronectin	P02751	402	24.1	32	5270	8
FZD7	Frizzled-7	O75084	44	5.2	2	631	8
GAS6	Growth arrest-specific protein 6	Q14393	3	1.1	1	47	1
GNAS	Guanine nucleotide-binding protein G(s) subunit alpha isoforms XLas	Q5JWF2	300	19.9	14	2299	8
GPNMB	Transmembrane glycoprotein NMB	Q14956	315	4.7	2	8738	8
GPR56	G-protein coupled receptor 56	Q9Y653	54	6.5	3	650	4
GPR98	G-protein coupled receptor 98	Q8WYG9	4	0.5	2	53	1
HABP2	Hyaluronan-binding protein 2	Q14520	6	5.5	2	83	2
HAPLN3	Hyaluronan and proteoglycan link protein 3	Q96S86	46	21.4	6	322	2
HBB	Hemoglobin subunit beta	P68871	952	88.4	7	11220	5
HSPB1	Heat shock protein beta-1	P04792	164	40	5	1587	8
ICAM1	Intercellular adhesion molecule 1	P05362	337	32.3	12	5942	8
IGFALS	Insulin-like growth factor-binding protein complex acid labile subunit	P35858	110	17.2	8	1287	6
IGFBP7	Insulin-like growth factor-binding protein 7	Q16270	2	7.1	1	44	1
ISLR	Immunoglobulin superfamily containing leucine-rich repeat protein	O14498	377	27.1	8	4596	8
ITGAM	Integrin alpha-M	P11215	49	4	3	793	5
ITGAV	Integrin alpha-V	P06756	9	5.1	3	40	1
ITGAX	Integrin alpha-X	P20702	25	7.2	4	211	2
ITGB1	Integrin beta-1	P05556	7	4.6	2	32	2
ITGB2	Integrin beta-2	P05107	49	12.7	7	715	7
JUP	Junction plakoglobin	P14923	45	12.6	3	249	5
KIF14	Kinesin-like protein KIF14	Q15058	3	2.3	3	37	1
KIT	Mast/stem cell growth factor receptor Kit	P10721	39	4.6	4	301	4
KRT18	Keratin, type I cytoskeletal 18	P05783	412	33.3	9	3448	8
LAMA3	Laminin subunit alpha-3	Q16787	8	0.7	1	46	2
LAMA5	Laminin subunit alpha-5	O15230	129	7.1	17	671	7
LAMB2	Laminin subunit beta-2	P55268	172	13.3	15	1647	8
LAMC1	Laminin subunit gamma-1	P11047	39	4.7	5	1221	7
LAMC2	Laminin subunit gamma-2	Q13753	35	7.2	5	310	6
LGALS3BP	Galectin-3-binding protein	Q08380	6960	55.4	25	112765	8
Gene ID	Protein Name	UniProt Accession	Spectra	Coverage	Unique Peptides	Score Mascot	Frequency

MCAM	Cell surface glycoprotein MUC18	P43121	16	1.5	1	128	2
MFGE8	Lactadherin	Q08431	3120	74.9	24	44479	8
MPDZ	Multiple PDZ domain protein	O75970	5	3.5	4	67	1
MSN	Moesin	P26038	1139	43.3	13	14451	8
MUC16	Mucin-16	Q8WXI7	3150	11.2	56	31323	8
MUC4	Mucin-4	Q99102	799	19.4	23	12749	6
MYH10	Myosin-10	P35580	374	7	2	2702	1
MYH9	Myosin-9	P35579	1231	33.3	42	12532	8
NCAM2	Neural cell adhesion molecule 2	O15394	84	12.2	6	1174	6
NEO1	Neogenin	Q92859	22	4.2	4	167	5
NME1	Nucleoside diphosphate kinase B	P22392	60	27	1	249	3
NOV	Protein NOV homolog	P48745	38	14.3	3	426	7
NPNT	Nephronectin	Q6UXI9	114	19.5	8	1114	8
NPTN	Neuroplastin	Q9Y639	2	4.5	1	29	1
NRXN3	Neurexin-3	Q9Y4C0	24	3.2	3	90	4
NT5E	5'-nucleotidase	P21589	245	40.4	16	3462	7
OLFM4	Olfactomedin-4	Q6UX06	3540	49.4	20	58285	8
PAK4	Serine/threonine-protein kinase PAK 4	O96013	2	2.5	1	43	1
PCDHA10	Protocadherin alpha-10	Q9Y5I2	31	6.3	1	489	2
PCDHA4	Protocadherin alpha-4	Q9UN74	21	6.9	3	245	2
PCDHB14	Protocadherin beta-14	Q9Y5E9	30	9	4	165	3
PCDHB4	Protocadherin beta-4	Q9Y5E5	34	5.3	1	452	1
PCDHB5	Protocadherin beta-5	Q9Y5E4	51	7.7	2	596	3
PCDHGA12	Protocadherin gamma-A12	O60330	4	1	1	57	3
PCDHGA3	Protocadherin gamma-A3	Q9Y5H0	2	1.3	1	29	1
PCDHGB7	Protocadherin gamma-B7	Q9Y5F8	17	6.9	4	110	2
PCDHGC3	Protocadherin gamma-C3	Q9UN70	21	6.9	4	369	2
PDLIM1	PDZ and LIM domain protein 1	O00151	13	4	1	90	4
PDLIM5	PDZ and LIM domain protein 5	Q96HC4	28	9.6	4	357	4
PLXNB2	Plexin-B2	O15031	93	7.2	10	764	8
PLXNB3	Plexin-B3	Q9ULL4	10	2.3	2	67	3
PLXNC1	Plexin-C1	O60486	9	3.6	4	91	2
PODXL	Podocalyxin	O00592	156	5.6	3	1364	8
PODXL2	Podocalyxin-like protein 2	Q9NZ53	22	10.1	3	127	3
POSTN	Periostin	Q15063	85	17.9	8	919	6
PPP1CA	Serine/threonine-protein phosphatase PP1-alpha catalytic subunit	P62136	120	26.7	1	1752	4
PTK7	Inactive tyrosine-protein kinase 7	Q13308	64	7.9	5	882	4
PTPRF	Receptor-type tyrosine-protein phosphatase F	P10586	240	16	18	2480	7
PTPRK	Receptor-type tyrosine-protein phosphatase kappa	Q15262	166	9.5	8	1739	7
PTPRS	Receptor-type tyrosine-protein phosphatase S	Q13332	58	6.2	6	516	7
PVR	Poliovirus receptor	P15151	64	12.9	4	465	6
PVRL1	Nectin-1	Q15223	49	16.6	5	245	5
PVRL2	Nectin-2	Q92692	82	10.6	4	918	6
PVRL4	Nectin-4	Q96NY8	54	9.4	4	360	7
RAB10	Ras-related protein Rab-10	P61026	194	22.5	2	2867	8
RAB1A	Ras-related protein Rab-1A	P62820	245	51.2	2	3099	4
RAC1	Ras-related C3 botulinum toxin substrate 1	P63000	177	39.6	5	1097	8
RAC2	Ras-related C3 botulinum toxin substrate 2	P15153	43	12.5	1	543	2
RAP2B	Ras-related protein Rap-2b	P61225	26	29	3	366	4
RGMB	RGM domain family member B	Q6NW40	20	4.3	1	85	6
RPSA	40S ribosomal protein SA	P08865	81	38.6	7	843	6
S100A11	Protein S100-A11	P31949	310	63.8	4	7761	8
S100A8	Protein S100-A8	P05109	827	45.2	6	8206	8
S100A9	Protein S100-A9	P06702	3300	75.4	8	54496	8
SHC1	SHC-transforming protein 1	P29353	6	2.1	1	24	3
SIRPA	Tyrosine-protein phosphatase non-receptor type substrate 1	P78324	470	28.8	9	4641	8
SLURP1	Secreted Ly-6/uPAR-related protein 1	P55000	3	24.3	1	47	1
SPP1	Osteopontin	P10451	982	55.1	13	11049	3
STXBP1	Syntaxin-binding protein 1	P61764	14	7.2	3	62	3
TGFB2	Transforming growth factor beta-2	P61812	12	12.8	3	71	3
TGFBI	Transforming growth factor-beta-induced protein igh3	Q15582	327	30.7	13	2474	8
THBS1	Thrombospondin-1	P07996	1294	40.9	31	17622	8

Gene ID	Protein Name	UniProt Accession	Spectra	Coverage	Unique Peptides	Score Mascot	Frequency
THBS3	Thrombospondin-3	P49746	6	2.8	2	55	2
THY1	Thy-1 membrane glycoprotein	P04216	14	24.2	3	93	4
TINAGL1	Tubulointerstitial nephritis antigen-like	Q9GZM7	31	13.9	5	212	5
TLN1	Talin-1	Q9Y490	60	3.6	6	860	7
TM9SF4	Transmembrane 9 superfamily member 4	Q92544	6	3	1	44	3
TMEM8A	Transmembrane protein 8A	Q9HCN3	23	7.8	5	181	3
TNC	Tenascin	P24821	3757	57.7	81	56044	8
TOR1A	Torsin-1A	O14656	17	6.9	2	112	6
TSTA3	GDP-L-fucose synthase	Q13630	9	11.5	2	43	5
VCAN	Versican core protein	P13611	158	3.4	7	1918	6
VCL	Vinculin	P18206	75	15.3	11	945	8
VTN	Vitronectin	P04004	491	26.4	9	8572	8
VWA2	von Willebrand factor A domain-containing protein 2	Q5GFL6	6	3.7	2	72	1
VWF	von Willebrand factor	P04275	89	7.2	15	783	2

4.3.6 Expression of cancer related proteins

The normalised peak area mean for each protein for each matched pair was calculated to establish significant expression changes between individuals. Whilst many proteins detected are case/patient specific, there were some important trends observed that reflect changes in disease-related physiology. The proteomic profiles from the healthy volunteer (Case 1) had distinctly different protein compositions from the other 3 cases. A Student t-test was applied to identify proteins that were significantly different compared to the other three cases. A comparison of non-cancer (healthy/benign) to malignancies (DCIS/invasive carcinoma), identified 225 proteins (Supplementary Table 4.3) that were differentially expressed ($p < 0.05$). Of these 76 are normally found in the extracellular space, including basement membrane and extracellular matrix proteins (STRING, 28 proteins, FDR – $6.68e-14$, DAVID, 33 proteins, $2.4e-19$, Benjamini corrected), focal adhesion (DAVID, 35 proteins, $8.8e-18$, Benjamini corrected) and those involved in wound healing (STRING 40 proteins, FDR – $4.48e-15$) cell-cell adhesion junction (DAVID, 18 proteins, $7.7e-6$, Benjamini corrected).

In this study, 15 growth factors were also identified including insulin-like growth factor I and II (IGF1 and IGF2), interleukin-8, 9 and 34, transforming growth factor beta 1, 2, and 3, and vascular endothelial growth factor A (VEGFA) (Table 4-3). Although some are case-specific, amongst the receptors detected were those associated with tumour growth, including epidermal growth factor receptor (EGFR), transforming growth factor beta receptor type II and III (TGF β R 2 and 3), hepatocyte growth factor receptor (HGFR), vascular endothelial growth factor receptor (VEGFR1), mast/stem cell growth factor receptor (KIT) and receptor tyrosine-protein kinase ErbB-3 (HER3) (Table 4-4). Furthermore, a number of cancer associated ligands were also detected including pro-epidermal growth factor (EGF), platelet-derived growth factors C and D (PDGF-C and PDGF-D), placenta growth factor, neuropilin-1 and pleiotropin (Table 4-4).

Accession	GeneID	Description	MW [kDa]	Score Mascot	Coverage	# Spectra	# Unique Peptides	Frequency	Comments	Detected in plasma
P05019	IGF1	Insulin-like growth factor I	21.8	35	5.1	4	1	2	IC pair only	ELISA
P01344	IGF2	Insulin-like growth factor II	20.1	31	5.0	3	1	2		ELISA
P10145	IL8	Interleukin-8	11.1	123	16.2	13	1	2	PD only	CIA
Q9UHD0	IL19	Interleukin-19	20.4	155	22.0	39	3	2	HV only	MS
Q6ZMJ4	IL34	Interleukin-34	27.5	358	6.6	26	1	8		not detected
Q969H8	MYDGF	Myeloid-derived growth factor	18.8	177	15.6	14	2	4		MS
P49763	PGF	Placenta growth factor	24.8	82	19.0	30	2	6	all except HV	ELISA
Q9NRA1	PDGFC	Platelet-derived growth factor C	39.0	1960	15.7	117	5	8		not detected
Q9GZP0	PDGFD	Platelet-derived growth factor D	42.8	968	10.8	51	3	8		not detected
P21246	PTN	Pleiotrophin	18.9	1536	23.8	92	3	8		MS
P01133	EGF	Pro-epidermal growth factor	133.9	17275	28.4	1201	25	8		MIA
P01137	TGFB1	Transforming growth factor beta-1	44.3	41	3.3	2	1	2		ELISA
P61812	TGFB2	Transforming growth factor beta-2	47.7	71	12.8	12	3	4		not detected
P10600	TGFB3	Transforming growth factor beta-3	47.3	39	3.6	2	1	1		MS-MRM
P15692	VEGFA	Vascular endothelial growth factor A	27.0	69	8.2	32	1	6		ELISA

Table represents UniProt accession number, Gene ID, Protein name, Molecular weight of protein, Mascot score, Sequence coverage of protein, number of spectra identified in MS analysis, number of unique peptides, Frequency of detection in samples, detected in plasma by mass spectrometry (MS), Enzyme-linked immunosorbent assay (ELISA), multiple reaction monitoring mass spectrometry (MRM-MS), Immunoassay (IA).

Table 4-3: Selected mitogenic factors identified in NAF

Accession	Gene ID	Description	MW [kDa]	Score Mascot	Coverage	# Spectra	# Unique Peptides	Frequency	Comments	Detected in plasma
P20273	CD22	B-cell receptor CD22	95.3	87	1.4	45	1	6	all except HV	MS
Q16620	NTRK2	BDNF/NT-3 growth factors receptor	91.9	988	3.9	54	3	8		MS
P00533	EGFR	Epidermal growth factor receptor	134.2	132	3.1	18	3	4	PB and PI pairs	IA
P08581	MET	Hepatocyte growth factor receptor	155.4	54	2.5	10	2	2		ELISA
P17181	IFNAR1	Interferon alpha/beta receptor 1	63.5	305	2.3	17	1	5		MS
P14778	IL1R1	Interleukin-1 receptor type 1	65.4	405	9.0	53	3	6	all except HV	MS
P78552	IL13RA1	Interleukin-13 receptor subunit alpha-1	48.7	278	5.2	18	2	4		MS
P40189	IL6ST	Interleukin-6 receptor subunit beta	103.5	4320	12.0	307	8	8		ELISA
P16871	IL7R	Interleukin-7 receptor subunit alpha	51.5	260	7.0	23	3	2	HV only	MS
P42702	LIFR	Leukemia inhibitory factor receptor	123.7	1751	13.6	176	12	6	all except HV	IA
P01130	LDLR	Low-density lipoprotein receptor	95.3	331	6.5	23	3	6		MS
P10721	KIT	Mast/stem cell growth factor receptor Kit	109.8	301	4.6	39	4	4	HV and PI pairs	ELISA
O00264	PGRMC1	Membrane-associated progesterone receptor component 1	21.7	67	7.7	8	2	3		MS
O14786	NRP1	Neuropilin-1	103.1	2910	23.7	240	15	7		MS - MRM
P21860	ERBB3	Receptor tyrosine-protein kinase erbB-3	148.0	164	1.8	15	1	5		MS
P37173	TGFBR2	TGF-beta receptor type-2	64.5	49	2.6	24	1	7		MS
Q03167	TGFBR3	Transforming growth factor beta receptor type 3	93.4	279	9.4	26	5	2		MS
P17948	FLT1	Vascular endothelial growth factor receptor 1	150.7	295	0.8	58	1	6	all except HV	ELISA

Table represents UniProt accession number, Gene ID, Protein name, Molecular weight of protein, Mascot score, Sequence coverage of protein, number of spectra identified in MS analysis, number of unique peptides, Frequency of detection in samples, detected in plasma by mass spectrometry (MS), Enzyme-linked immunosorbent assay (ELISA), multiple reaction monitoring mass spectrometry (MRM-MS), Immunoassay (IA).

Table 4-4: Selected mitogenic receptors identified in NAF.

The IGF signalling pathway plays an important role in regulating cell proliferation and inhibiting apoptosis in breast and prostate cancer. Expression of IGF1 and insulin-like growth factors-binding proteins 3 (IGFBP3) may be related to mammographic density, a high-risk factor for breast cancer as it correlates with the volume of ductal/lobule anatomy and consequentially the epithelial cell content which in turn is connected with proliferation and possibly mutagenesis [24]. Circulating IGF1 has been shown to be positively associated with ER-positive breast cancer risk [25]. In this study, IGF1 was only detected in the invasive carcinoma patient, however, IGFBP3 was detected in all samples and exhibited increasing levels with disease stage. Basic fibroblast growth factor, a marker of angiogenesis in tumour growth and metastatic breast cancer, has been measured in NAF by immunoassay [26], but was not detected in our samples. However, three other proteins (MMP9, collagen alpha-2(IV) and VEGFR1) associated with angiogenesis, were detected.

A subgroup of extracellular matrix proteins exhibited a variation on the above profile with a decrease in DCIS (PD) expression compared to benign (PB) but with highest levels in invasive carcinoma (PI) - glypicans-1 (GPC1) and -4 (GPC4), and syndecan-1 (SDC1), and junctional adhesion molecule A (F11R). The degradation of the ductal lining, is most likely due to the action of the large number of proteases present in NAF. We have previously observed that laminin beta2 decreases in invasive carcinoma tissue (Shaheed et al., 2013b), most likely due to proteolytic degradation and I was now able to see these products secreted in NAF.

Mannello *et al* identified the importance of exploring NAF for established biomarkers such as urokinase-dependent plasminogen activator

(uPA) and plasminogen activator inhibitor (PAI-I), particularly for their role in extracellular matrix turnover, which might be expected from processes associated with cancer invasiveness (Mannello and Ligi, 2013). Analysis of NAF identified nearly 233 proteolytic enzymes and inhibitors (1.94 fold enriched, p value $58E^{-18}$), many of which are naturally secreted by cells to modulate stromal composition. Within this group were 10 members of the kallikrein family, including prostate specific antigen (PSA/KLK3) (Table 4-5). An inverse correlation of KLK3 levels in NAF with breast cancer stage (DCIS to metastatic) has previously been described (Alexander et al., 2004, Sauter et al., 2004b). KLK3 was detected in 3 NAF samples (Table 4-5), most significantly representing a decrease in DCIS compared to the matched normal. Of the other kallikreins, KLK5, KLK6, KLK8, KLK11 were detected in all NAF samples.

There were 6 proteins (kallikrein 6, ATP-binding cassette sub-family C member 11, secretoglobin family 3A member 1, mammaglobin-A, prolactin-inducible protein [PIP], and mucin-like protein 1) strongly associated with breast tissue expression (compared to all other tissues). The expression of PIP, was also verified by western blot in NAF samples (Figure 4-8). Although these proteins are not cancer-specific, changes in these proteins may prove useful indicators of breast health.

Table 4-5: List of proteolysis proteins identified in the NAF samples.

Table represents Gene ID, Protein name, UniProt accession number, Mascot score, Sequence coverage, protein, number of spectra identified in MS analysis, number of unique peptides, Frequency of detection in samples.

Gene ID	Protein Name	UniProt Accession	# Spectra	Coverage	# Unique Peptides	Score Mascot	Frequency
ABHD5	1-acylglycerol-3-phosphate O-acyltransferase ABHD5	Q8WTS1	45	14.3	3	721	3
ACE	Angiotensin-converting enzyme	P12821	1813	33.1	34	22210	8
ACE2	Angiotensin-converting enzyme 2	Q9BYF1	74	22.0	12	690	6
ADAM10	Disintegrin and metalloproteinase domain-containing protein 10	O14672	4	2.3	1	83	1
ADAM15	Disintegrin and metalloproteinase domain-containing protein 15	Q13444	55	11.9	6	630	7
ADAM9	Disintegrin and metalloproteinase domain-containing protein 9	Q13443	17	9.9	5	152	5
ADAMDEC1	ADAM DEC1	O15204	8	10.2	3	92	1
AGA	N(4)-(beta-N-acetylglucosaminy)-L-asparaginase	P20933	335	34.7	5	8616	8
AGT	Angiotensinogen	P01019	734	31.5	10	14508	8
ANPEP	Aminopeptidase N	P15144	4833	48.6	41	80648	8
APEH	Acylamino-acid-releasing enzyme	P13798	13	3.1	2	166	4
APOH	Beta-2-glycoprotein 1	P02749	1245	53.3	12	14435	8
ARRB1	Beta-arrestin-1	P49407	13	7.9	2	31	3
ASPRV1	Retroviral-like aspartic protease 1	Q53RT3	4	3.8	1	62	2
ATP6AP2	Renin receptor	O75787	88	26.9	6	1159	5
AZU1	Azurocidin	P20160	271	48.6	7	3268	5
BLMH	Bleomycin hydrolase	Q13867	27	7.9	2	207	4
BMP1	Bone morphogenetic protein 1	P13497	5	5.2	3	118	2
C1QA	Complement C1q subcomponent subunit A	P02745	2	9.4	1	24	1
C1QB	Complement C1q subcomponent subunit B	P02746	4	5.5	1	73	1
C1QC	Complement C1q subcomponent subunit C	P02747	17	8.6	2	77	3
C1R	Complement C1r subcomponent	P00736	335	39.0	16	4114	8
C1RL	Complement C1r subcomponent-like protein	Q9NZP8	287	16.8	4	2220	8
C1S	Complement C1s subcomponent	P09871	198	28.9	12	4227	8
C2	Complement C2	P06681	167	18.9	11	1666	8
C3	Complement C3	P01024	18775	77.6	113	297198	8
C4A	Complement C4-A	POC0L4	32891	65.6	5	556429	8
C4B	Complement C4-B	POC0L5	33808	65.9	6	572440	8
CAPN1	Calpain-1 catalytic subunit	P07384	43	11.5	7	394	7
CAPN2	Calpain-2 catalytic subunit	P17655	31	8.0	4	136	5
CAPN7	Calpain-7	Q9Y6W3	5	3.3	2	29	3
CAPNS1	Calpain small subunit 1	P04632	24	28.7	4	223	4
CASP14	Caspase-14	P31944	178	39.3	10	2406	7
CD2AP	CD2-associated protein	Q9Y5K6	10	2.8	1	39	1
CD46	Membrane cofactor protein	P15529	22	3.8	1	32	7
CFB	Complement factor B	P00751	1977	46.1	27	20896	8
CFD	Complement factor D	P00746	572	53.0	10	5978	8
CFI	Complement factor I	P05156	900	40.5	19	17877	8
CHMP4B	Charged multivesicular body protein 4b	Q9H444	34	25.4	3	636	8
CLCA2	Calcium-activated chloride channel regulator 2	Q9UQC9	212	19.3	12	2570	6
CLCA4	Calcium-activated chloride channel regulator 4	Q14CN2	360	35.5	20	4336	8
CLN5	Ceroid-lipofuscinosis neuronal protein 5	O75503	60	24.0	7	457	7
CNDP2	Cytosolic non-specific dipeptidase	Q96KP4	161	56.2	16	1574	8
COPS4	COP9 signalosome complex subunit 4	Q9BT78	3	3.4	1	34	1
CORIN	Atrial natriuretic peptide-converting enzyme	Q9Y5Q5	7	2.1	1	67	1
CPA1	Carboxypeptidase A1	P15085	52	5.3	1	679	0
CPA4	Carboxypeptidase A4	Q9UI42	15	5.0	2	67	4
CPB1	Carboxypeptidase B	P15086	2045	78.9	24	30392	8
CPD	Carboxypeptidase D	O75976	65	3.6	4	1644	7
CPE	Carboxypeptidase E	P16870	128	21.0	6	2328	8
CPM	Carboxypeptidase M	P14384	111	16.5	5	1376	8
CPNE1	Copine-1	Q99829	2	1.7	1	47	1
CPQ	Carboxypeptidase Q	Q9Y646	122	19.3	6	2421	8

Gene ID	Protein Name	UniProt Accession	# Spectra	Coverage	# Unique Peptides	Score Mascot	Frequency
CSNK1A1	Casein kinase I isoform alpha	P48729	7	3.0	1	62	3
CTNNB1	Catenin beta-1	P35222	31	6.4	1	150	2
CTSA	Lysosomal protective protein	P10619	14	7.9	3	200	5
CTSB	Cathepsin B	P07858	594	38.1	10	9400	8
CTSC	Dipeptidyl peptidase 1	P53634	383	22.0	7	4139	8
CTSD	Cathepsin D	P07339	1842	48.1	16	31145	8
CTSF	Cathepsin F	Q9UBX1	6	6.4	2	38	2
CTSG	Cathepsin G	P08311	167	39.6	7	1646	6
CTSH	Pro-cathepsin H	P09668	60	9.0	3	459	8
CTSL	Cathepsin L1	P07711	10	8.4	2	44	3
CTSO	Cathepsin O	P43234	11	13.1	3	81	4
CTSS	Cathepsin S	P25774	137	30.8	6	2146	7
CTSZ	Cathepsin Z	Q9UBR2	206	38.0	7	1272	8
CUL3	Cullin-3	Q13618	4	4.9	2	32	2
CUL4B	Cullin-4B	Q13620	9	0.9	1	82	2
DAG1	Dystroglycan	Q14118	480	23.5	12	3883	7
DCD	Dermcidin	P81605	36	12.7	2	370	5
DDB1	DNA damage-binding protein 1	Q16531	13	4.3	3	48	6
DHCR24	Delta(24)-sterol reductase	Q15392	22	7.4	3	227	2
DNAJB2	DnaJ homolog subfamily B member 2	P25686	2	4.6	1	37	1
DNAJC3	DnaJ homolog subfamily C member 3	Q13217	28	11.5	4	220	5
DNPEP	Aspartyl aminopeptidase	Q9ULA0	14	5.1	2	43	2
DPP3	Dipeptidyl peptidase 3	Q9NY33	21	7.6	4	165	5
DPP4	Dipeptidyl peptidase 4	P27487	21	5.1	4	124	4
DPP7	Dipeptidyl peptidase 2	Q9UHL4	161	26.4	10	1944	7
ELANE	Neutrophil elastase	P08246	313	36.7	6	4135	5
ERAP1	Endoplasmic reticulum aminopeptidase 1	Q9NZ08	662	32.8	25	7162	8
ERAP2	Endoplasmic reticulum aminopeptidase 2	Q6P179	17	2.9	2	97	2
F12	Coagulation factor XII	P00748	95	15.6	6	653	7
F2	Prothrombin	P00734	650	43.6	21	7797	8
F9	Coagulation factor IX	P00740	3	2.0	1	47	2
FAM125A	Multivesicular body subunit 12A	Q96EY5	101	34.1	6	1960	8
FCN2	Ficolin-2	Q15485	34	14.7	3	352	6
FGA	Fibrinogen alpha chain	P02671	97	13.4	9	811	8
FGB	Fibrinogen beta chain	P02675	628	46.6	15	6466	8
FGG	Fibrinogen gamma chain	P02679	503	53.0	16	5667	8
FOLH1	Glutamate carboxypeptidase 2	Q04609	56	13.1	6	690	4
FURIN	Furin	P09958	19	9.9	4	80	3
GAS6	Growth arrest-specific protein 6	Q14393	3	1.1	1	47	1
GCA	Grancalcin	P28676	4	4.6	1	42	1
GGH	Gamma-glutamyl hydrolase	Q92820	291	35.2	8	2786	8
GGT1	Gamma-glutamyltranspeptidase 1	P19440	1852	34.8	11	33170	8
GGT5	Gamma-glutamyltransferase 5	P36269	591	33.8	13	7608	8
GGT6	Gamma-glutamyltransferase 6	Q6P531	47	16.0	4	519	4
GGTLC3	Putative gamma-glutamyltransferase light chain 3	B5MD39	231	19.1	2	4964	1
GTSE1	G2 and S phase-expressed protein 1	Q9NYZ3	6	1.8	1	78	4
HABP2	Hyaluronan-binding protein 2	Q14520	6	5.5	2	83	2
HP	Haptoglobin	P00738	1339	56.7	11	17693	8
HPN	Serine protease hepsin	P05981	54	13.2	3	374	6
HPR	Haptoglobin-related protein	P00739	457	27.9	2	7609	4
HSP90B1	Endoplasmic	P14625	243	18.8	11	3487	8
HSPA5	78 kDa glucose-regulated protein	P11021	440	39.4	20	5711	8
HTRA1	Serine protease HTRA1	Q92743	32	4.0	2	594	5
ITCH	E3 ubiquitin-protein ligase Itchy homolog	Q96J02	9	7.9	4	107	1
KIF14	Kinesin-like protein KIF14	Q15058	3	2.3	3	37	1
KLK10	Kallikrein-10	Q43240	71	14.1	3	364	7
KLK11	Kallikrein-11	Q9UBX7	394	46.8	9	3633	8
KLK13	Kallikrein-13	Q9UKR3	12	4.0	1	32	5
KLK14	Kallikrein-14	Q9P0G3	129	50.9	9	2439	5
KLK2	Kallikrein-2	P20151	3	6.5	1	45	1
KLK3	Prostate-specific antigen	P07288	294	55.6	8	2457	3
KLK5	Kallikrein-5	Q9Y337	164	27.3	5	1503	8
KLK6	Kallikrein-6	Q92876	558	48.4	8	5116	8
KLK7	Kallikrein-7	P49862	7	9.9	1	36	3
KLK8	Kallikrein-8	O60259	136	20.8	4	1029	8
KLKB1	Plasma kallikrein	P03952	8	4.2	2	146	2
LAP3	Cytosol aminopeptidase	P28838	48	16.8	6	328	7
LGMN	Legumain	Q99538	263	41.6	11	2064	8
LNPEP	Leucyl-cystinyl aminopeptidase	Q9UIQ6	2	1.3	1	31	1
LONP2	Lon protease homolog 2, peroxisomal	Q86WA8	17	8.8	4	78	2

Gene ID	Protein Name	UniProt Accession	# Spectra	Coverage	# Unique Peptides	Score Mascot	Frequency
LTF	Lactotransferrin	P02788	76302	94.9	106	1379153	8
MAN1B1	Endoplasmic reticulum mannosyl-oligosaccharide 1,2-alpha-mannosidase	Q9UKM7	42	11.7	6	192	5
MBTPS1	Membrane-bound transcription factor site-1 protease	Q14703	100	3.8	2	705	7
MME	Neprilysin	P08473	182	31.2	17	2828	3
MMP7	Matrilysin	P09237	203	34.1	6	1606	8
MMP9	Matrix metalloproteinase-9	P14780	32	9.5	5	503	4
MYH9	Myosin-9	P35579	1231	33.3	42	12532	8
NCSTN	Nicastrin	Q92542	205	18.3	7	2392	8
NPEPPS	Puromycin-sensitive aminopeptidase	P55786	52	7.6	6	997	7
OS9	Protein OS-9	Q13438	11	10.9	5	115	2
OTUB1	Ubiquitin thioesterase OTUB1	Q96FW1	26	18.1	4	107	6
PAPPA2	Pappalysin-2	Q9BXP8	13	4.2	4	43	1
PARK7	Protein deglycase DJ-1	Q99497	171	64.6	8	2073	8
PCBP2	Poly(rC)-binding protein 2	Q15366	4	4.4	1	25	3
PCOLCE	Procollagen C-endopeptidase enhancer 1	Q15113	35	12.2	4	172	8
PCYOX1	Prenylcysteine oxidase 1	Q9UHG3	22	11.9	4	476	4
PDCD6	Programmed cell death protein 6	O75340	76	40.8	5	981	8
PDCD6IP	Programmed cell death 6-interacting protein	Q8WUM4	1354	50.9	33	20305	8
PDIA3	Protein disulfide-isomerase A3	P30101	277	34.9	13	2729	8
PEF1	Peflin	Q9UBV8	6	8.5	2	119	2
PEPD	Xaa-Pro dipeptidase	P12955	64	22.7	6	1243	5
PGC	Gastricsin	P20142	163	5.2	2	2189	6
PIP	Prolactin-inducible protein	P12273	56864	77.4	21	122387	8
PLAT	Tissue-type plasminogen activator	P00750	1229	58.0	23	15190	8
PLAU	Urokinase-type plasminogen activator	P00749	8	4.4	2	30	3
PLG	Plasminogen	P00747	927	49.0	26	8068	8
PROS1	Vitamin K-dependent protein S	P07225	65	12.6	7	321	8
PRSS1	Trypsin-1	P07477	7	8.1	1	56	3
PRSS22	Brain-specific serine protease 4	Q9GZN4	66	18.6	4	414	7
PRSS23	Serine protease 23	Q95084	20	10.2	3	248	4
PRSS8	Prostasin	Q16651	1340	41.4	7	18478	8
PRTN3	Myeloblastin	P24158	356	27.7	4	4200	7
PSMD1	26S proteasome non-ATPase regulatory subunit 1	Q99460	12	1.3	1	161	5
PSMD2	26S proteasome non-ATPase regulatory subunit 2	Q13200	12	1.7	1	37	2
PSMD3	26S proteasome non-ATPase regulatory subunit 3	O43242	5	2.4	1	28	2
PSMD5	26S proteasome non-ATPase regulatory subunit 5	Q16401	10	9.7	3	89	3
PSMD6	26S proteasome non-ATPase regulatory subunit 6	Q15008	9	9.3	3	71	2
PSME1	Proteasome activator complex subunit 1	Q06323	16	24.1	4	303	5
PSME2	Proteasome activator complex subunit 2	Q9UL46	3	5.9	1	111	1
RAD23B	UV excision repair protein RAD23 homolog B	P54727	9	2.7	1	69	2
RFFL	E3 ubiquitin-protein ligase rifylin	Q8WZ73	4	5.8	1	40	1
RHOA	Transforming protein RhoA	P61586	364	65.3	2	3306	8
RNPEP	Aminopeptidase B	Q9H4A4	170	32.6	13	1500	8
RPS27A	Ubiquitin-40S ribosomal protein S27a	P62979	537	30.1	4	6930	8
RUVBL1	RuvB-like 1	Q9Y265	3	3.5	1	45	2
SCPEP1	Retinoid-inducible serine carboxypeptidase	Q9HB40	158	21.0	7	1582	8
SFRP1	Secreted frizzled-related protein 1	Q8N474	485	42.7	9	3313	8
SHMT1	Serine hydroxymethyltransferase, cytosolic	P34896	25	15.1	5	360	2
SPPL2A	Signal peptide peptidase-like 2A	Q8TCT8	11	4.4	2	201	2
SRI	Sorcin	P30626	7	6.1	1	110	2
ST14	Suppressor of tumorigenicity 14 protein	Q9Y5Y6	130	16.1	8	1623	8
TBC1D10A	TBC1 domain family member 10A	Q9BXI6	6	10.6	3	130	1
TGFB1	Transforming growth factor beta-1	P01137	2	3.3	1	41	1
THSD4	Thrombospondin type-1 domain-containing protein 4	Q6ZMP0	3	1.1	1	61	2
TINAGL1	Tubulointerstitial nephritis antigen-like	Q9GZM7	31	13.9	5	212	5
TMEM59	Transmembrane protein 59	Q9BXS4	38	7.4	2	134	5
TMPRSS11E	Transmembrane protease serine 11E	Q9UL52	4	8.7	2	56	2
TMPRSS13	Transmembrane protease serine 13	Q9BYE2	52	12.1	6	450	4
TMPRSS2	Transmembrane protease serine 2	O15393	32	15.4	5	288	6

Gene ID	Protein Name	UniProt Accession	# Spectra	Coverage	# Unique Peptides	Score Mascot	Frequency
TOM1L1	TOM1-like protein 1	O75674	73	13.2	5	374	7
TOR1A	Torsin-1A	O14656	17	6.9	2	112	6
TPP1	Tripeptidyl-peptidase 1	O14773	185	30.9	9	2260	8
TPSAB1	Tryptase beta-2	P20231	190	22.9	5	1274	5
TRABD2B	Metalloprotease TIKI2	A6NFA1	16	5.2	2	113	2
TSG101	Tumor susceptibility gene 101 protein	Q99816	140	20.0	7	1419	8
UBA3	NEDD8-activating enzyme E1 catalytic subunit	Q8TBC4	3	3.2	1	82	3
UBE2L3	Ubiquitin-conjugating enzyme E2 L3	P68036	24	24.0	2	86	6
UBE2N	Ubiquitin-conjugating enzyme E2 N	P61088	51	34.2	4	132	7
UQCRC1	Cytochrome b-c1 complex subunit 1, mitochondrial	P31930	7	14.2	3	44	2
UQCRC2	Cytochrome b-c1 complex subunit 2, mitochondrial	P22695	10	11.0	3	90	3
USP5	Ubiquitin carboxyl-terminal hydrolase 5	P45974	4	1.5	1	28	2
VCP	Transitional endoplasmic reticulum ATPase	P55072	129	25.8	15	1336	8
VPS25	Vacuolar protein-sorting-associated protein 25	Q9BRG1	41	21.6	4	155	6
VPS28	Vacuolar protein sorting-associated protein 28 homologs	Q9UK41	126	37.6	9	1966	8
VPS36	Vacuolar protein-sorting-associated protein 36	Q86VN1	8	14.5	4	100	2
VPS4A	Vacuolar protein sorting-associated protein 4A	Q9UN37	50	14.9	3	377	3
VPS4B	Vacuolar protein sorting-associated protein 4B	O75351	106	21.8	5	735	7
WFDC2	WAP four-disulfide core domain protein 2	Q14508	1142	67.7	6	18605	8

4.3.7 Pathway analysis of differently expressed proteins in the NAF proteome.

The KEGG pathway (<http://www.genome.jp/kegg/>) analysis of NAF proteome identified 11 significant proteins (highlighted in red) which play a role in breast cancer proliferation, survival, and progression. These proteins control four important pathways in breast cancer; MAPK signalling pathway, PI3k-Akt signalling pathway, Notch signalling pathway and Wnt signalling pathway. The cellular interaction of these proteins is shown in Figure 4-7.

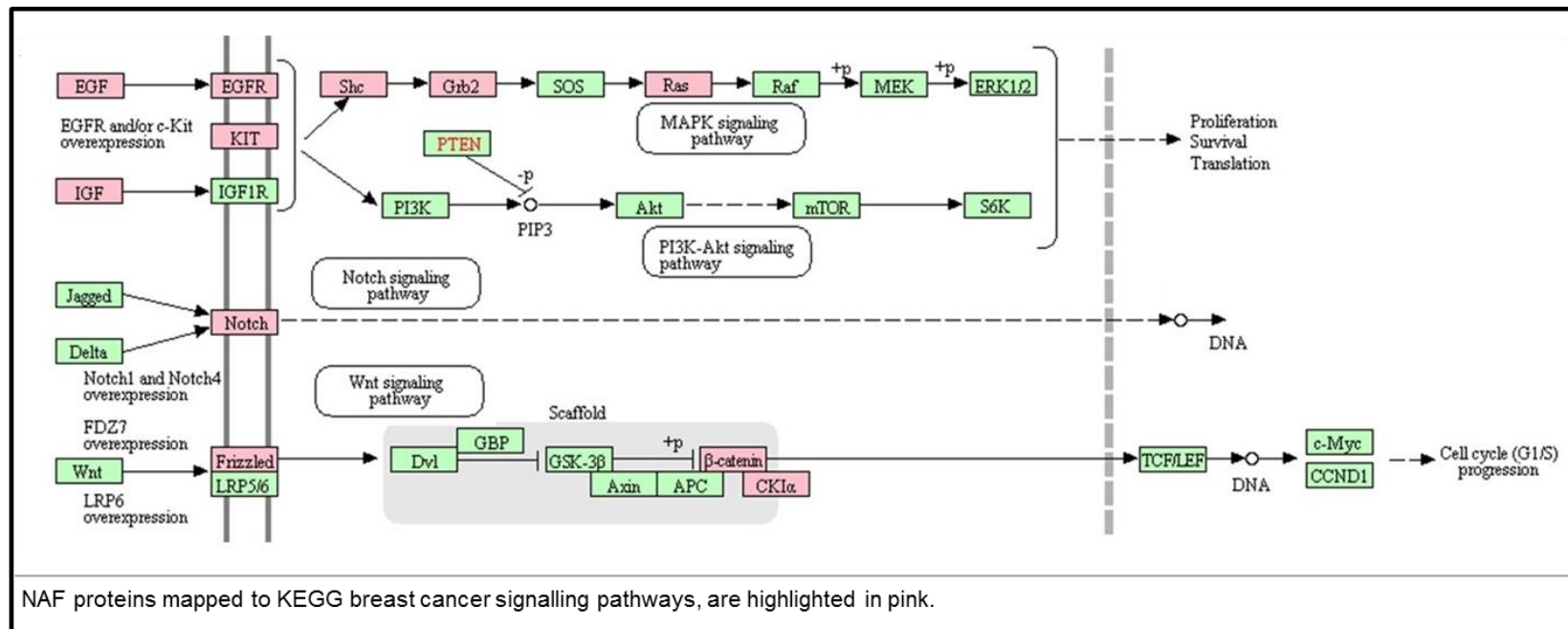


Figure 4-7: NAF proteins detected in breast cancer signalling pathways.

4.3.8 Verification of presence of selected targets by Western blot.

The expression of prolactin-inducible protein (PIP), CD44 antigen (CD44), vimentin (VIM), actin, cytoplasmic 1 (ACTB), cofilin-1 (CFL1), calreticulin (CALR), epidermal growth factor receptor (EGFR), and cytochrome P450 3A4 (CYP3A4), was verified in NAF by Western blotting. All immunoblots were analysed by GelAnalyzer 2010, to determine the MW of proteins (Figure 4-8). The target proteins; CD44, CALR, EGFR, ACTB and cofilin-1, were detected with single band at 81kDa, 48kDa, 44kDa, 42kDa and 18kDa respectively.

The theoretical mass of EGFR is 134kDa (Uniprot.org, 2015) but in NAF samples was detected at 44kDa. EGFR has 4 isoforms; P00533-1 (134kDa), P00533-2 (44kDa), P00533-3 (77kDa) and P00533-4 (69kDa) (Uniprot.org, 2015). On this basis, anti-EGFR antibody seems to be detecting isoform 2 at 44kDa, which can be further verified by GEL-LC-MS analysis of the 44kDa band (Shevchenko et al., 1996) VIM was detected with 2 bands at 53kDa and 51kDa while theoretical mass of VIM is 53.65kDa. PIP was present in all NAF samples at 16kDa (theoretical MW-16.57kDa) except those of the healthy volunteers, while a weak band at 14kDa was also detected in Patient with DCIS and patient with invasive carcinoma (Figure 4-8). Actin cytoplasmic 1 (ACTB) was used a control for western blot assay, but ACTB was observed at different level in NAF samples (Figure 4-8). There was also a difference for expression in left and right breasts of same patient. These findings suggest that ACTB is not the right control for analysis of NAF

samples by Western blot analysis. The limited volume of NAF also makes Western blot a less than ideal approach for validation of breast cancer biomarkers. Therefore, a multiplex assay like MRM-MS, will be ideal to validate and quantify potential biomarkers in NAF samples.

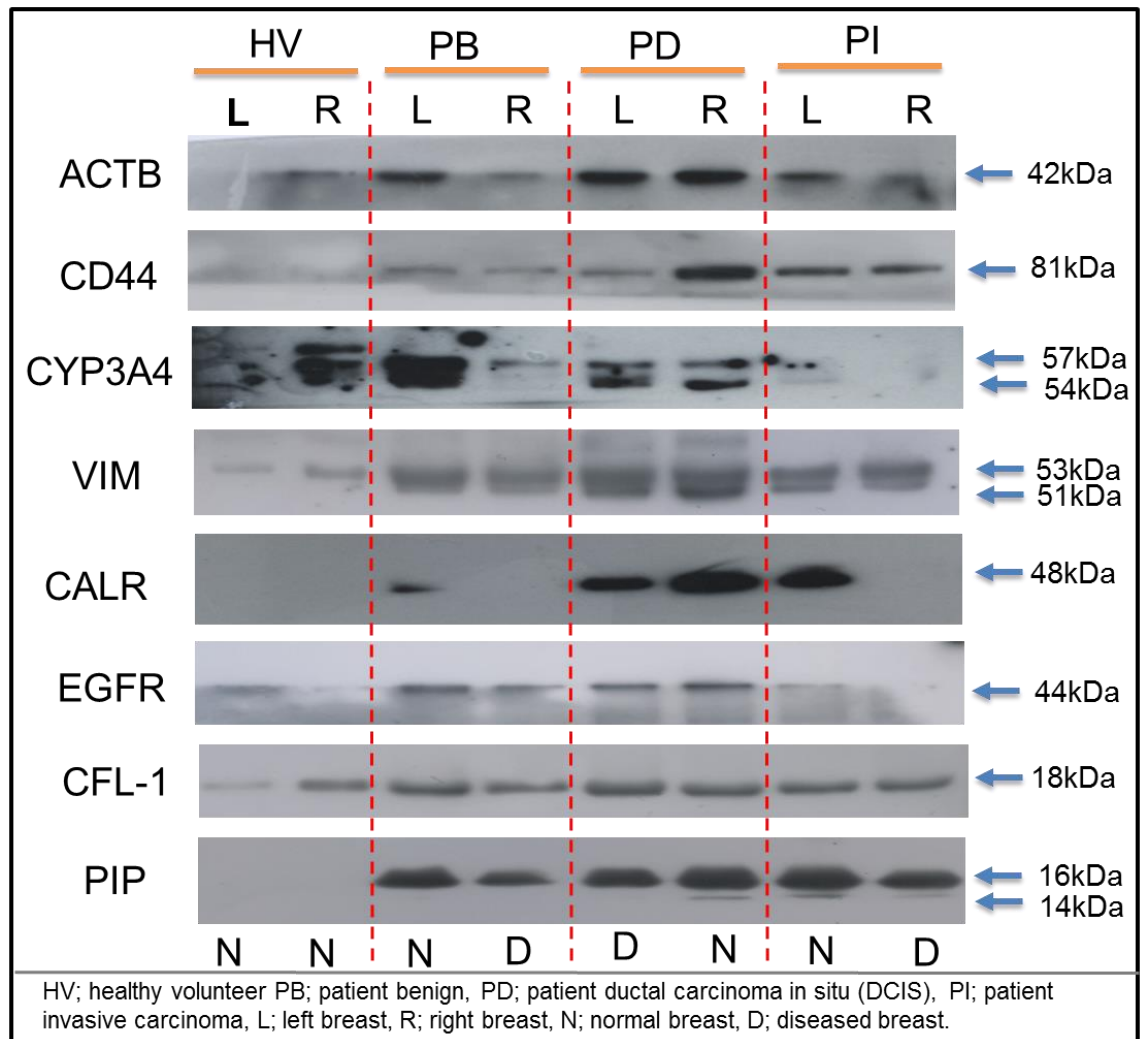


Figure 4-8: Western blot analysis of ACTB, CD44, CYP3A4, VIM, CALR, EGFR, CFL1 and PIP in NAF.

4.4 DISCUSSION

4.4.1 Comparison of identified NAF proteome with published NAF data.

Prior to this study, the most complete proteomics profile of NAF was that of Pavlou *et al* (Pavlou *et al.*, 2010). A total of 854 proteins were identified from 3 healthy volunteers and 3 cancer patients (2 samples from the cancerous breast and one from the non-cancer breast). Comparison with our dataset, based on gene identity (691 entries Pavlou *et al*, and 1919 for our set), indicated an overlap of 563 proteins Figure 4-9: Overlapping identities with the data from the NAF study by Pavlou *et al* 2010.(Figure 4-9), however our current study illustrated substantial progress in NAF characterisation identifying 1374 new proteins not previously seen in NAF.

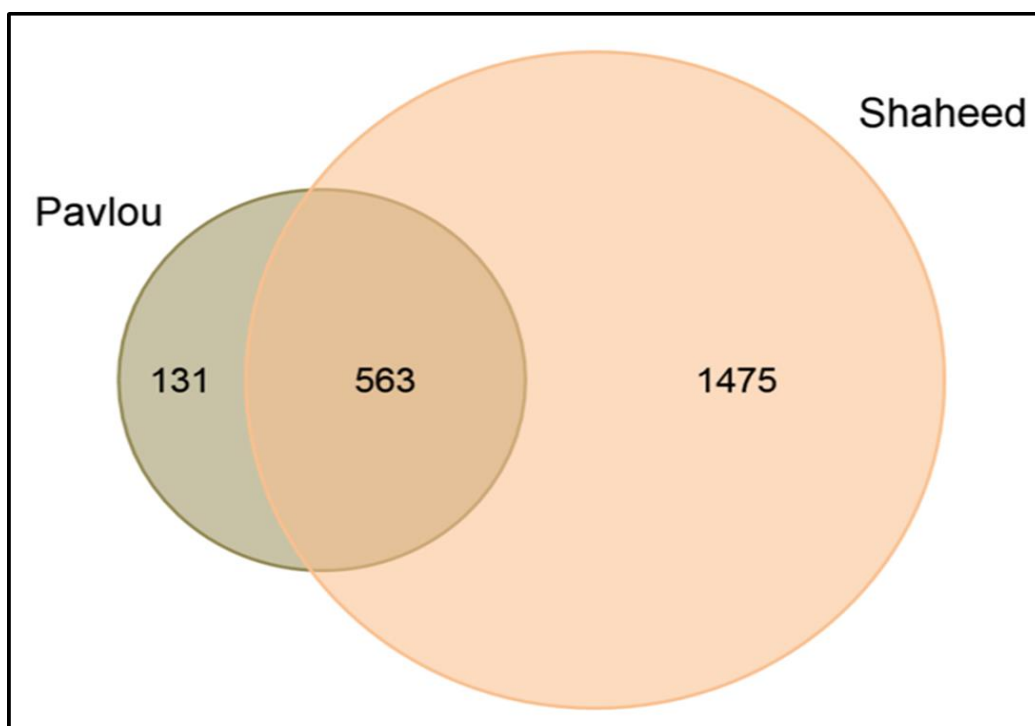


Figure 4-9: Overlapping identities with the data from the NAF study by Pavlou *et al* 2010.

4.4.2 Comparison of NAF proteome with Plasma proteome

As plasma is by far the most commonly used and most completely characterised liquid biopsy, including many breast cancer biomarker studies (Zeidan et al., 2015), I wanted to establish if the NAF proteome is likely to provide unique insights. The Plasma Proteome Database comprises the collated quantitative data for 10546 proteins that have been detected in plasma and serum (<http://www.plasmaproteomedatabase.org/>) (Omenn et al., 2005). A comparison of the NAF profile with the Plasma Proteome Database identified 1578 proteins in common (Figure 4-10), but 332 proteins (21% of the total NAF profile) were unique to NAF indicating it has great potential to provide molecular information specific to breast health.

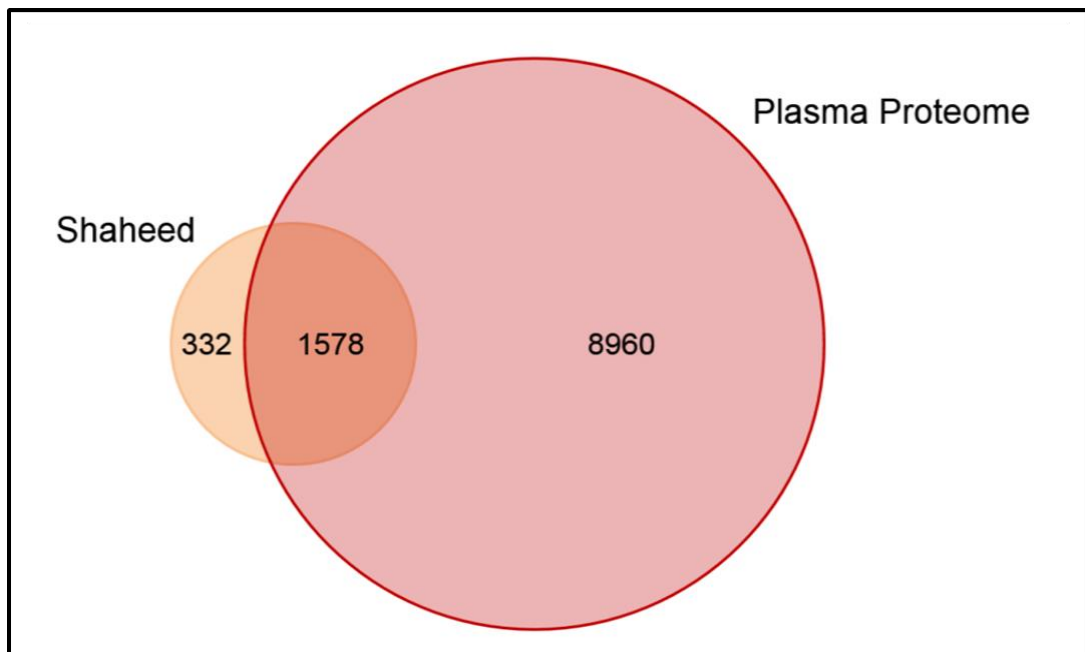


Figure 4-10: overlapping identities with the Human Plasma Proteome database.

In this study, the high plasma composition was interpreted as blood contamination of the tissue during surgery as high levels of haemoglobin isoforms were also detected. Interestingly, haemoglobin alpha and beta

isoforms were only detected in patient NAF samples, and with significant higher levels in the diseased breast compared to the normal.

4.4.3 Phenotype specific protein expression changes

A two-tailed Student t-test of HV, using the average normalised sum of the three strongest peak areas for each protein, from two breast analyses, compared to the equivalent data for the three cancer samples, identified 331 proteins that were present at significantly different levels ($p < 0.05$) (Supplementary Table 4.1). The proteomic profile of Case 1 (healthy volunteer, HV), indicated high levels of milk proteins (Figure 4-11). Of the 40 most abundant proteins observed by Beck et al, the most complete proteome study of human milk to date, identifying 1606 gene products, 25 were also the most abundant in the NAF samples of case 1 (Beck et al., 2015). Case 1 presented at the outpatient clinic with a spontaneous milky-pus discharge when either nipple was squeezed. The reproductive history of the volunteer, aged 48, indicated, she had achieved parity 3 times (with the first birth at age 38). Small amounts of milk or serous fluid expression can persist for months or years after weaning, but Case 1 did not engage in breastfeeding. Further investigation of her medical records, however, indicated that she had been prescribed amitriptyline, for depression and stress-management, and omeprazole and lansoprazole for gastric oesophageal reflux. In rare cases, these may cause breast tissue enlargement and nipple discharge, which is associated with galactorrhea rather than cancer (Pipaliya et al., 2016).

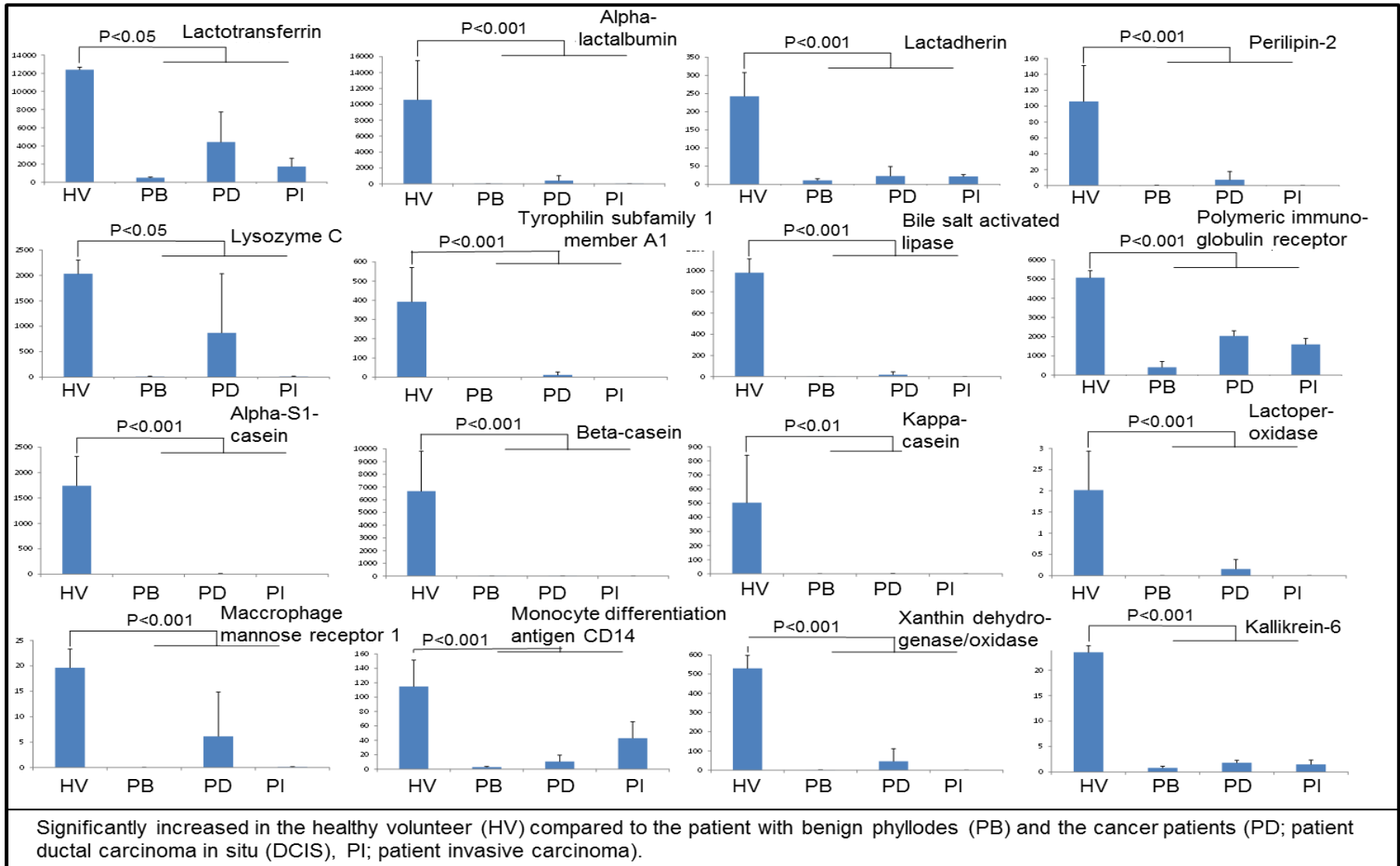


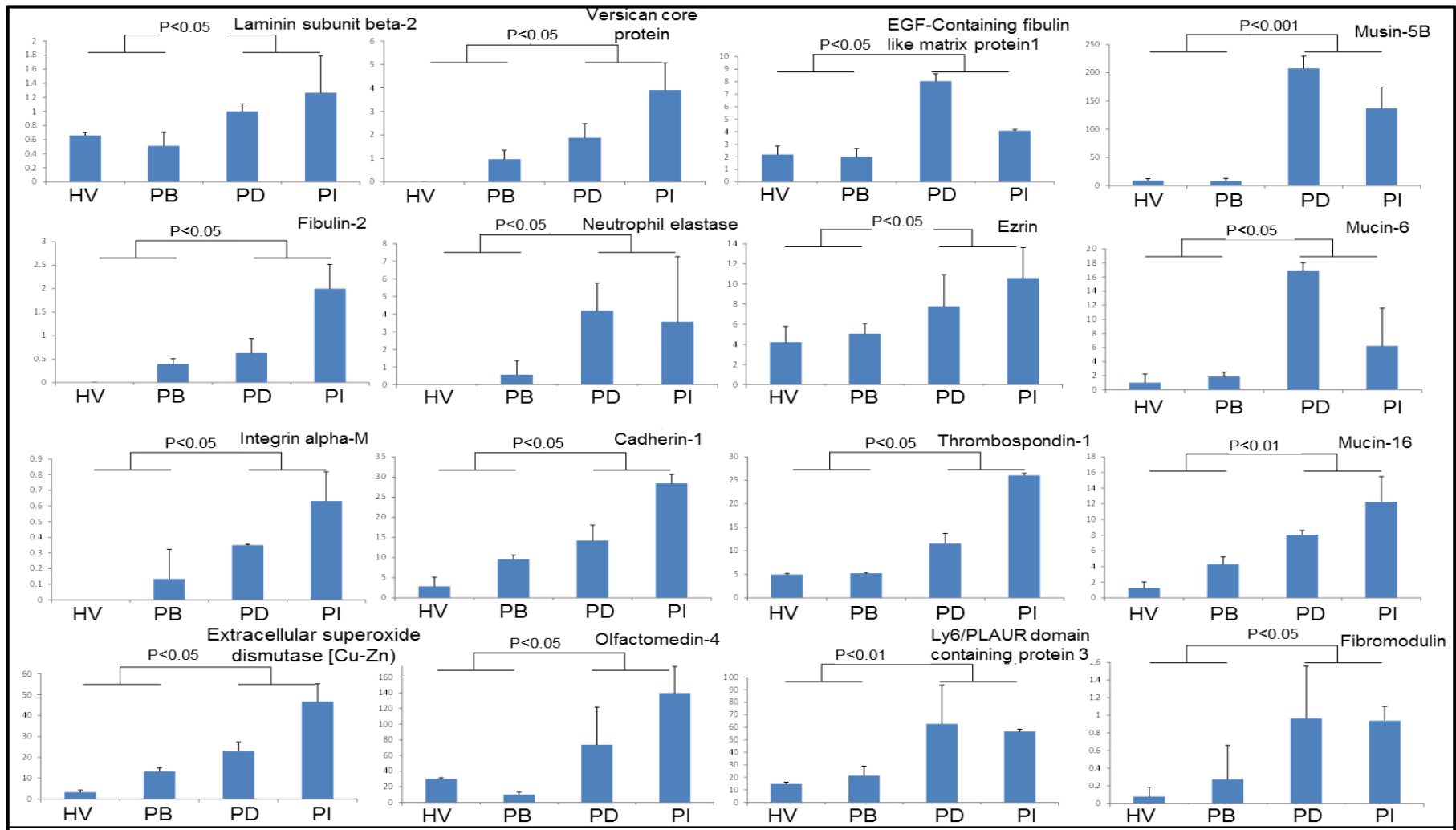
Figure 4-11: Significantly increased in the healthy volunteer (HV), compared to diseased samples.

Phyllodes tumours constitute less than 1% of all breast tumours and are characterised by atypical stromal histological features. The phyllodes patient (Case 9, PB) investigated here was a benign form (Figure 4-1). Of the 462 proteins (Supplementary Table 4.2) previously identified as benign phyllodes markers, predominantly through immunohistochemistry methods, EGFR (malignant phyllodes marker) (Tse et al., 2009), neprilysin (CD10) (Vilela et al., 2014), VEGF (Tse et al., 2004), alpha catenin and beta-catenin (Ho et al., 2013, Tsang et al., 2012, Lacroix-Triki et al., 2010) were most abundant in the NAF samples from Case 9 confirming previous observations. High expression of EGFR was also observed in patients with benign phyllodes and ductal carcinoma in situ (DCIS), by Western blot analysis (Figure 4-8). E-cadherin or cadherin 1 has also been linked to benign phyllodes, and was significantly high in the benign patient, but was also high in the invasive cancer patient. c-kit (CD117) another marker of benign phyllodes (Sawyer et al., 2003) was detected in NAF, but only in the healthy and invasive cases. Other important markers of benign phyllodes, p53 and Ki-67 (CD34) were not detected in NAF, but as they are found in the nucleus and nuclear component of NAF is under-represented compared to normal, this would be expected. Ubiquitin carboxy-terminal hydrolase L1 has been associated with the tumour development including benign phyllodes (Lien et al., 2013). Although not detected in this study, a related isoform Ubiquitin carboxy-terminal hydrolase 5, which has a role in p53 de-ubiquitination, was solely detected in the benign case.

4.4.4 Cancer specific protein expression changes

In this study, the proteome of health volunteer/benign was compared against DCIS/IC, to understand cancer specific protein expression changes. Laminins form a complex and constitute a key component of basement membranes and have an essential role the structure and function of extracellular matrix (Glentis et al., 2014). Each complex comprising a heterotrimer of subunits (alpha, beta and gamma) linked by disulphide bridges. Of the 5 laminin isoforms detected in NAF samples, the three most abundant were alpha5, beta2 and gamma1, which suggests the expression of the specific heterotrimer laminin-11 (or laminin-521) normally expressed/found in the glomerular basement membrane in the kidney, the neuromuscular synaptic cleft and placenta (Miner and Patton, 1999). Laminin alpha5, beta2 and gamma1, along with collagen alpha 1(VI), alpha 2(VI), fibulin(s), versican were increased in absolute levels from healthy volunteer to invasive carcinoma suggesting increasing disruption of the basement membrane with disease progression (Figure 4-12). Extracellular matrix proteins, including ICAM1 (which has been implicated in cell migration and metastasis), mucin-1 (MUC1), mucin-like protein-1 (MUCL1), E-cadherin (cadherin-1, CDH1), cell adhesion molecule 4 (CADM4), extracellular superoxide dismutase (SOD3), moesin (MSN), tenascin (TNC), thrombospondins-1 (THBS1) and -2 (THBS2), vitronectin (VTN), ezrin (EZR), olfactomedin-4 (OLFM4), extracellular matrix protein 1 (ECM1), EGF-containing fibulin-like extracellular matrix protein 1 (EFEMP1) and Ly6/PLAUR domain-containing protein 3 (LYPD3) also exhibited a similar

pattern of expression in the NAF samples. Proteins associated with inflammation, cell growth and cell migration were also significantly increased and included attractin, protein S100-A8 and A9, cathepsin D, CSF1, insulin-like growth factors-binding proteins 3 (IGFBP3), EGF and PDGF-C (Figure 4-13).



Significantly increased in the cancer patients (PD; patient ductal carcinoma in situ (DCIS), PI; patient invasive carcinoma) compared to healthy volunteer (HV) and patient with benign phyllodes (PB) .

Figure 4-12, A: Significantly increased in the cancer patients compares to healthy volunteer and patient with benign phyllodes.

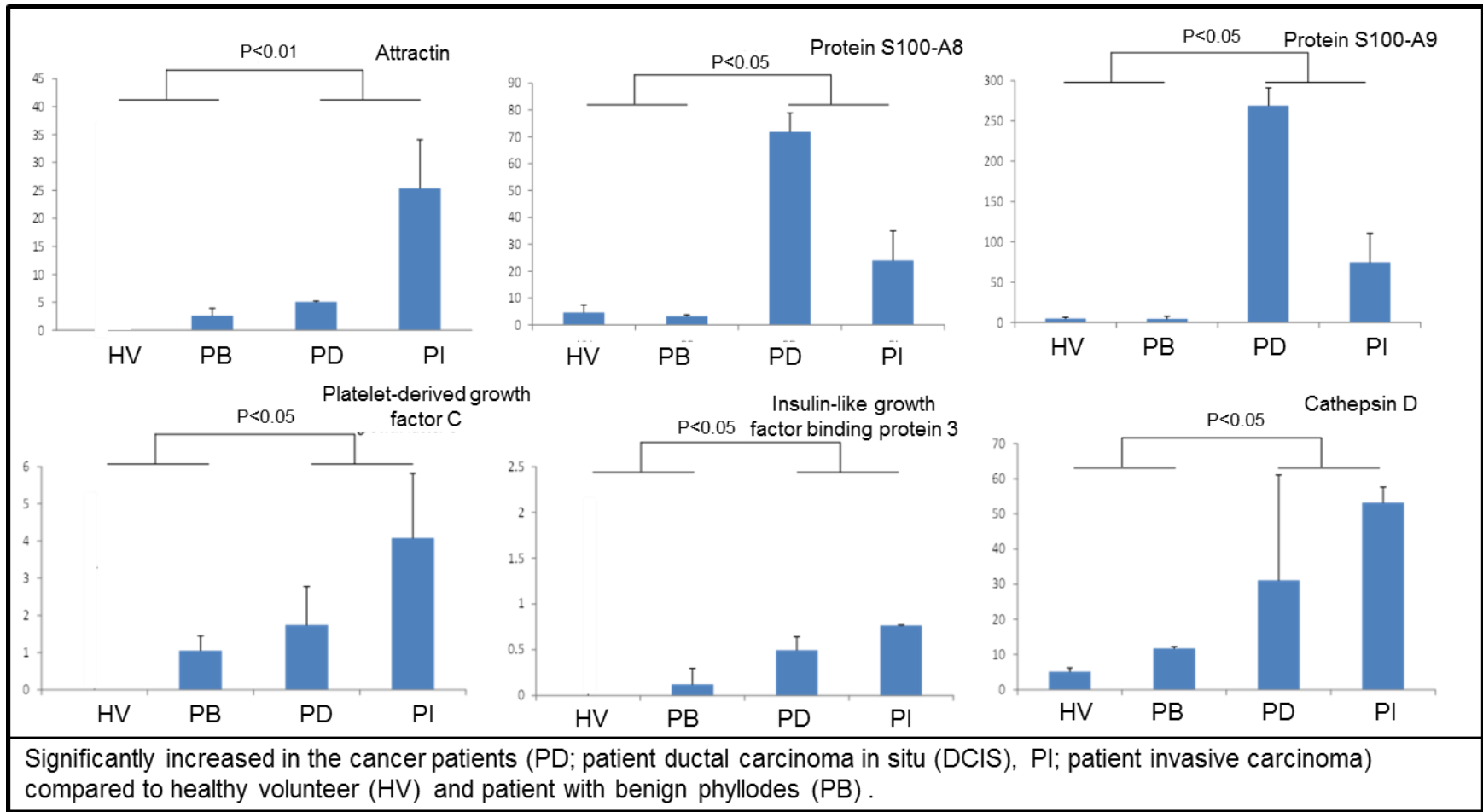


Figure 4-13 B: Significantly increased in the cancer patients compares to healthy volunteer and patient with benign phyllodes.

4.4.5 Diagnostic application

One of my objectives was to understand whether analysing matched pairs would provide a more specific approach to detecting disease compared to normal. SDS PAGE showed that most matched pairs have similar protein band patterns, which was corroborated by the high correlation of proteomics profiles. Whilst some proteins were significantly different in bilateral samples, a more extensive longitudinal study is required to determine statistically valid differences between disease and healthy breast of an individual.

The National Cancer Institute has coordinated the research of many institutions to accelerate the identification and validation of early stage cancer testing and detection (<https://edrn.nci.nih.gov/>). Of the 195 breast cancer proteins and genes under investigation by the NCI Early Detection Research Network, 46 are present in NAF of which 22 were detected in all 8 samples and 7 were not found in plasma (Table 4-6). Among the candidate biomarkers detected in NAF were C-C motif chemokine 28, CSF1, EGFR, VEGFR1, VEGFA, ICAM1, KIT, HGFR, MMP9, metalloproteinase inhibitor 1 (TIMP1), osteopontin and Toll-like receptor 2.

Gene ID	Description	Accession	Mascot Score	# Spectra	# Unique Peptides	Plasma	Case profile
ADH5	Alcohol dehydrogenase class-3	P11766	252	25	4	MS	
AKR1C2	Aldo-keto reductase family 1 member C2	P52895	612	51	4	ND	
AKR1B1	Aldose reductase	P15121	184	42	1	MS	
ALPL	Alkaline phosphatase, tissue-nonspecific isozyme	P05186	3082	168	7	MS	
ANXA1	Annexin A1	P04083	9026	362	16	MS	
CDH1	Cadherin-1	P12830	10468	590	14	MS	
CEACAM5	Carcinoembryonic antigen-related cell adhesion molecule 5	P06731	1328	104	2	MS	NQ
CCL28	C-C motif chemokine 28	Q9NRJ3	318	30	1	ND	
MYCBP	C-Myc-binding protein	Q99417	25	9	2	MS	
CRP	C-reactive protein	P02741	38	8	1	MS/ELISA	
ENG	Endoglin	P17813	2738	196	9	MS/ELISA	
EFNA5	Ephrin-A5	P52803	660	121	6	ND	
EGFR	Epidermal growth factor receptor	P00533	132	18	3	MS/IA	
STOM	Erythrocyte band 7 integral membrane protein	P27105	22404	1392	14	MS	
FABP5	Fatty acid-binding protein, epidermal	Q01469	554	60	4	MS	
GPI	Glucose-6-phosphate isomerase	P06744	6832	334	11	MS	
GSTM1	Glutathione S-transferase Mu 1	P09488	7599	544	7	MS	
GSTM2	Glutathione S-transferase Mu 2	P28161	2427	197	6	ND	
GNB4	Guanine nucleotide-binding protein subunit beta-4	Q9HAV0	317	77	2	ND	
MET	Hepatocyte growth factor receptor	P08581	54	10	2	MS	
ITGB1	Integrin beta-1	P05556	32	7	2	MS	
ICAM1	Intercellular adhesion molecule 1	P05362	5942	337	12	MS/ELISA	
GLO1	Lactoylglutathione lyase	Q04760	253	24	3	MS	
LBP	Lipopolysaccharide-binding protein	P18428	15772	899	9	MS	
CSF1	Macrophage colony-stimulating factor 1	P09603	17423	1181	7	MS	
KIT	Mast/stem cell growth factor receptor Kit	P10721	301	39	4	MS	
MMP9	Matrix metalloproteinase-9	P14780	503	32	5	MS	
TIMP1	Metalloproteinase inhibitor 1	P01033	4627	378	7	MS	
NDUFA10	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 10, mitochondrial	O95299	107	7	1	MS	
PVRL4	Nectin-4	Q96NY8	360	54	4	ND	
SPP1	Osteopontin	P10451	11049	982	13	MS/ELISA	
PRDX4	Peroxisiredoxin-4	Q13162	3107	334	9	MS	
PDCD6IP	Programmed cell death 6-interacting protein	Q8WUM4	20305	1354	33	MS	
S100A4	Protein S100-A4	P26447	33	17	1	MS	
RAC1	Ras-related C3 botulinum toxin substrate 1	P63000	1097	177	5	MS	
RAB13	Ras-related protein Rab-13	P51153	961	75	2	MS	
RAB5A	Ras-related protein Rab-5A	P20339	1219	131	3	MS	
SERPINB3	Serpin B3	P29508	173	18	3	MS	
SSBP1	Single-stranded DNA-binding protein, mitochondrial	Q04837	33	5	2	MS	
TLR2	Toll-like receptor 2	O60603	2408	199	15	MS	
RHOA	Transforming protein RhoA	P61586	3306	364	2	ND	
TNFRSF11B	Tumor necrosis factor receptor superfamily member 11B	O00300	27	2	1	MS/ELISA	
TNFRSF1A	Tumor necrosis factor receptor superfamily member 1A	P19438	107	5	1	MS/ELISA	
VEGFA	Vascular endothelial growth factor A	P15692	69	32	1	EUSA/IA	
FLT1	Vascular endothelial growth factor receptor 1	P17948	295	58	1	EUSA	
ATP6AP1	V-type proton ATPase subunit S1	Q15904	3922	202	6	MS	

Table represents UniProt accession number, Gene ID, Protein name, Mascot score, number of spectra identified in MS analysis, number of unique peptides, detected in plasma by mass spectrometry (MS), Enzyme-linked immunosorbent assay (ELISA), multiple reaction monitoring mass spectrometry (MRM-MS), Immunoassay (IA), Case profile; histogram of 4 data sets; HV, PB, PD and PI. ND; not detected, NQ; Not Quantified.

Table 4-6: Biomarkers under investigation by the NCI Early Detection Research Network.

The expression of CD44 antigen was verified by western blot; higher expression was observed in NAF samples derived from patients with breast cancer (PD and PI) (Figure 4-8). The altered expression of CD44 had been reported in inflammatory responses and cellular malfunctioning during tumour progression (Basakran, 2015) but the exact role of CD44 is not clear. The high expression of vimentin (VIM) in cancer, correlates well with accelerated tumour growth, invasion, and poor prognosis (Satelli and Li, 2011). VIM has also been reported as a marker for epithelial-mesenchymal transition (EMT); a process by which epithelial cells lose their cell polarity and cell-cell adhesion characteristics (Kong et al., 2011). VIM was detected in all diseased NAF samples with two bands as 53kDa and 51kDa while in healthy volunteer only a single band at 53kDa was detected (Figure 4-8). VIM has three positions for glycosylation (Wang et al., 2012) and increased of glycosylation modification had also be reported in different cancers (Pinho and Reis, 2015). VIM also contains many modified residues with phosphorylation and acetylation (Uniprot.org, 2015). Hence, these modifications, may be the reason for two bands of VIM, in NAF samples derived from patients.

Cytochrome P450 3A4 (CYP3A4) was detected in 2 NAF samples by mass spectrometric analysis, which was verified by Western blotting (Figure 4-8). CYP3A4 has an important role in converting tamoxifen to N-desmethyl-4-hydroxytamoxifen, resulting in a 30 to 100-fold higher affinity for estrogen receptor than tamoxifen (Desta et al., 2004). CYP3A4, measured by immnuohistochemistry in normal and cancer breast tissue biopsies was found to be prognostic for patient response to docetaxel (Sakurai et al., 2011, Miyoshi et al., 2005) and by activity assay and western blot to correlate with ifosfamide activation (Schmidt et al., 2004). CYP3A4 is most strongly expressed in liver,

but its presence in NAF may provide a unique opportunity to screen for patients who are most likely to respond to prophylactic tamoxifen treatment.

In our previous study, we found the increased expression of cofilin-1 (CFL1) and calreticulin (CALR), in matched normal and disease breast tissues from DCIS and IC patients (Shaheed et al., 2013b), but no change in breast cancer cell lines and primary cells (Chapter 3). Validation studies of NAF samples also indicated no significant difference was observed for CFL1, by western blot, between matched pair of NAF samples (Figure 4-8). CALR is an endoplasmic reticulum chaperone protein and involved in tumour development but expression of CALR is mostly dependent on cell types and clinical stages of disease (Lu et al., 2015). CALR was detected as a single band (48kDa) in patient-derived NAF samples, with decreased expression in diseased breast as compare to normal breast (Figure 4-8). CALR is expressed on the cell membrane, promotes phagocytic uptake of apoptotic and cancer cells immune system (Obeid et al., 2007) and on cancer cell surface is important for immunogenic cell death (Lu et al., 2015, Gardai et al., 2005). Hence, decreased expression of CALR in diseased breast may correlate with failure of the immune system, one of hallmark of breast cancer (Hanahan and Weinberg, 2000).

The total protein complement encapsulates a rich source of information about breast tissue composition and more specifically uncovers changes in expression relating to disease. Overall, the composition of the selected ND/NAFs was dominated by proteins representative of the basement membrane, extracellular milieu and interstitial fluid surrounding breast cells, with roles in tissue stability, cell adhesion and cell-cell communication. Significantly, quantitative changes in these proteins may be diagnostic of disease-related

changes. Their increased presence in NAF of the cancer patients may be indicative of stromal disruption and degradation as cells migrate or proliferate into the surrounding normal tissue environment. Furthermore, many of the proteins were associated with plasma membrane activity including growth factors, receptors, signal transduction accessory proteins, and ion and solute transporters. Intracellular components, particularly mitochondrial and nuclear proteins were under-represented, suggesting that there is little evidence of cellular lysis. Serum proteins, many of which are also present in milk as part of normal breast function, were also prevalent and indicative of transport between the blood stream, interstitial fluid surrounding the cells and ductal secretions (Beck et al., 2015).

The collection of NAF remains challenging. In this study, samples were collected by manual massage techniques or when the patients presented with a natural discharge. Of those volunteers who consented, the success rate in NAF expression was approximately 50%. In order to screen for breast cancer specific biomarkers across a larger cohort increased success in expression and collection will be required. Improved expression rates up to 90% can be achieved with the aid of oxytocin nasal spray (Suijkerbuijk et al., 2010), supported by application of manual or mechanical pumps normally used for milk expression by mothers with preterm infants (Jones and Spencer, 2007).

4.5 CONCLUSIONS

My aim was to determine whether NAF has the potential to provide diagnostic value in screening for breast cancer. The study has identified many physiologically and oncologically important proteins which warrant a more expansive study of a larger cohort of patients and healthy volunteers. One of my aims was to understand whether analysing matched pairs would provide a more specific approach to detecting disease compared to normal. SDS PAGE showed that most matched pairs have similar protein band patterns, which was corroborated by the high correlation of proteomics profiles. Whilst some proteins were significantly different in bilateral samples, a more extensive longitudinal study is required to determine if these are associated with specific disease stages.

An average of 1265 proteins per sample, were identified by 2D LCMS, with a total of 1990 unique gene products, of which 567 were common to all 8 samples. The complement of proteins included growth factors (including EGF, VEGF), growth factor receptors (including EGFR, Erbb3, VEGFR) and 35 established breast cancer markers and breast associated proteins. Matched sample pairs shared the greatest similarity in composition (average Pearson correlation coefficient of 0.91), compared to those from other cases. The paired samples from the healthy volunteer had a unique profile dominated by milk proteins. Quantitative comparison of the non-cancer with cancer cases showed a significant increase in extracellular matrix and cell adhesion associated proteins. I identified double the number of proteins previously detected in NAF, including 300 not detected in plasma and 24% of the proteins currently part of the NCI Early Detection Research Network studying breast cancer.. The profiling of breast cancer markers like growth factors and receptors, required

immunoassays to prepare plasma samples for LC-MS analysis, but a single quantitative, multiplexed, target method by multiple reaction monitoring mass spectrometry can utilise the NAF samples efficiently by decreasing the sample preparation steps.

CHAPTER 5. DEVELOPMENT OF A NOVEL MULTIPLEX MRM-MS ASSAY FOR DETECTION OF BREAST CANCER.

5.1 Introduction

The new developments in the field of proteomics provide a more focused quantitative approach to verify and validate disease-related biomarkers and to understand the micro-environment of a cell. Although relative quantitative approaches; iTRAQ, SILAC and TMT, have their merits (Elliott et al., 2009), “absolute” quantitative approaches are particularly important for clinical implementation across large patient cohorts/ national medical services. An emerging technique that could satisfy this requirement for targeted quantitative analysis of a single or panel of tumour markers, is bottom-up Multiple reaction monitoring mass spectrometry (MRM-MS). This approach is more precise and accurate for quantitative analysis of disease related biomarkers in complex biomaterials, such as blood plasma and tissues. MRM-MS based targeted protein assay, offers two main advantages over immune-assays; (i) ability to develop a specific assay for any protein or post-translationally modified protein form, without the requirement for an antibody, (ii) multiplex assay (analysis of many peptides representative of many proteins in a single assay). Targeted peptide-based protein assays provide high throughput, quantitative proteomic data and have common application in the systematic development of assays for protein biomarker verification and validation studies (Liebler and Zimmerman, 2013). MRM-MS approach has become increasingly popular analytical technique in the last few years, with 126 publications in PubMed for “MRM-MS

and Proteomics” and average of 17 papers per month in 2017 (Figure 5-1) (NCBI, 2017). The advancement of technologies in MS methods and software for targeted quantitation, provide the analytical scientist with a transformative platform for systematic, reliable, and essentially universal protein quantitation.

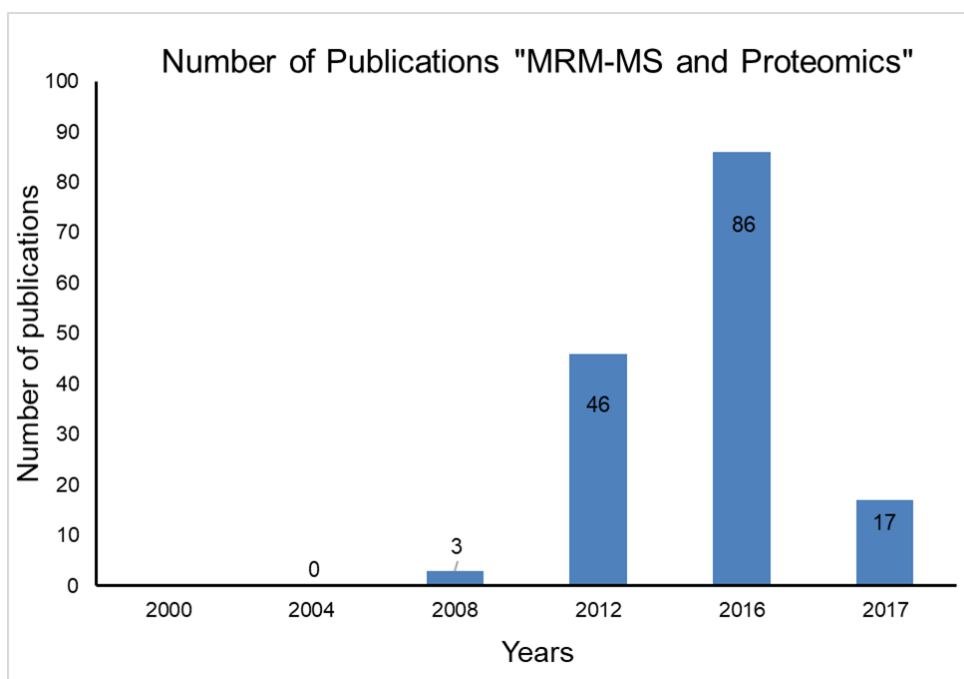


Figure 5-1: Number of PubMed Publications on MRM-MS and Proteomics.

In the MRM-MS approach, synthetic standard peptides for target proteins are used to optimise the LC-MS conditions then the optimised multiplex assay is used to calculate the concentration of the target proteins in biological samples from standard calibration curves of the synthetic standards. Peptides with unique sequences to the target protein are selected as representatives for the parent protein. Ideally tryptic peptides are selected for MRM-MS assay with sequence-specific digestion rules (see below selection of proteins and peptides). Each peptide ion (precursors) undergoes fragmentation in a tandem MS instrument, to produce b- and y- ions (products), which are N- and C-terminal fragments of peptide ions respectively (Liebler, 2001). Both precursors

and resulting specific products, contain transitions that are specific for the monitored peptide sequences (Kitteringham et al., 2009, Elliott et al., 2009). The analysis of peptides through the signals created by precursors and their products ions, is called multiple reaction monitoring (MRM). The peak area of precursors and their products ions are determined as abundance of peptides and serve as the basis for quantitative comparisons. MRM-MS assay can measure multiple transitions in single method, which allows multiplexed analysis of hundreds of peptides. The multiplexity of a MRM-MS assays depends on three things; (i) type of MS instrument (ii) number of transitions for each peptide, and (iii) number of peptides monitored for each target protein (Zhao et al., 2009b, Doerr, 2013). The multiplexity of MRM-MS assay can be increased by monitoring the MRM transitions during specific elution time windows for specific peptides. The multiplex analysis, allow systematic quantification of multiple proteins networks and pathways in single LC-MS run, which is not possible with immunoblotting assays (Zhao et al., 2009b).

The gold standard MRM-MS assay is on based stable isotope dilution (SID), in which isotope labelled peptides are used as internal standard to measure the abundance of target peptides. Most common type of labelled peptides have incorporation of [$^{13}\text{C}_6^{15}\text{N}_2$] lysine or [$^{13}\text{C}_6^{15}\text{N}_4$] arginine (Doerr, 2013). SID-based MRM-MS assay, provide high sensitivity, specificity and <20% coefficient variations because standards are usually spiked into samples after trypsin digestion and thus serve to decrease the subsequent analytical steps variations. But synthesis of peptide standards for SID is challenging to get isotopic purity (>95%) and also very expensive (£500-£800 per peptide), if monitoring large number of peptides in a multiplex assay (Liebler and Zimmerman, 2013).

The simplest and easy MRM-MS methods is based on light peptides (LP) or label free peptides, where standard peptides are used to draw the calibration curves then abundance of target peptides in biological samples, is measured by using these curves(Doerr, 2013). Light standard peptides are very easy to synthesizes with >99% purity and also very cheap (£100-£300 per peptide) but are analysed separately in MRM-MS assay, resulting in higher measurement variation, because of undetected or uncorrected variations in differences in peptide recovery and MS instrument performance (Liebler and Zimmerman, 2013). The LP-based MRM-MS assay, has been compared to the SID method and performed surprisingly well, with measurement CVs ranging from 20 to 30% while on SID method measurement CVs was 15 to 20% (Zhang et al., 2011).

MRM-MS assay is performed on a mass spectrometer with multiple mass analysers; the first mass analyser (Q1) is set to only transmit the target mass (parent), the collision energy is optimized to produce charged daughters of the selected parent in the second mass analyser (Q2), and the third mass analyser (Q3) is set to transmit these daughters only. Therefore, only this exact parent transition is detected. In addition, the method preparation on MRM analysis involves parameters (collision energy, cone voltage, retention time) of mass spectrometry that need to be optimized for the best transmission and sensitivity of each transition ion and to avoid any overlaps during detection (Liebler and Zimmerman, 2013).

In this study, three Waters instruments; Xevo® TQD, Micromass® Quattro Premier XE and Micromass® Quattro Ultima with different MassLynx versions (Table 5-1), were evaluated for performance (dynamic range,

sensitivity, multiplexity), based on availability at different stages during the project timeline, to obtain a highly sensitive and selective method for the targeted biomarker proteins in complex biological samples. Waters Quattro Ultima is a standard triple Quadrupole mass spectrometer with Electrospray (ES) and Chemical ionisation (APCI) sources and Z-shaped ions pathway, which allow charged particles only while the neutrals and the solvent go straight. Micromass® Quattro Premier XE was first generation of mass spectrometer with T-Wave collision cell, which provide high speed MRM data acquisition. The Waters Xevo TQD is an advance benchtop mass spectrometer with multiple ionisation options, T-wave collision cell and Intellistart features. Intellistart features of Xevo TQD provide; automated SIR and MRM method development, mass calibration, column performance test, integrated sample and calibrant delivery system (WatersCorporation, 2017a). The comparison of main features of three instruments, used in this study to develop multiplex MRM-MS assay, are presented in Table 5-1 (WatersCorporation, 2017a).

Specification	Quattro Ultima	Quattro Premier XE	Xevo TQD
Source and ionization modes	API source	ESI/API source	Multi-mode source-ESI/API/APCI/ESCI
Digital dynamic range of detector	1x10 ⁵	1x10 ⁵	4x10 ⁶
Collision cell (T-Wave enabled)	NO	YES	YES
Simultaneous full scan and MRM (RADAR)	NO	NO	YES
Mass range	2 to 1,500 m/z	2 to 1,500 m/z	2 to 2,048 m/z
Mass stability	>0.5 Da / 24 hour	>0.5 Da / 24 hour	<0.1 Da / 24 hour
Polarity switching time	20 ms	20 ms	20 ms
MS to MS/MS switching	10 ms	10 ms	3 ms
MRM acquisition rate	5 ms / MRM channel	5 ms / MRM channel	1 ms / MRM channel
Number of MRM channels	up to 30 /method	up to 1000 /method	up to 32,768 /method
Mass resolution	>1.0 Da	>1.0 Da	>0.5 Da
Intellistart Technology (acquisition rate assignment)	NO	NO	YES
Automated MRM scheduling (acquisition rate assignment)	NO	NO	YES
Automated MRM scheduling (acquisition window assignment)	NO	NO	YES
System supported software	MassLynx NT	MassLynx version 4.0	MassLynx version 4.1
ESI; Electrospray ionization, API; Atmospheric Pressure Ionization, APCI; Atmospheric pressure chemical ionization, ESCI; Combination of ESI and APCI.			

Table 5-1: Comparison of Xevo TQD, Quattro Premier XE and Quattro Ultima, used for multiplex MRM-MS assay.

The advantages of MRM-MS assays; multiplexing capability, low development cost (without cost of mass spectrometer), sensitivity (limit of quantification; LOQ), specificity, and high sample throughput, eliminate the need for antibody based assays for verification and validation of protein biomarkers (Liebler and Zimmerman, 2013). Recently, a large-scale MRM-MS study was performed on depleted human plasma to detect more than 100 putative cancer biomarkers like Alpha-fetoprotein (hepatocellular carcinoma), PSA (prostate cancer), CA-125 (non-small cell lung cancer), CA19.9 (pancreatic cancer), CA 15.3 (breast cancer), leptin, prolactin, osteopontin, and IGF-II (ovarian cancer), CD98 and fascin (lung cancer), CEA (malignant pleural effusion), Her-2/neu (stage IV breast cancer), Bladder Tumor Antigen (urothelial cell carcinoma), Thyroglobulin (thyroid cancer metastasis) (Hüttenhain et al., 2012). This approach was impressive but depletion step added some drawbacks; cost to the assay, potential removal of clinically relevant high-abundance proteins, potential increase in experimental variation and bottleneck in sample throughput. Furthermore, depletion can remove low molecular weight proteins due to weak, non-covalent binding to the depletion column or carrier proteins such as albumin (Percy et al., 2013). The method used by Domanski et al. for determination of 67 cardiovascular disease (CVD) biomarkers like fetuin A, factor VII, eotaxin-1, cystatin C, macrophage inflammatory protein-1 β , matrix metalloproteinase-2 and matrix metalloproteinase-9, level in un-depleted human plasma, was robust (<20% CV), sensitive (accuracy 80-100%), and high-throughput in a 30-min analysis (Domanski et al., 2012b). While the MRM-MS assay used by Chen et al. for the quantitation of 63 putative protein biomarkers of bladder cancer in human urine, had 71.1% sensitivity and 75.0%

specificity for differentiating bladder cancer from non-cancerous patients (Chen et al., 2012).

From the published literature, there have been three MRM-MS based studies for the differentiation of breast cancer. First MRM-MS based study was conducted by Whiteaker and his co-workers in 2007 on HER2/Neu-driven mouse model and identified fibulin-2 and osteopontin as potential diagnostic biomarkers for breast cancer (Whiteaker et al., 2007). A large-scale SI-MRM-MS assay was conducted by Kennedy et al 2014, for 319 target proteins expressed in 30 breast cancer cell lines. The median precision of MRM assay was 5.4% and the quantity of specific target proteins in breast cell lines was able to differentiate the molecular subtypes of breast cancer (Kennedy et al., 2014). Marchi et al 2016, applied MRM-MS on ER+ breast cancer tissues and serum and proposed programmed cell death protein 4 (PDCD4), cingulin (CGN), Ras GTPase-activating protein-binding protein 2 (G3BP2), and OCIA domain-containing protein 1 (OCIAD1) as potential biomarker for ER+ breast cancer (De Marchi et al., 2016). But this study was more focused on expression of PDCD4, CGN, G3BP2, and OCIAD1 in tamoxifen resistance ER+ breast cancer.

The lengthy time and laborious procedures, are two main draw backs for using cancer protein biomarkers as diagnostic approach, especially their verification and validation in large numbers of patient samples. A rapid, sensitive, and robust analytical method is required for breast cancer biomarkers, to bridge this gap between discovery and pre-clinical validation. Ideally, this method should also be multiplexed; allowing multiple proteins to be analysed simultaneously in a single LC-MS experiment, to increase the throughput and minimize sample consumption (Liebler and Zimmerman, 2013). The focus of

method development in this study, was to develop a multiplex quantitative proteomics assay for the detection and differentiation of breast cancer and its phenotypes, without pre-analytical affinity-based depletion or enrichment steps.

5.2 Selection of proteins and peptides for MRM-MS assay

The first step in MRM-MS assay development was the selection of target proteins inferred from previous discovery experiments and correlating scientific literature. For this study, in addition to the 6 target proteins, 2 control proteins were included (Table 5-2).

In our discovery project (Shaheed et al., 2013b) on tissue biopsies, we identified an average of 826 proteins, of which 402 were common in fibroadenoma (benign tumors, three patients), DCIS (noninvasive cancer, three patients), and invasive ductal carcinoma (four patients). After excluding those originating from blood, 59 proteins were significantly changed in tumor compared with normal tissues, with the majority associated with invasive carcinomas. We found increased expression of cofilin-1 (CFL1), prostaglandin E synthase 3 (PTGES3), SH3 domain-binding glutamic acid-rich-like protein 3 (SH3BGRL3), and 10 kDa heat shock protein, mitochondrial (HSPE1). The expression of membrane primary amine oxidase (AOC3) and hormone-sensitive lipase (LIPE) was decreased in tissue with advance stages of breast cancer (Shaheed et al., 2013a). B-Actin and serum albumin were included as indicators of the cellularity and serum component present in each sample.

The next step was to select the peptides that will be suitable for MRM analysis. Key steps in the workflow for configuring MRM-MS assays for proteins are summarized in Figure 5-2. The specificity and accuracy of the MRM-MS assay requires selection of proteotypic peptides, whose sequences are unique

to each target protein biomarker. The MRM-MS assay is based on tryptic peptides, which typically range from 8 to 25 amino acids in length and should be reproducibly generated by a standard trypsin digestion procedure. The tryptic peptides usually form multiply charged positive ions to collect useful sequence information through MS/MS fragmentations (Lange et al., 2008, Picotti and Aebersold, 2012).

Sr. no	UniProt accession	Gene	Protein	Location of Protein	Breast cancer response*	Methods of Analysis
1	P23528	CFL1	Cofilin-1	intracellular	Increased in DCIS/IC	MS, WB, IHC
2	P61604	HSPE1	10 kDa heat shock protein, mitochondrial	intracellular	Increased in DCIS/IC	MS, WB, IHC
3	Q15185	PTGES3	Prostaglandin E synthase 3	intracellular	Increased in IC	MS, WB, IHC
4	Q9H299	SH3BGRL3	SH3 domain-binding glutamic acid-rich-like protein 3	intracellular	Increased in IC	MS, WB
5	Q16853	AOC3	Membrane primary amine oxidase	membrane	Decreased in IC	MS, WB
6	Q05469	LIPE	Hormone-sensitive lipase	membrane	Decreased in IC	MS, WB
7	P60709	ACTB	Actin, cytoplasmic 1	all cell types	no change	MS, WB
8	P02768	ALB	Serum albumin	blood	N/A	MS, WB

*: Shaheed et al 2013; **N/A**: Not Applicable, **IC**; Invasive carcinoma, **DCIS**; Ductal carcinoma in situ, **MS**; Mass spectrometry, **WB**; Western blotting, **IHC**; Immunohistochemistry

Table 5-2: List of proteins selected for MRM-MS assay and expression in breast cancer.

Peptides which may have post translational modifications, such as phosphorylation or acetylation, were avoided. Peptides with residues susceptible to artefactual modifications during sample preparation; methionine (oxidation), cysteine (oxidation and carbamidomethylation), were also avoided for MRM-MS assay. Peptides with missed cleavages containing sequences adjacent basic residues, e.g. Lys-Lys, Arg-Arg, Lys-Arg, were also avoided to decrease digestion variability as these are inefficiently cleaved (Lange et al., 2008, Picotti and Aebersold, 2012). Proteotypic peptides are a very specific signature for the protein biomarker. Ideally, three proteotypic peptides are used representing the full protein sequence – N-terminal, C-terminal and middle regions, however this was not always possible.

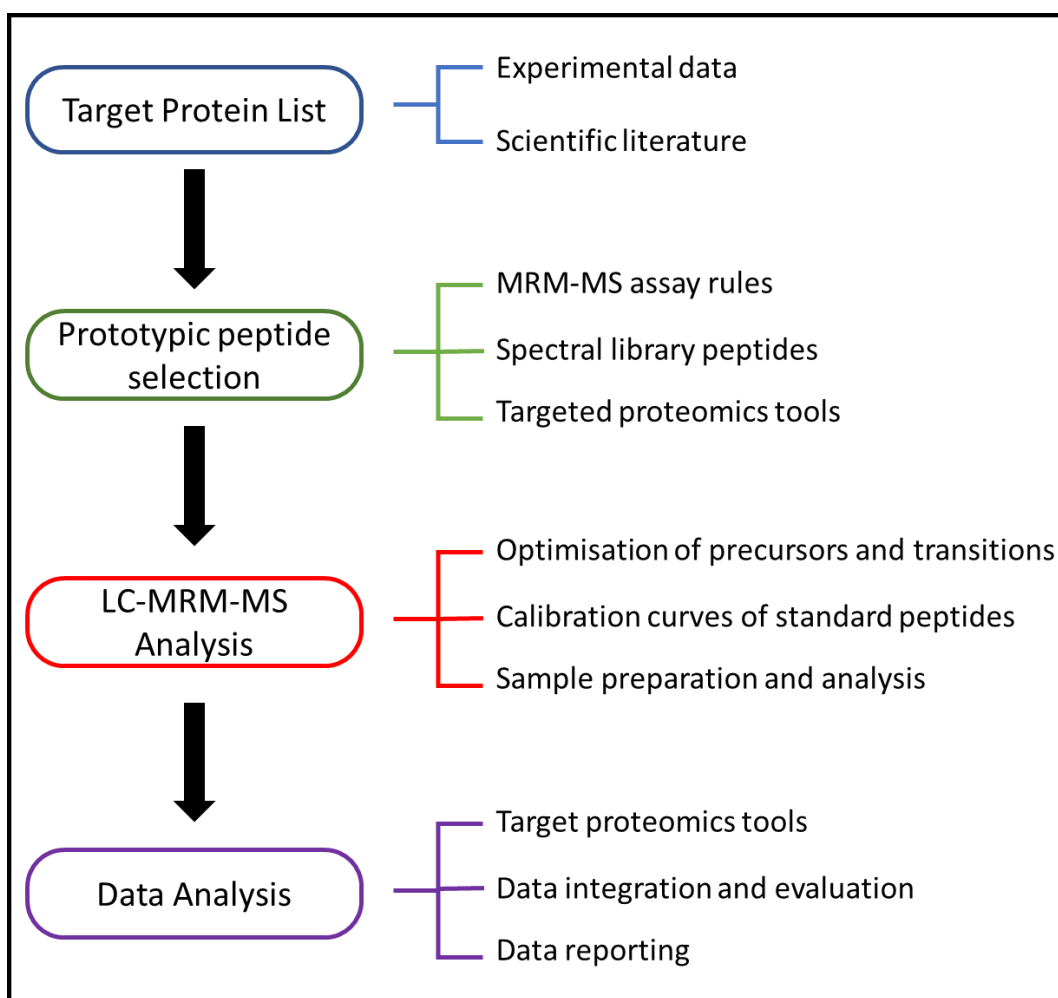


Figure 5-2: Workflow for the design of a MRM-MS assay.

The selection of peptides is an empirical exercise that is the combination of both ideal characteristics and practical limitation. Larger proteins yield more signature peptides than smaller protein, so sequence of proteins play important role in peptide selection. If the target proteins have significant sequence homology or members of closely related isoforms, then it can be difficult to follow to the peptide selection criteria described above. Our previous breast cancer study provided peptide identification of the target proteins (Shaheed et al 2013) (Shaheed et al., 2013a).

An important aspect of peptides selection is to include those that have been previously observed in LC-MS analyses and thus are known to be detectable in MRM-MS assay (Picotti and Aebersold, 2012). This information can be gathered from previous discovery data or online repositories such as PeptideAtlas (PeptideAtlas, 2017), the Global Proteome Machine Database (GPM, 2016), the Human Proteome map (Kim et al., 2014), the MaxQuant Database (Schaab et al., 2012) and PRIDE (Vizcaíno et al., 2016). For proteins or peptides not found in any Spectral library databases, computational software tools; Skyline (MacLean et al., 2010), ESP predictor (Fusaro et al., 2009) and PeptideSieve (Thermo Scientific) can predict the most likely MS-observable peptides. These software packages are trained with MS data sets and fragmentation models to link characteristics of peptides to the probability of peptide formation and detection. Proteotypic peptide candidates for each protein were identified by following the steps in Figure 5-2 and described above. Table 5-3 represents the final panel of proteins and peptides for MRM-MS analysis.

Sr. no	accession	Gene	Protein	Peptides ID	Peptide Sequence	Avg. MW	Hydrophobicity	MRM-MS Rules	Tissue Data	BCL Data	NAF Data	Peptides Atlas	MaxQB Data	HPM Data	Skyline
1	P23528	CFL1	Cofilin-1	CFL1-1	LGGSAVISLEGKPL	1339.77	32.53	YES	YES	YES	YES	YES	YES	YES	YES
				CFL1-2	EILVGDVGQTVDDPYATFVK	2165.09	40.32	YES	YES	YES	YES	YES	YES	YES	YES
				CFL1-3	YALYDATYETK	1336.62	23.12	YES	YES	YES	YES	YES	YES	YES	YES
2	P61604	HSPE1	10 kDa heat shock protein, mitochondrial	HSPE1-1	FLPLFDR	906.5	33.11	YES	YES	YES	ND	YES	YES	YES	YES
				HSPE1-2	SAAETVTK	805.42	6.42	YES	ND	YES	ND	YES	YES	YES	YES
				HSPE1-3	VLQATVVAVGSGSK	1314.75	22.31	YES	YES	YES	YES	YES	YES	YES	YES
3	Q15185	PTGES3	Prostaglandin E synthase 3	PTGES3-1	DVNVNFEK	963.47	15.69	YES	ND	YES	ND	YES	YES	YES	YES
				PTGES3-2	LTFSCLOGGSDNFK	1387.64	28.64	NO	YES	YES	YES	YES	YES	YES	NO
4	Q9H299	SH3BGL3	SH3 domain-binding glutamic acid-rich-like protein 3	SH3BGL3-1	VYIASSSGSTAIK	1282.68	21.34	YES	YES	YES	ND	YES	YES	YES	YES
				SH3BGL3-2	QQDVLGFLEANK	1360.7	36.23	YES	YES	ND	ND	YES	ND	ND	YES
				SH3BGL3-3	GDYDAFFEAR	1189.5	26.99	YES	YES	ND	ND	YES	ND	ND	YES
5	Q16853	AOC3	Membrane primary amine oxidase	AOC3-1	YQLAVTQR	977.53	17.18	YES	YES	ND	ND	YES	ND	ND	YES
				AOC3-2	SPVPPGPAPPLQFYPPGPR	2001.05	34.9	YES	YES	ND	ND	YES	ND	YES	YES
				AOC3-3	HGGPLPYHR	1032.53	17.8	YES	YES	ND	ND	YES	ND	YES	YES
6	Q05469	LIPE	Hormone-sensitive lipase	LIPE-1	EQALGLEPALGR	1252.68	27.93	YES	ND	ND	ND	YES	YES	YES	YES
				LIPE-2	ALVYYAQR	982.52	19.99	YES	ND	ND	ND	YES	ND	YES	YES
7	P60709	ACTB	Actin, cytoplasmic 1	ACTB-1	GYSFTTTAER	1131.52	17.07	YES	YES	YES	YES	YES	YES	YES	YES
8	P02768	ALB	Serum albumin	ALB-1	LVNEVTEFAK	1148.61	24.51	YES	YES	ND	YES	YES	YES	YES	YES

Avg. MW; average molecular weight of peptide, **Hydrophobicity** of peptide; calculated by Sequence-specific retention calculator, **MRM-MS**; Multiple Reaction Monitoring Mass spectrometry, **Tissue data**; Shaheed et al 2013, **BCL data**; Breast cell lines data (chapter 3), **NAF data**; Nipple aspirate fluid data (chapter 4), **Peptide Atlas**; online database for MRM-MS assay, **MaxQB data**; The MaxQuant DataBase of label-free quantification, **HPM** ; The Human Proteome Map database with high resolution and high accuracy acquired on Orbitrap instruments, **Skyline**; a freely-available and open source to develop MRM-MS assay, **YES**; peptide is ideal for MRM-MS assay and detected in MS analysis, **NO**; peptide is not Ideal for MRM-MS assay, **ND**; Not detected in MS analysis.

Table 5-3: Panel of proteins and peptides for MRM-MS analysis.

5.3 Materials and methods

5.3.1 Chemicals and Reagents

Where ever possible, the highest-grade chemicals and reagents were used in this study, and are described in Chapter 2 General Materials and Methods; Section 2.1 Chemicals and Reagents. All mobile phases and solutions including water, acetonitrile, methanol, and formic acid, were prepared with LC-MS grade solvents from Sigma Aldrich.

5.3.2 Synthetic peptides

5.3.2.1 Stock peptides

For each targeted protein, one to three associated peptides were selected using the rules in section 5.2 above. These peptides were purchased from Severn Biotech Ltd UK, with >99% purity in quantities of 2-4 mg. Prior to MRM-MS analysis, these peptides were resuspended in HPLC grade water and aliquoted as 1 nmol/ μ L, lyophilized and stored in -20°C .

5.3.2.2 Working solution

Each aliquoted peptide was resuspended in HPLC grade water by vortexing to produce a working solution of 100 pmol/ μ L and then further 2-fold serial diluted to 3 fmol/ μ L, with HPLC water prior to LC-MS analysis. MALDI-MS analysis and all optimisation of MRM-MS assay was conducted with 100 pmol/ μ L working peptides solution.

5.3.3 MALDI Mass spectrometry

All the synthetic peptides were analysed manually on the Ultraflex II (Bruker), to confirm the sequence and purity of peptides. A mixture containing; 1 μ L of synthetic peptide (at a concentration of xx pmol/ μ l) and 1 μ L of α -cyano-4-hydroxycinnamic acid MALDI Matrix, was applied on the MTP AnchorChip (Bruker) and analysed on the Ultraflex II. An MS spectrum was generated (100 shots) in reflectron time-of-flight mode, to confirm the correct mass. Acquired MS/MS spectrum (200 shots, in LIFT mode) for individual peptide, was searched against SwissProt data base, using Mascot server 2.4.0, to unambiguously confirm the sequence of peptide and identity of the protein.

5.3.4 Sample preparation

All biological samples (cell lines, tissue biopsies, serum and NAF) (Table 5-4), were prepared, using classical bottom-up proteomics workflows, as previously described in Chapter 2 (Section 2.9). Briefly, urea extraction buffer containing protease inhibitor cocktail, was used to extract protein from cell lines, but for primary human tissue biopsies, a dual buffer system (RIPA and urea) was used followed by probe sonication (Shaheed et al., 2013a).

5.3.5 Protein determination

The protein concentration of each sample was measured using the Bradford assay (Bradford, 1976) protein quantification kit, according to the manufacturer's instructions (see Chapter 2, Section X). Each protein extract (200 μ g of protein) from cell lines and tissue biopsies, was precipitated overnight with 100% acetone, to remove extraction buffer, at -20°C and centrifuged for 20 min at 13 400 rpm at 4°C . The pellet was resuspended in 8M urea in 400mM

ammonium bicarbonate (AMBIC) and protein concentration was again measured by Bradford assay, to check the efficiency of acetone precipitation.

Each protein sample (50 µg of protein) from cell lines, tissue biopsies, NAF and serum, was reduced with 50 mM DTT for 15 min at 60°C, alkylated with 100 mM IAA at ambient temperature for 15 min in the dark, and digested with 3 µL of a 1 mg/mL solution of modified sequencing grade trypsin (Roche Diagnostics) at 37°C for 20 hrs. After digestion, each sample was desalted on an Isolute C18 desalting column according to the manufacturer's instructions and lyophilized. A simple work flow for the preparation of samples for proteomics analysis is presented in Figure 5-3.

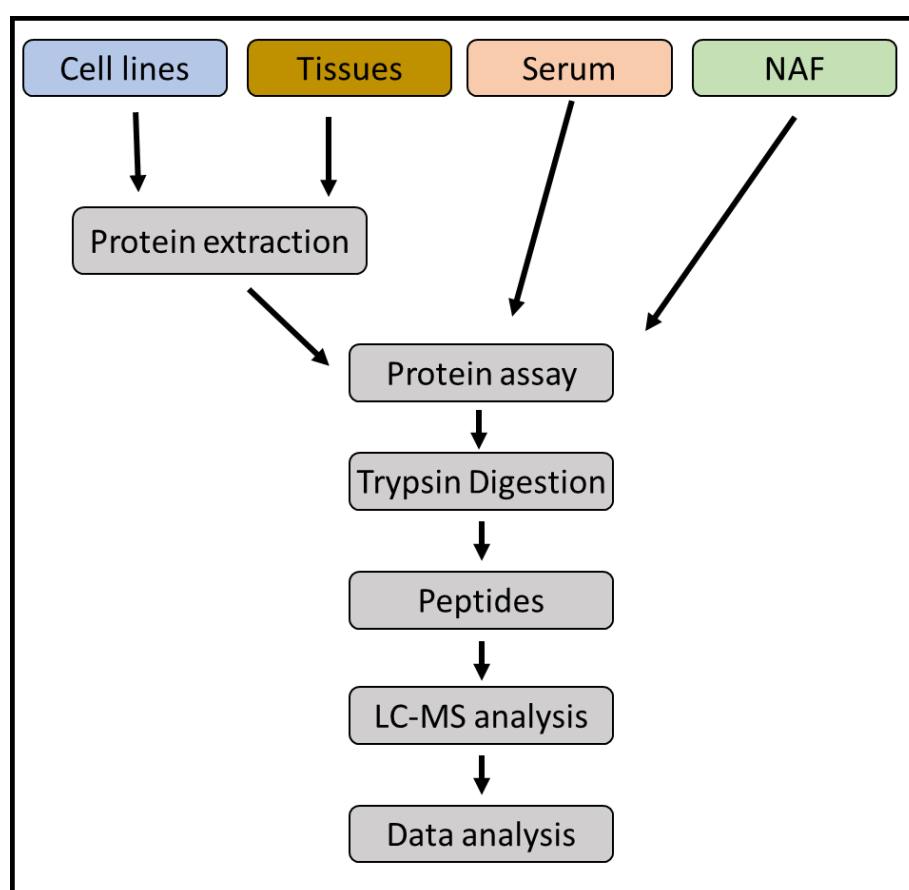


Figure 5-3: Representation of proteomics work flow for the preparation of samples.

Sr. No	Sample Type	Sample ID	Characteristics
1	Breast cell lines	MCF-7	Luminal A breast cancer
2		ZR-75	Luminal B breast cancer
3		MDA-MB-453	HER2+ breast cancer
4		MDA-MB-468	Basal-like breast cancer
5		MDA-MB-231	Claudin-low breast cancer
6		HB2	Normal-Like
7		MCF-10A	Normal-Like
8		HMEC	Normal-human mammary epithelial cells
9	Serum	HVS-1	Healthy volunteer
10		HVS-2	Healthy volunteer
11		HVS-3	Healthy volunteer
12		HVS-4	Healthy volunteer
13		HVS-5	Healthy volunteer
14		PTS-1	Papillary Carcinoma
15		PTS-2	Ductal carcinoma in situ (DCIS)
16		PTS-3	Ductal tubular carcinoma
17		PTS-4	Invasive ductal carcinoma
18		PTS-5	Invasive ductal carcinoma
19	Nipple aspirate fluid	PTN1-L	Fibrocystic change
20		PTN1-R	Normal
21		PTN2-L	Normal
22		PTN2-R	Lobular carcinoma
23		PTN3-L	Ductal carcinoma
24		PTN3-R	Normal
25		PTN4-L	Ductal carcinoma
26		PTN4-R	Normal
27	Tissues biopsy	PTT1-N	Normal
28		PTT1-T	Fibroadenoma
29		PTT2-N	Normal
30		PTT2-T	Ductal carcinoma in situ (DCIS)
31		PTT3-N	Normal
32		PTT3-T	Invasive Carcinoma
33		PTT4-N	Normal
34		PTT4-T	Invasive Carcinoma

HVS; Healthy volunteer serum, PTS; patient serum, PTN; patient nipple aspirate fluid, L; left breast, R; right breast, PTT: patient tissue, N; Normal, T; Tumour.

Table 5-4: List of samples used in this study with biopsy types and characteristics.

5.3.6 LC-MRM-MS analysis

All experiments were performed on standard-flow LC-MRM/MS platforms with Xevo TQD, Quattro Premier XE or Quattro Ultima triple quadrupole mass spectrometers (Waters, Manchester, UK). Quattro Ultima was coupled to Waters Alliance 2695 HPLC Separation system, Quattro Premier XE with Waters Acquity UPLC Separation system and Xevo TQD with on-line Waters Acquity I-Class UPLC separation system. The same HPLC separation conditions were used for separating target peptides; solvent A and solvent B (please see Chapter 2, Table 2-3 for full details) on a Luna C18 column (2 mm ID x 25 cm length, 5µm particle size, Phenomenex Inc., Macclesfield, UK). A linear gradient of 3 to 40% solvent B was applied for 38 minutes then 40% to 90% for 2 minutes followed by a column wash for 10 minutes using 90% solvent B and column equilibration for 10 minutes using 10% solvent B prior to the next injection (Chapter 2, Table 2-3). To enhance chromatographic performance, the column was maintained at 30°C for all experiments, while the auto-sampler was set at 8°C. In the quantitation experiments, two blanks were run between each sample concentration level and the synthetic standard peptides were analysed, in order of increasing concentration to minimize sample carryover. All standard peptides and samples were analysed in triplicate. All instruments were controlled by MassLynx workstation software (Table 5-1) and were operated in the positive ion mode. The HPLC/UPLC retention time, cone voltage and collision energy were optimized using synthetic reference peptides (Table 5-6 and Table 5-7). Calibration curves were prepared with the synthetic peptides (Table 5-3) before quantitative analysis of trypsin digests of cell lines, tissue biopsies, serum and NAF.

5.3.7 MRM-MS Data analysis

All MRM data was processed, evaluated, and visualized with TargetLynx™ (Waters) Quantitative and Qualitative Analysis software, as described (WatersCorporation, 2017b). After automated chromatographic peak selection and integration by TargetLynx, MRM data was also manually verified to ensure correct chromatographic peak selection and integration. From the calibration curves of standard peptides, dynamic range, the lower limit of quantitation and peptides concentration values (in fmoles/ μ L) were determined in biological samples. The amount of each target protein, was calculated by average concentration of signature peptides in each sample.

5.4 Results

5.4.1 MALDI MS and MS/MS data

All the synthetic peptides had the expected molecular weights and the MS/MS data identified the peptide and protein correctly (Table 5-5). The Mascot score for all synthetic peptides was greater than 23 (significant, $p < 0.05$).

Sr. no	Peptide name	Peptides Sequence	Avg. MW	MH+ (mono)	MF ions	Mascot score
1	CFL1-1	LGGSAVISLEGKPL	1339.77	1340.778	29	101
2	CFL1-2	EILVGDVGQTVDDPYATFVK	2165.09	2166.096	21	119
3	CFL1-3	YALYDATYETK	1336.62	1337.626	21	95
4	HSPE1-1	FLPLFDR	906.5	907.504	14	60
5	HSPE1-2	SAAETVTK	805.42	806.425	18	71
6	HSPE1-3	VLQATVVAVGSGSK	1314.75	1315.758	34	113
7	PTGES3-1	DVNVNFEK	963.47	964.473	13	58
8	PTGES3-2	LTFSCGGSDNFK	1387.64	1388.651	22	80
9	SH3BGRL3-1	VYIASSSGSTAIK	1282.68	1283.684	18	132
10	SH3BGRL3-2	QQDVLGFLEANK	1360.7	1361.706	29	86
11	SH3BGRL3-3	GDYDAFFEAR	1189.5	1190.511	20	99
12	AOC3-1	YQLAVTQR	977.53	978.537	15	53
13	AOC3-2	SPVPPGPAPPLQFYPQGPR	2001.05	2002.055	38	114
14	AOC3-3	HGGPLPYHR	1032.53	1033.533	24	73
15	LIPE-1	EQALGLEPALGR	1252.68	1253.685	27	142
16	LIPE-2	ALVYYAQR	982.52	983.531	19	114
17	ACTB-1	GYSFTTTAER	1131.52	1132.527	18	95
18	ALB-1	LVNEVTEFAK	1148.61	1149.615	25	97

Avg MW; Average Molecular Weight, **MH+**; monoisotopic mass of peptide + 1.008, **MF ions**; number of experimental fragment ions matching theoretical fragment ions; **Mascot score**; calculated probability (p) of an MS/MS between experimental data and database sequence.

Table 5-5: Verification of synthetic peptides analysed by MALDI MS against SwissProt database.

5.4.2 Optimizing the MRM-MS transitions

Prior to multiplex MRM scheduling, all target peptides were analysed separately to optimise LC-MS conditions. The retention time (RT) of each peptide and its transitions, was checked and a retention time segment was set to 2-4 minutes in multiplex assay, for each targeted peptide with its expected RT in the centre based on the individual synthetic peptides analysis. The (Figure 5-4) represent the retention time in HPLC and charge status of target precursors selected for multiplex MRM-MS assay.

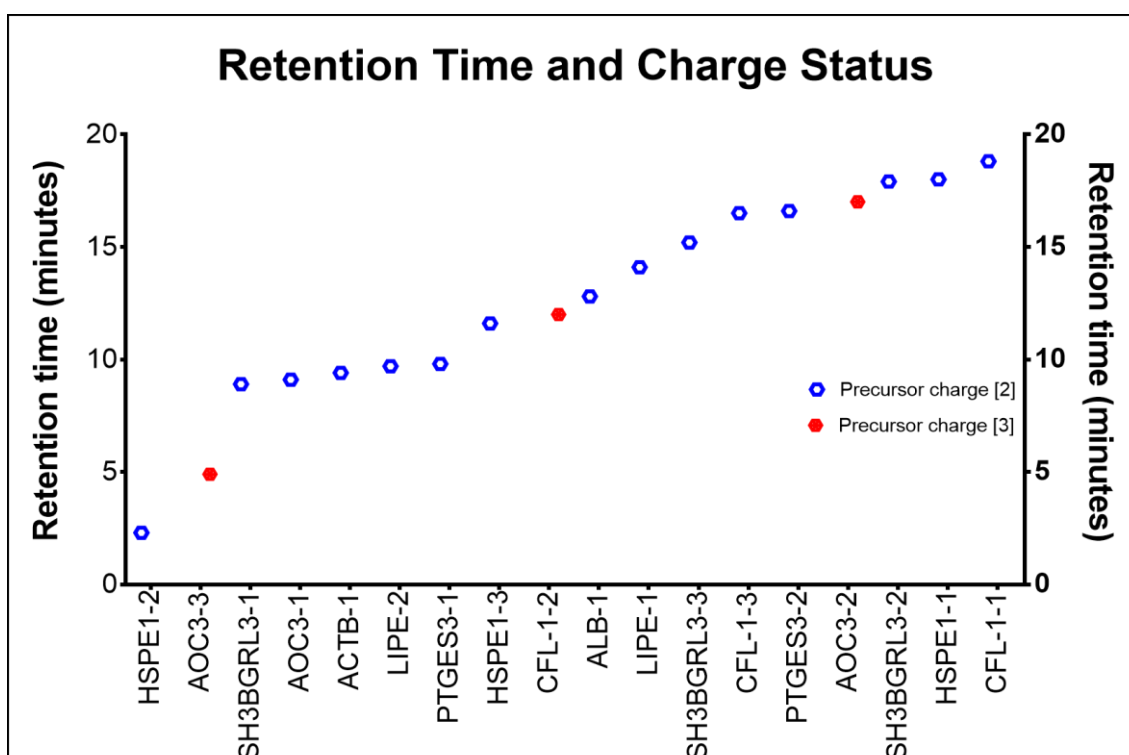


Figure 5-4: Retention time and charge status of target precursors in multiplex MRM-MS assay.

The optimised values of cone voltage (CV), collision energy (CE) and Dwell time for Quattro Ultima and Premier XE, were determined manually while on Xevo TQD, a fully automated option in MassLynx (WatersCorporation, 2017b) was used to calculate cone voltage, Dwell time and collision energy values for individual peptides were then verified manually (Table 5-6). From 18

target peptides; 15 precursors were selected with 2⁺-charge status and 3 precursors with 3⁺-charge status, on the base of peak intensity and MRM-MS rules. The transition ions of each peptide were selected by following rules:

- a) precursor ions with positive charge states of 2⁺ or 3⁺,
- b) y or b series of fragment ions with a charge state of one,
- c) the three most intense fragment ions in the MS/MS spectra from untargeted analysis, and
- d) m/z of precursor and transition ions between 200 and 1500.

All selected transitions were also manually checked by comparing the full scan MS² data from three Waters instruments against MS² data acquired on MALDI MS analysis, to confirm identity of transitions with the expected properties. The final list of all transitions selected for multiplex MRM-MS assay is presented in Table 5-7. The transitions that produced the highest corresponding peaks and free from co-eluting ions, was then selected as the representative transition to be used in the final MRM-MS assay. After optimising the LC-MS conditions (Table 5-6 and Table 5-7) for each peptide a multiplex MRM-MS assay was generated to check the LoD and LoQ of three instruments.

Peptide No.	Peptide ID	Peptides Sequence	Mono Isotopic Mass	Precursor m/z	Quattro Ultima			Premier XE			Xevo TQD		
					Dwell* time (Sec)	Cone* (V)	CE (V)	Dwell* time (Sec)	Cone* (V)	CE (V)	Dwell* time (Sec)	Cone* (V)	CE (V)
1	CFL-1-1	LGGSAVISLEGKPL	1339.77	723.00	0.300	30	30	0.300	30	30	0.06	40	30
2	CFL-1-2	EILVGDVGQTVDDPYATFVK	2165.09	669.61	0.300	30	25	0.300	30	25	0.06	40	25
3	CFL-1-3	YALYDATYETK	1336.62	671.21	0.300	30	26	0.300	30	26	0.03	40	26
4	HSPE1-1	FLPLFDR	906.50	454.26	0.300	30	18	0.300	30	18	0.05	30	18
5	HSPE1-2	SAAETVTK	806.43	403.21	0.300	30	20	0.300	30	20	0.13	30	20
6	HSPE1-3	VLQATVVAVGSGSK	1315.76	658.70	0.300	30	26	0.300	30	26	0.06	35	26
7	PTGES3-1	DVNVNFEK	963.47	482.98	0.300	30	17	0.300	30	17	0.04	30	17
8	PTGES3-2	LTFSCLGGSDFNK	1387.57	695.16	0.300	30	27	0.300	30	27	0.03	35	27
9	SH3BGRL3-1	VYIASSSGSTAIK	1282.68	642.68	0.300	30	28	0.300	30	28	0.04	40	28
10	SH3BGRL3-2	QQDVLGFLEANK	1360.70	681.36	0.300	30	28	0.300	30	28	0.05	40	28
11	SH3BGRL3-3	GDYDAFFEAR	1189.50	595.76	0.300	30	21	0.300	30	21	0.13	30	21
12	AOC3-1	YQLAVTQR	977.53	489.77	0.300	30	18	0.300	30	18	0.04	30	18
13	AOC3-2	SPVPPGPAPPLQFYPQGPR	2001.05	668.00	0.300	30	24	0.300	30	24	0.03	40	24
14	AOC3-3	HGGPLPYHR	1032.53	345.29	0.300	30	15	0.300	30	15	0.06	30	15
15	LIPE-1	EQALGLEPALGR	1253.68	626.84	0.300	30	24	0.300	30	24	0.13	40	24
16	LIPE-2	ALVYYAQR	983.53	492.62	0.300	30	18	0.300	30	18	0.04	30	18
17	ACTB-1	GYSFTTTAER	1132.53	566.24	0.300	30	20	0.300	30	20	0.04	40	20
18	ALB-1	LVNEVTEFAK	1149.62	575.63	0.300	30	22	0.300	30	22	0.13	40	22

Table 5-6: Optimised Dwell time, Cone voltage and Collision energy (CE) conditions for multiplex MRM-MS assay, using synthetic peptides of target proteins.

Peptide No.	Peptide ID	Peptides Sequence	Mono Isotopic Mass	Precursor m/z	RT (minutes)	Precursor charge state	Transition m/z
1	CFL-1-1	LGGSAVISLEGKPL	1339.77	723.00	18.8	2	215.21, 413.25, 825.77
2	CFL-1-2	EILVGDVGQTVDDPYATFVK	2165.09	669.61	12.0	3	827.55, 990.5, 1103.66
3	CFL-1-3	YALYDATYETK	1336.62	671.21	16.5	2	229.21, 743.65, 856.4
4	HSPE1-1	FLPLFDR	906.50	454.26	18.0	2	437.31, 550.54, 647.53
5	HSPE1-2	SAAETVTK	806.43	403.21	2.3	2	448.18, 648.4, 719.29
6	HSPE1-3	VLQATVVAVGSGSK	1315.76	658.70	11.6	2	341.23, 543.5, 605.29
7	PTGES3-1	DVNVNFEK	963.47	482.98	9.8	2	537.04, 636.43, 750.32
8	PTGES3-2	LTFSCLOGGSDNFK	1387.57	695.16	16.6	2	724.46, 837.19, 940.53
9	SH3BGRL3-1	VYIASSSGSTAIK	1282.68	642.68	8.9	2	838.25, 908.75, 1022.29
10	SH3BGRL3-2	QQDVLGFLEANK	1360.70	681.36	17.9	2	672.38, 778.37, 891.65
11	SH3BGRL3-3	GDYDAFFEAR	1189.50	595.76	15.2	2	522.17, 669.17, 740.62
12	AOC3-1	YQLAVTQR	977.53	489.77	9.1	2	503.34, 574.22, 687.38
13	AOC3-2	SPVPPGPAPPLQFYPQGPR	2001.05	668.00	17.0	3	400.91, 554.37, 650.39
14	AOC3-3	HGGPLPYHR	1032.53	345.29	4.9	3	462.13, 475.26, 572.26
15	LIPE-1	EQALGLEPALGR	1253.68	626.84	14.1	2	513.62, 642.29, 812.59
16	LIPE-2	ALVYYAQR	983.53	492.62	9.7	2	537.16, 700.15, 800.55
17	ACTB-1	GYSFTTTAER	1132.53	566.24	9.4	2	447.57, 678.18, 912.44
18	ALB-1	LVNEVTEFAK	1149.62	575.63	12.8	2	595.28, 694.34, 937.4

Table 5-7: Optimised Retention time (RT) and transitions for multiplex MRM/MS assay, using synthetic peptides of target proteins.

5.4.3 Calibration curves of the synthetic peptides

After optimising the LC-MRM/MS assay, calibration curves were generated by titrating a mixture of standard peptides (from 100 pmol/uL to 0.003 pmol/uL, 2-fold serial dilutions). Three peptides of cofilin-1 (CFL1); LGGSAVISLEGKPL (CFL1-1), EILVGDVGQTVDDPYATFVK (CFL1-2) and YALYDATYETK (CFL1-3), had dynamic range from 25 pmoles to 0.01 pmoles (Limit of Detection, LoD). The Limit of Quantitation (LoQ) on Xevo TQD, for CFL1-1, CFL1-2 and CFL1-3, was 0.08 pmoles, 0.04 pmoles and 0.02 pmoles respectively (Table 5-8, Figure 5-5, A, B, and C). The values of linear regression for CFL1-1 (R^2 : 0.9633), CFL1-2 (R^2 : 0.9898) and CFL1-3 (R^2 : 0.9945), were higher in Xevo TQD as compare to other two instruments, with ten points calibration curves (Figure 5-5.).

10 kDa heat shock protein, mitochondrial (HSPE1) was represented by three peptides; FLPLFDR (HSPE1-1), SAAETVTK (HSPE1-2), and VLQATVVAVGSGSK (HSPE1-3). The LoD of HSPE1-1 was up to 0.0002 pmoles (R^2 : 0.9846) on Xevo TQD while on Quattro Ultima was only 1.56 pmoles (R^2 : 0.9552) (Figure 5-5, D). Similar kind of dynamic range was also observed on Quattro Ultima for other two peptides; HSPE1-2 and HSPE1-3, with linear regression of R^2 : 0.995 and R^2 : 0.9814 respectively (Table 5-8, Figure 5-5, E, F).

The LoD for SH3BGRL3-1 was 0.001 pmoles Xevo TQD and other two peptides; SH3BGRL3-2 and SH3BGRL3-3, also have higher sensitivity on Xevo TQD, with ten and eleven points calibration curves respectively (Table 5-8, Figure 5-6, B, D, E).

The dynamic range (LoD and LoQ) for two peptides of prostaglandin E synthase 3 (PTGES3); DVNVNFEK (PTGES3-1) and LTFSCCLGGSDNFK

(PTGES3-2), was 0.01 pmoles and 0.3 pmoles on Xevo TQD with linear regression $R^2:0.999$ and $R^2:0.9568$ respectively (Figure 5-6, E, F).

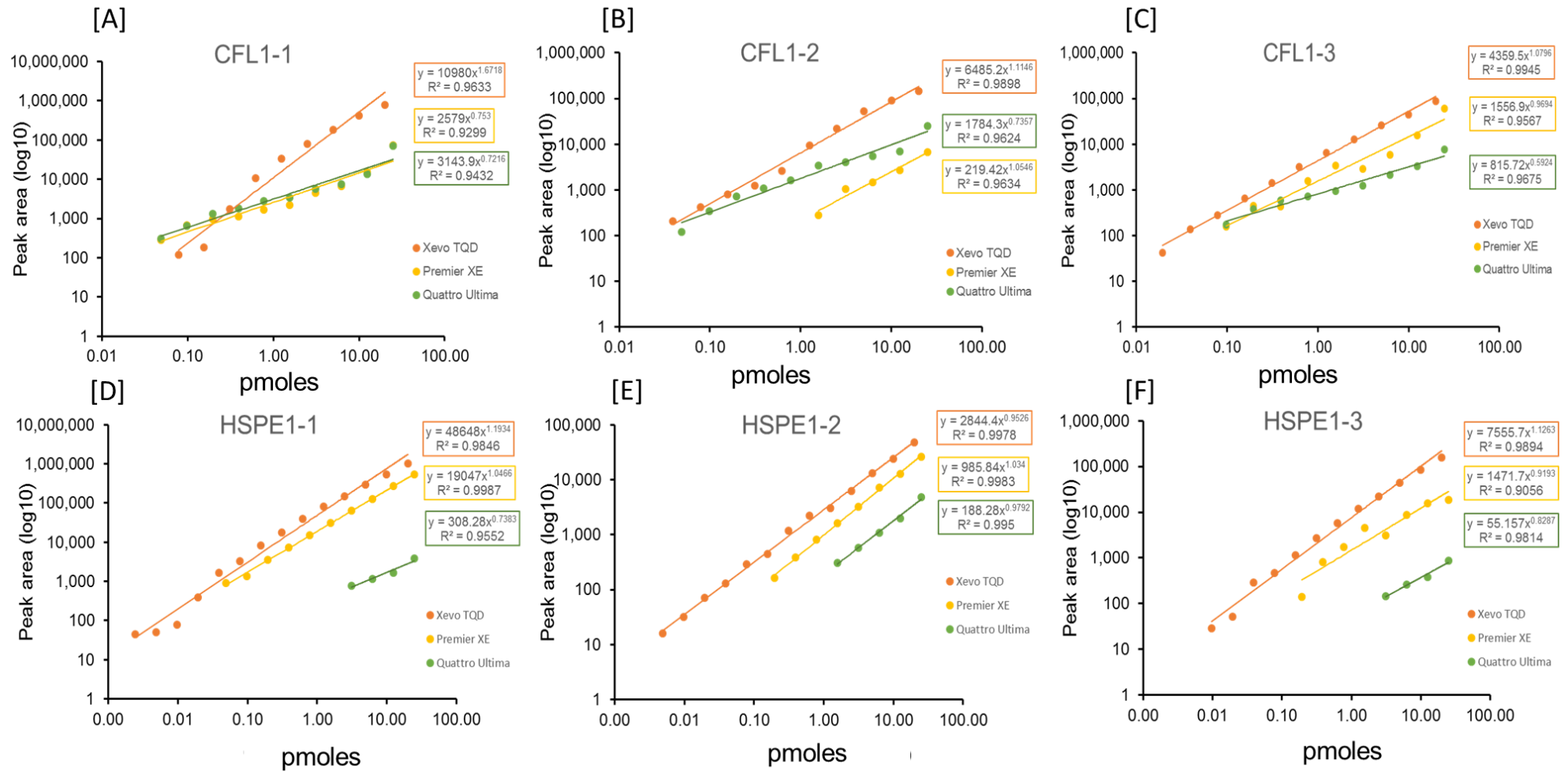


Figure 5-5: Representative standard curves from the multiplex LC-MRM/MS analysis of 6 peptides (CFL1 and HSPE1).

Serum albumin (ALB) was represented by a single peptide; LVNEVTEFAK (ALB-1), in this study. The LoD for ALB-1 on Xevo TQD, was 2.4 fmoles ($R^2:0.9944$), with fourteen points calibration curve (Figure 5-6, F). The LoQ for AOC3-1 was 0.6 fmoles, while AOC3-2 and AOC3-3, had 40 fmoles, on Xevo TQD (Table 5-8, Figure 5-7, A, B, C). Two peptides of hormone-sensitive lipase (LIPE); EQALGLEPALGR (LIPE-1) and ALVYYAQR (LIPE-2), had 0.02 pmoles as LoQ on Xevo (Figure 5-7, D, E) while single peptide of actin, cytoplasmic 1 (ACTB); GYSFTTTAER (ACTB-1), was up to 0.02 pmoles (LoD) on Xevo TQD, 0.2 pmoles on Premier XE (Figure 5-7, F).

Peptide ID	Peptides Sequence	Quattro Ultima		Premier XE		Xevo TQD	
		LoD (pmoles/ μ l)	LoQ (pmoles/ μ l)	LoD (pmoles/ μ l)	LoQ (pmoles/ μ l)	LoD (pmoles/ μ l)	LoQ (pmoles/ μ l)
CFL-1-1	LGGSAVISLEGKPL	0.05	0.05	0.05	0.05	0.0400	0.0800
CFL-1-2	EILVGDVGQTVDDPYATFVK	0.05	0.05	0.78	1.56	0.0200	0.0400
CFL-1-3	YALYDATYETK	0.05	0.10	0.05	0.10	0.0100	0.0200
HSPE1-1	FLPLFDR	1.56	3.12	0.05	0.05	0.0025	0.0025
HSPE1-2	SAAETVTK	0.78	1.56	0.20	0.20	0.0025	0.0050
HSPE1-3	VLQATVVAVGSGSK	1.56	3.12	0.20	0.20	0.0050	0.0100
PTGES3-1	DVNVNFEK	0.39	0.78	0.19	0.39	0.0010	0.0010
PTGES3-2	LTFSCCLGGSDNFK	0.78	1.56	0.39	0.78	0.0800	0.1600
SH3BGRL3-1	VYIASSSGSTAIK	0.05	0.10	0.05	0.05	0.0010	0.0010
SH3BGRL3-2	QQDVLGFLEANK	0.39	0.78	0.10	0.20	0.0200	0.0400
SH3BGRL3-3	GDYDAFFEAR	1.56	3.12	0.19	0.39	0.0100	0.0200
AOC3-1	YQLAVTQR	0.10	0.20	0.5	0.10	0.0006	0.0006
AOC3-2	SPVPPGPAPPLQFYPPGPR	0.39	0.78	0.78	1.56	0.0200	0.0400
AOC3-3	HGGPLPYHR	0.05	0.10	0.39	0.78	0.0200	0.0400
LIPE-1	EQALGLEPALGR	1.56	3.12	0.05	0.05	0.0050	0.0100
LIPE-2	ALVYYAQR	0.20	0.40	0.05	0.10	0.0050	0.0050
ACTB-1	GYSFTTTAER	0.78	1.56	0.10	0.20	0.0100	0.0200
ALB-1	LVNEVTEFAK	0.20	0.40	0.05	0.05	0.0020	0.0020

LoD; Limit of detection, LoQ; limit of quantification

Table 5-8: Summary of LoD and LoQ all standard peptides on three platforms.

Overall, Xevo TQD provided better dynamic range and sensitivity for multiplex MRM-MS assay. The dynamic range of multiplex assay was from average 0.2 fmoles to 20 pmoles for standard peptides, with 14 points calibration curves. On this basis, the biological samples (Table 5-4) were analysed on Xevo TQD and concentration of target peptides were determined by using the standards calibration curves of synthetic peptides.

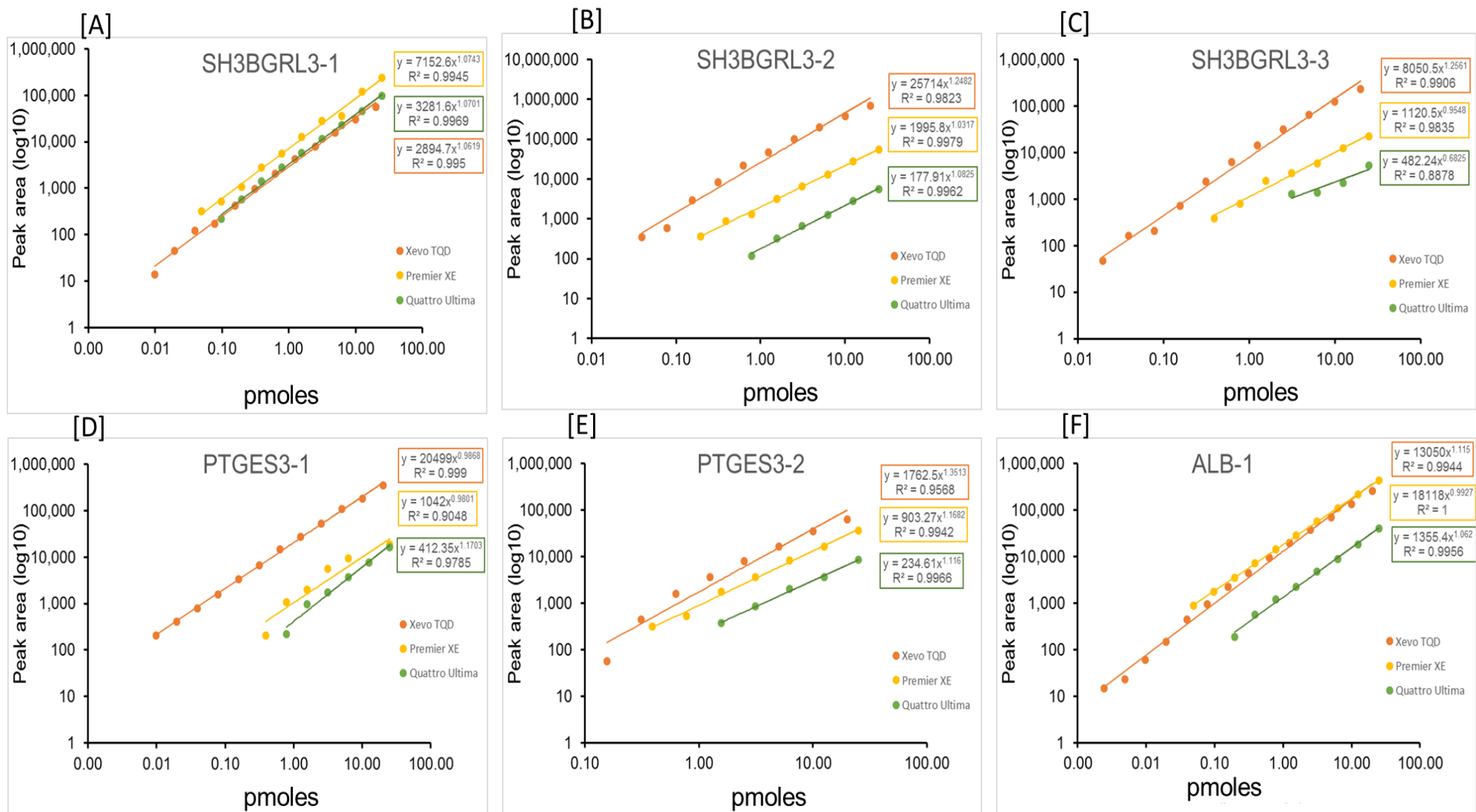


Figure 5-6: Representative standard curves from the multiplex LC-MRM/MS analysis of 6 peptides (SH3BGRL3, PTGES3 and ALB).

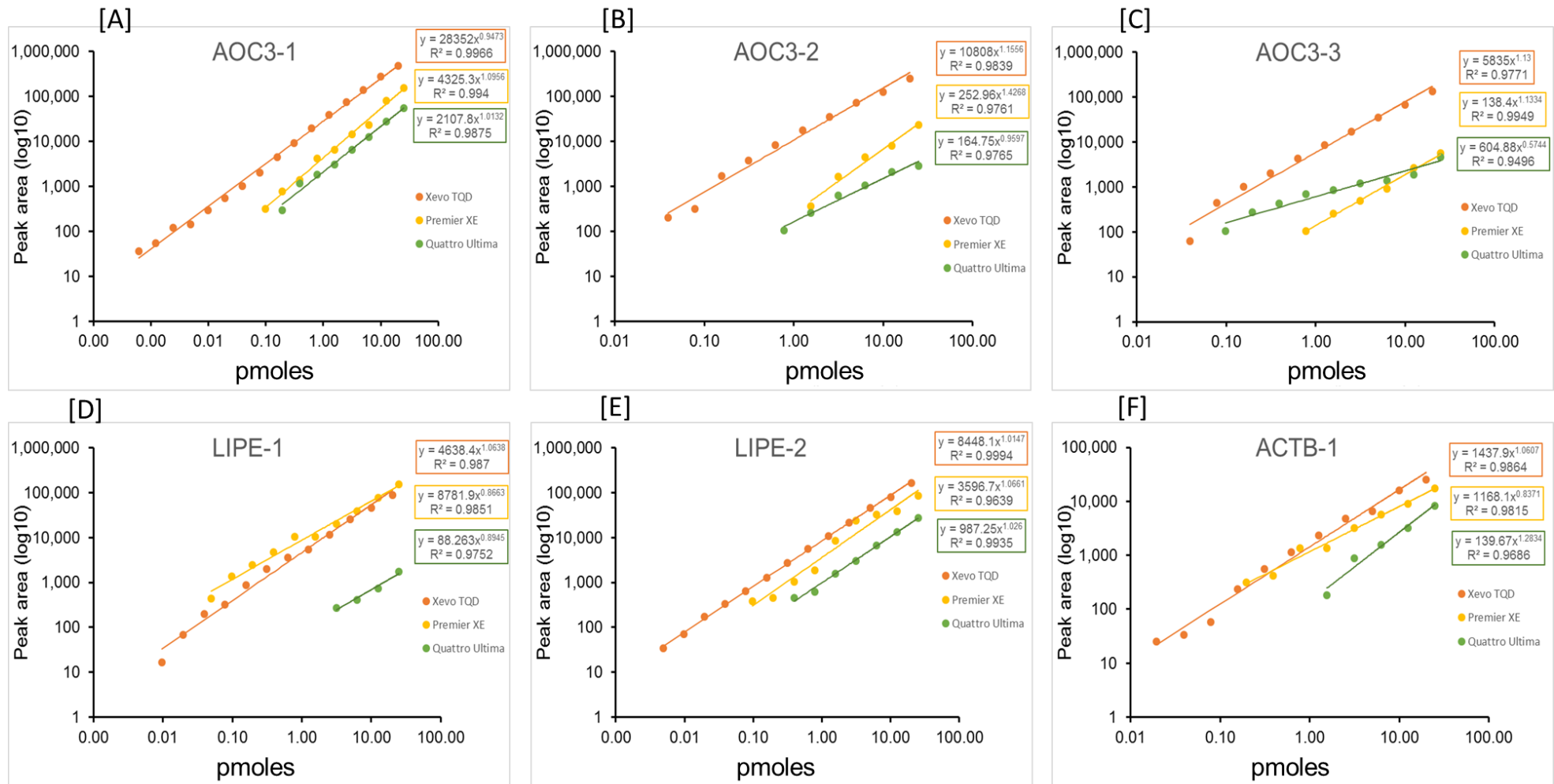


Figure 5-7: Representative standard curves from the multiplex LC-MRM/MS analysis of 6 peptides (AOC3, LIPE and ACTB).

A comparison of the LOD across the 3 platforms indicated that the Ultima was least sensitive (HSPE1-3 = 3 pmoles), then the Premier was next (HSPE1-3 = 0.2 pmoles), and the Xevo was most sensitive (HSPE1-3 = 0.01 pmoles). Based on measurements at 3 pmoles the sensitivity of the three instruments was compared for all target peptides. T-Test statistics was applied on peak areas of all 18-peptides at 3 pmoles. There was significant difference between the sensitivity of three instruments; Xevo-TQD vs Premier XE (p value 0.0290), Xevo-TQD vs Quattro Ultima (p value 0.0006), and Premier XE vs Quattro Ultima (p value 0.0108) (Figure 5-8).

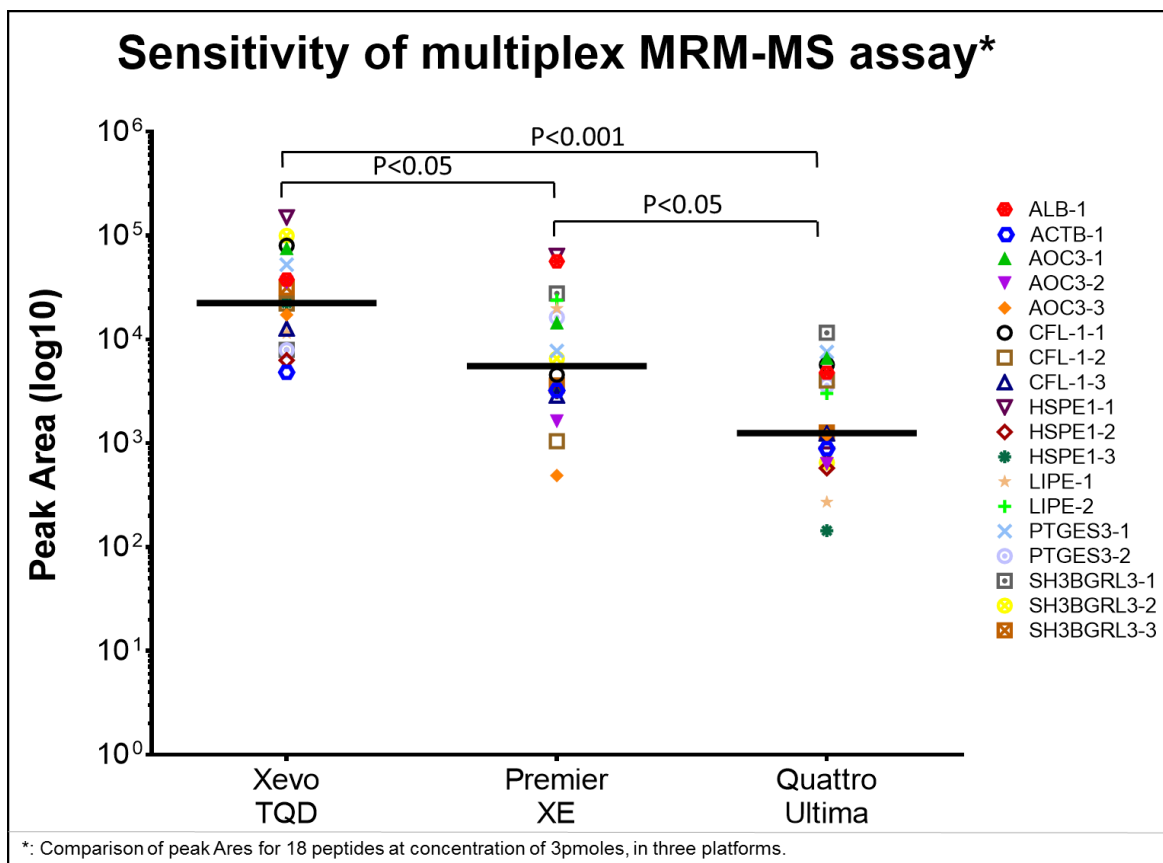


Figure 5-8: Sensitivity comparison of multiplex MRM-MS assay on three different platforms.

Overall, the Xevo TQD provided better dynamic range and sensitivity for a multiplex MRM-MS assay compared to the other two platforms. The dynamic range was 6 orders of magnitude (0.2 fmoles to 20 pmoles, 14 points

calibration curves, triplicate analyses) for standard peptides, compared to The Premier XE (3 orders of magnitude) and Quattro Ultima (2 orders of magnitude). On this basis, the biological samples were analysed, and target peptide concentrations determined, by using the standards calibration curves prepared on Xevo TQD (Table 5-4).

5.4.4 Multiplex analysis of biological materials

Using the optimised MRM-MS assay for the 6 targets and 2 control proteins, the proteotypic peptide levels were analysed in biological materials - breast cell lines, tissues, serum, and nipple aspirate fluid, to determine; -

- a) whether the peptides can be detected
- b) whether the levels between healthy and tumour can be differentiated, and
- c) whether a multiplex assay for detection of breast cancer could be developed.

5.4.4.1 Breast cell lines

Initially, the multiplex assay was applied to eight breast cell lines; MCF-7, ZR-75, MDA-MB-453, MDA-MB-468, MDA-231, HB2, MCF-10A and HMEC. All cell lines were analysed in triplicate on the Xevo TQD by optimised MRM-MS assay (Supplementary Table 5.1).

All, three peptides of cofilin-1; CFL1-1, CFL1-2 and CFL1-3, were detected and quantified in breast cell lines with dynamic range of 128.5 fmoles to 425.90 fmoles in 1.25µg of digest used for LC-MRM-MS analysis (Table 5-9). The expression of cofilin-1 was significantly higher ($p < 0.05$) in a cell line representative of claudin-low type breast cancer (MDA-MB-231) as compared

to HMEC (Figure 5-9,A). The coefficient of variations (CV) of all peptides of cofilin-1 in triplicate analysis, was less than 8% (Table 5-9).

Cell line	CFL-1-1		CFL-1-2		CFL-1-3		Cofilin-1			
	Average amount of peptide	Std. Dev. of peptide	Average amount of peptide	Std. Dev. of peptide	Average amount of peptide	Std. Dev. of peptide	Average amount of protein	Average Std. Dev. of peptides	p value	CV%
MCF-7	391.60	13.45	263.67	19.08	276.27	7.91	310.51	13.48	0.118	4.34
ZR-75	290.73	3.69	116.30	2.01	162.70	5.65	189.91	3.78	0.912	1.99
MDA-MB-453	378.20	14.70	378.47	22.94	259.90	6.41	338.86	14.68	0.065	4.33
MDA-MB-468	263.13	8.33	216.03	14.98	145.97	10.70	208.38	11.34	0.847	5.44
MDA-MB-231	425.90	12.57	326.70	10.50	308.60	12.00	353.73	11.69	0.044	3.30
HB2	271.60	6.58	230.73	16.69	128.50	9.80	210.28	11.02	0.838	5.24
MCF-10A	291.93	13.24	261.33	22.72	189.60	16.30	247.62	17.42	0.373	7.04
HMEC	267.60	9.09	194.63	3.60	130.60	8.50	197.61	7.06	1.000	3.57

Table represents average amount of CFL-1-1, CFL-1-2 and CFL-1-3, in fmoles per μL of sample ($\sim 1.25\mu\text{g}$ of digest injected for LC-MRM-MS). Std. Dev. of peptides: standard deviation of peptides in triplicate analysis. The average amount of Cofilin-1 was determined from average amount of peptides. T-Test statistics was applied between control (HMEC) and all other cell lines to calculate the p values. The percentage of coefficient of Variation (CV) of protein was determined by average peptides amount and average standard deviation of peptides in triplicate analysis.

Table 5-9: Replicate analyses of three peptides of Cofilin-1 in breast cell lines.

Three peptides of 10 kDa heat shock protein, mitochondrial (HSPE1), were detected in all breast cell lines with high expression ($p < 0.05$) in luminal type (MCF-7, ZR-75, and MDA-MB-453 breast cancer cell lines, as compare to HMEC. There was also significant difference (p value was 0.01) between luminal and basal type breast cancer phenotype (Figure 5-9, B, Supplementary Table 5.1). In MCF-7, SH3BGRL3-1, SH3BGRL3-2 and SH3BGRL3-3 were detected with highest amount 25.77 fmoles, 16.93 fmoles and 26.13 fmoles respectively (Supplementary Table 5.1). So, the average amount of SH3BGRL3 was significantly higher (p value 0.018), in luminal A (MCF-7) type breast cancer compared to HMEC (Figure 5-9, D Supplementary Table 5.1). Only one peptide of prostaglandin E synthase 3 (PTGES3), was detected in all cell lines. T-test analysis, of the five cell lines; MCF-7, MDA-MB-453, MDA-MB-231, HB2 and MCF-10A, had p values < 0.0001 compared to HMEC, while ZR-75 and MDA-MB-468 had p values < 0.001 and < 0.01 respectively (Figure 5-9, C Supplementary Table 5.1).

Membrane primary amine oxidase (AOC3) levels, were determined by a single peptide; AOC3-3, because AOC3-1 and AOC3-2 were not detected in any breast cell lines. On the base of triplicate values of AOC3-3, I determined the amount of AOC3 in breast cell lines, with highest amount of 68.93 fmoles in MCF-7 and 56.73 fmoles in MDA-MB-231 and were significantly different, $p < 0.0001$ and $p = 0.04$, respectively compared to HMEC levels (Figure 5-10, A). From two peptides of hormone-sensitive lipase (LIPE), only LIPE-2 was detected in cell lines with highest amount of 12.07 fmoles in the Claudin-low breast cell line (MDA-MB-231, p value < 0.01) (Figure 5-10, B). In breast cell lines, the amount of actin, cytoplasmic 1 (ACTB), was in range from 350.80 fmoles to 1990.0 fmoles for ZR-75 and MDA-MB-231 respectively. On T-Test analysis, I found that only ZR-75, has significant difference ($p < 0.05$), from control cell line (HMEC) Figure 5-10, C).

As expected, serum albumin (ALB), was not detected in any analysis of Breast cell lines. Using this multiplex MRM-MS assay, I quantified 6 target proteins and one control protein in breast cell lines.

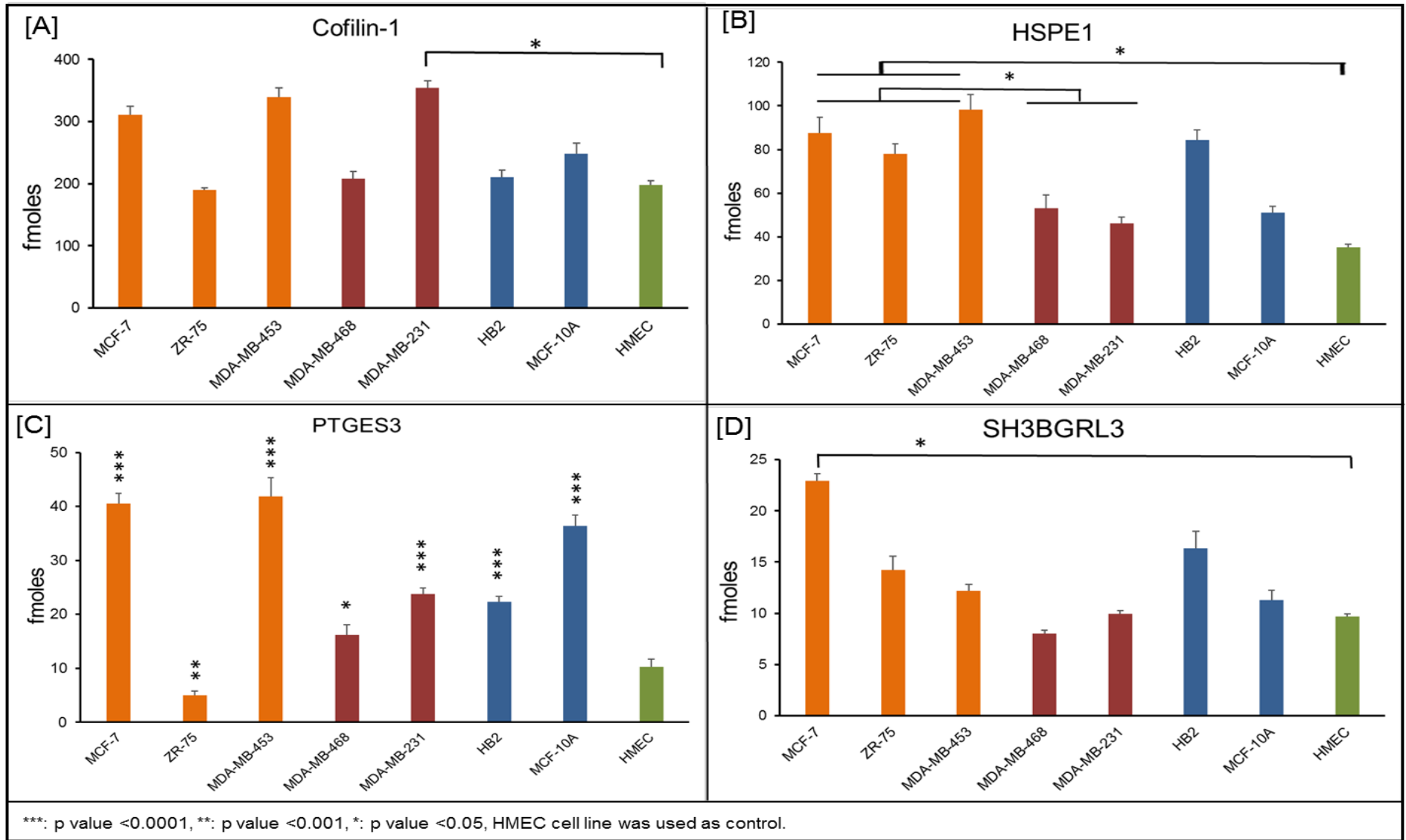


Figure 5-9: Expression of Cofilin-1, HSPE1, SH3BGRL3 and PTGES3 in breast cell lines.

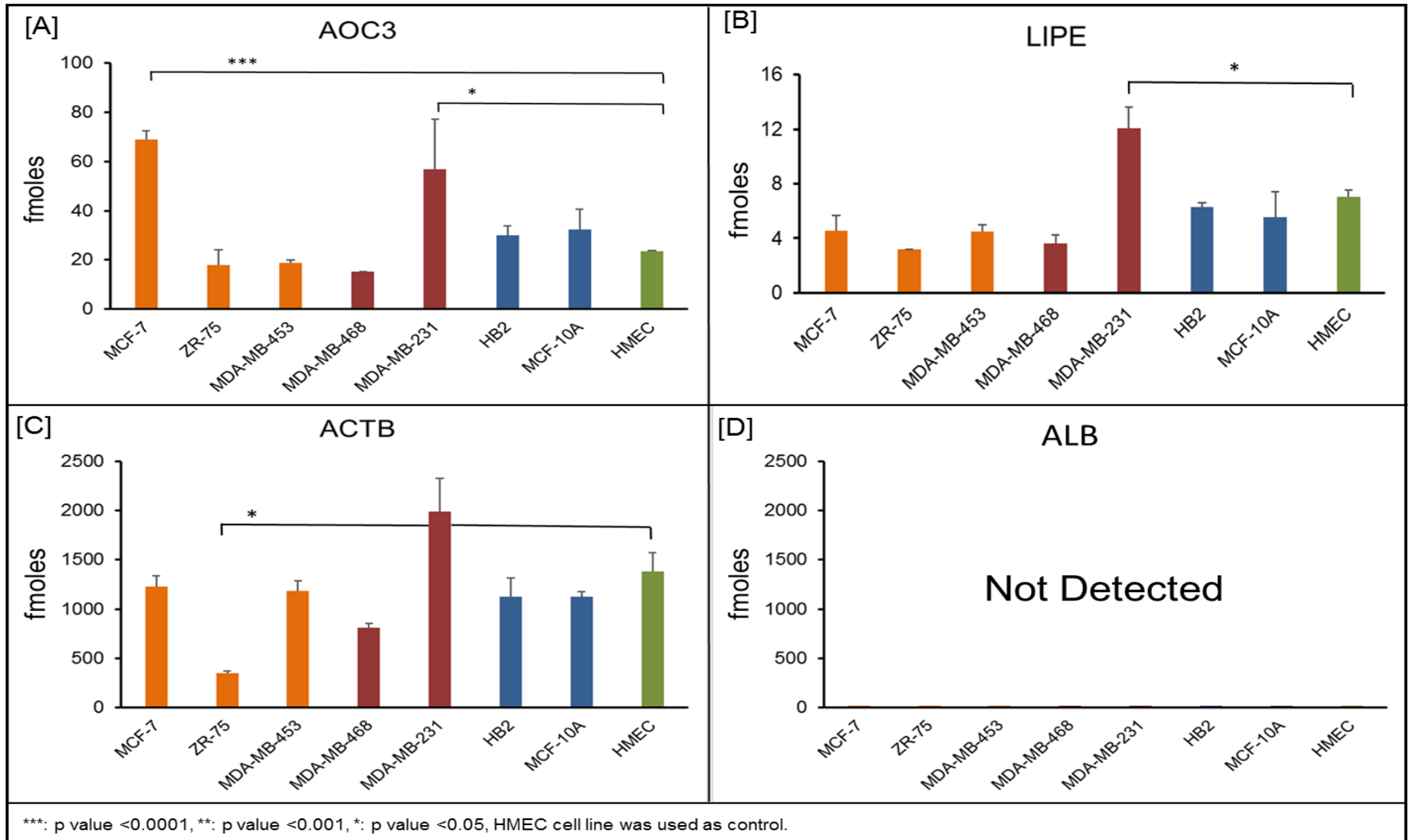


Figure 5-10: Expression of AOC3, LIPE, ACTB and ALB in breast cell lines.

5.4.4.2 Tissue biopsies

In this study, I used a panel of 8 breast tissue samples from 4 patients with breast cancer. From each individual, 2 biopsies were collected, one cancerous and the other non-cancerous as determined by histopathological examination. Among the 4 patients, 1 patient had fibroadenoma (PTT1), 2 were classified as having invasive carcinoma (IC) (PTT3 and PTT4), and 1 as mixed ductal carcinoma in situ (DCIS) and invasive carcinoma (IC) (PTT2). Detailed information about these tissues is provided in Table 5-4 and each tissue biopsy was measured in technical triplicates. I calculated the dynamic range of target proteins in breast tissues based on their prospective proteotypic peptides and calibration curves (Supplementary Table 5.2).

The average amount of three peptides of cofilin-1 was high in tumour tissues, collected from invasive carcinoma patient (PTT3) compared to matched normal tissue (p value <0.0001). but significantly decreased in fibroadenoma patient (PTT1) (p value 0.0027). (Figure 5-11). Two peptides of 10 kDa heat shock protein; HSPE1-1 and HSP1-3 were detected in all tissue biopsies while HSPE1-2 was only detected in PTT1 (fibroadenoma) and PTT3 (invasive carcinoma). The highest amount of HSPE1-3 (23.93 fmoles), was observed in PTT3-T while HSPE1-2 was also detected in same sample at a high amount of 7.00 fmoles (Supplementary data section 5.1). The average amount of HSPE1, was significantly higher in cancerous tissues of invasive carcinoma patient (PTT3, P value 0.0098) as compare to normal tissues (Figure 5-11, B). The expression of SH3 domain-binding glutamic acid-rich-like protein 3 (SH3BGR3), was significantly different between matched diseased and normal tissues for the fibroadenoma patient (PTT1) (p<0.004) where it was decreased, but was

increased in invasive carcinoma patients PTT3 ($p < 0.0001$) and PTT4 ($p < 0.05$)

Figure 5-11, D).

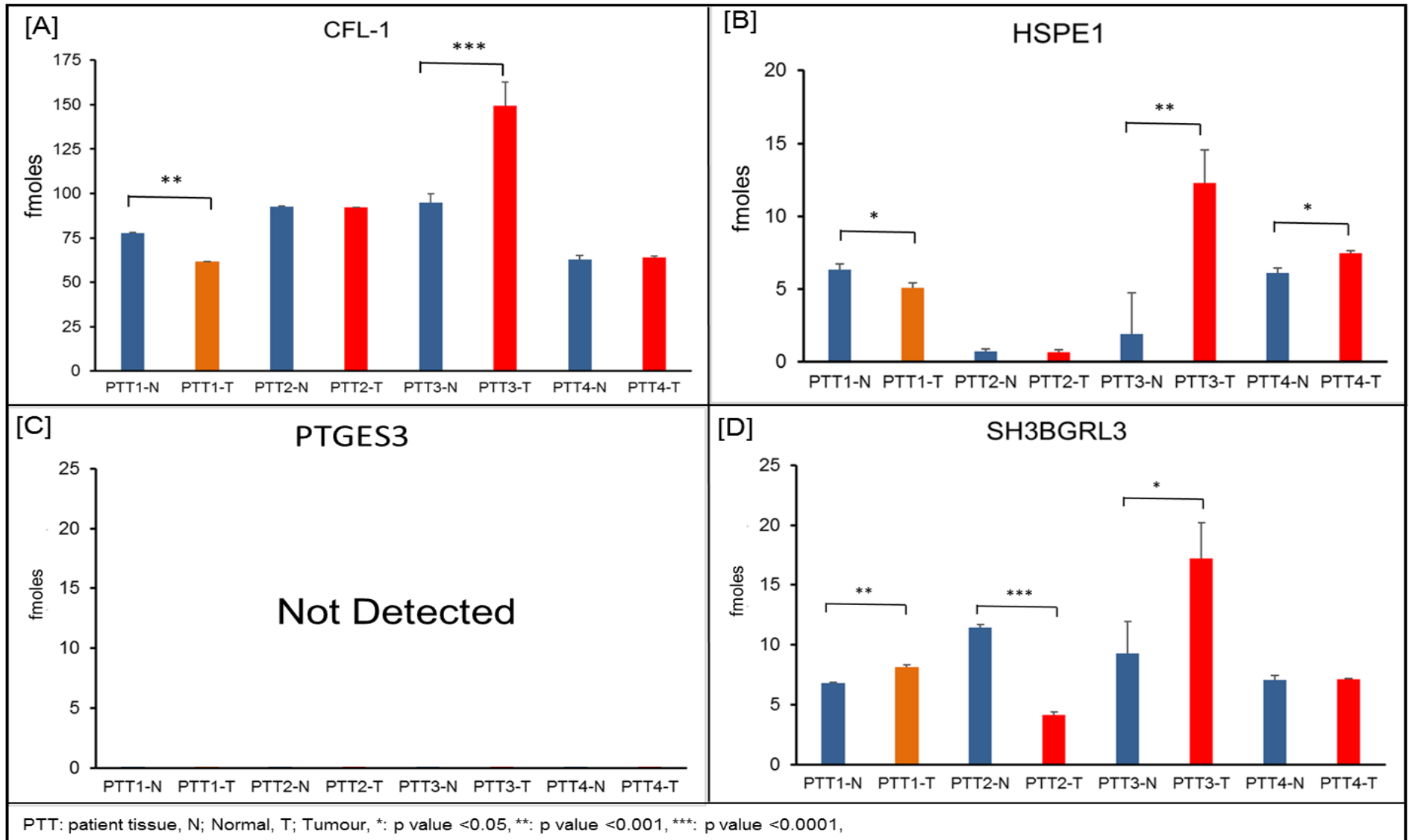


Figure 5-11: Expression of CFL-1, HSPE1, PTGES3 and SH3BGRL3 in breast tissue biopsies.

Hormone-sensitive lipase (LIPE) and Membrane primary amine oxidase (AOC3), were also quantified in all tissue sample. No significant difference was observed between normal and tumour tissues for both proteins (LIPE and AOC3), but expression level was different between patients (Figure 5-12, A, B, Supplementary table 5.2).

The amount of Actin, cytoplasmic 1 (ACTB), was high in PTT2, both normal (142.37 fmoles) and tumour (111.13 fmoles) sample but significant difference (p value 0.003), was only observed in PTT4 with amount of 9.20 fmoles in normal and 52.93 fmoles in tumour (Figure 5-12, C). The average dynamic range (LoD and LoQ) of Serum albumin in tissue biopsies was from 3.33 fmoles to 232.1 fmoles in triplicate analysis. In three patients, PTT1, PTT3 and PTT4, a significance difference was observed in matched normal and tumour pairs (Figure 5-12, D) (Supplementary table 5.2).

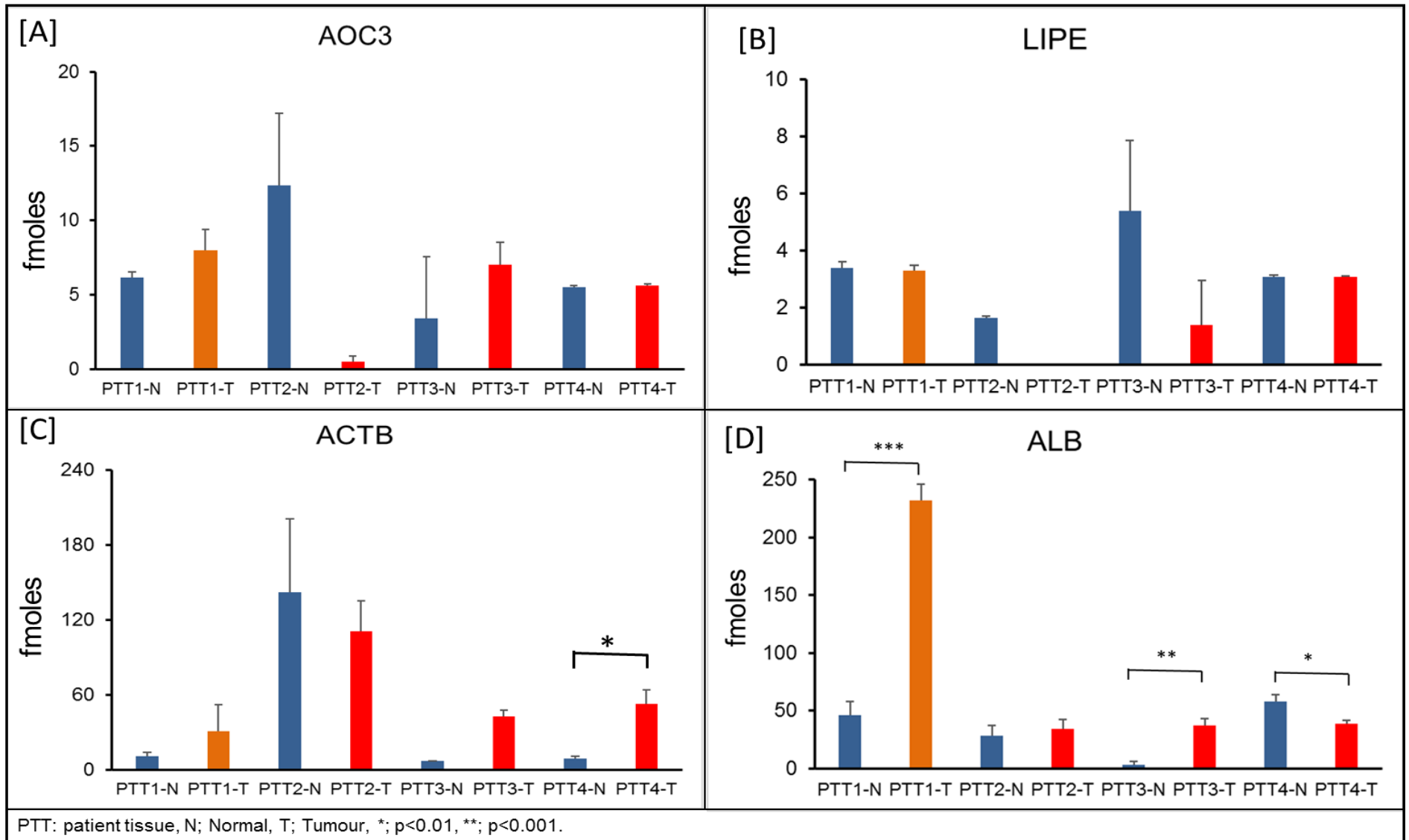


Figure 5-12: Expression of ACTB, ALB, LIPE and AOC3 in breast tissue biopsies.

5.4.4.3 Serum

In serum samples (Table 5-4), only 8 peptides for 6 target proteins were detected (Supplementary table 5.3). Furthermore, only single peptides were detected for cofilin-1, prostaglandin E synthase 3, and heat shock protein 10 - CFL1-2, PTGES3-1 and HSPE1-3 respectively. There was no significance difference between healthy volunteers and breast cancer patients, for Cofilin-1 and HSPE1 (Figure 5-13, A,B) and AOC3 (p value 0.4127, Figure 5-14, A). The amount of PTGES3 was high (p value 0.0357) in serum derived from cancer patients; 2.62 fmoles compared to healthy volunteers; 1.34 fmoles (Figure 5-13, C).

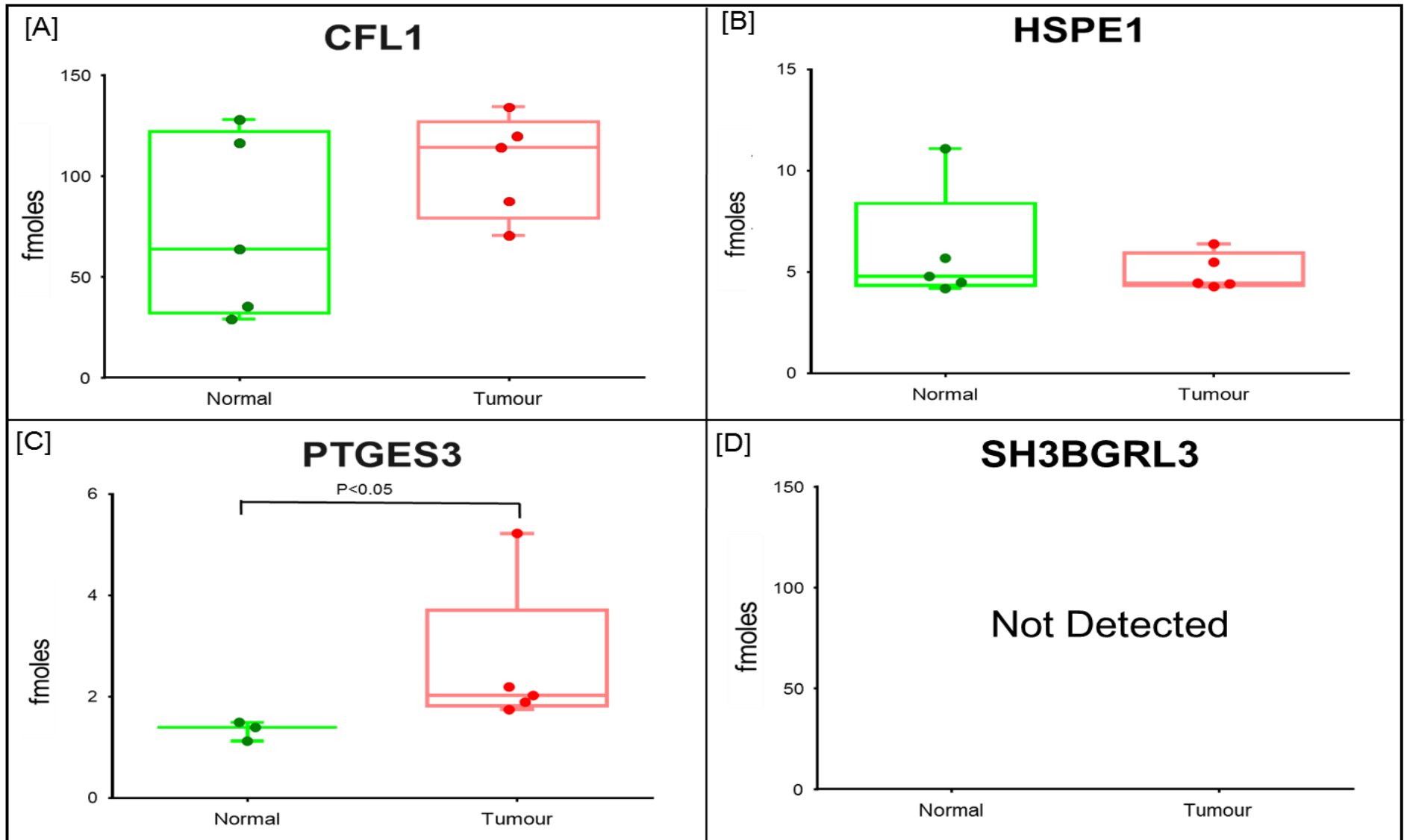


Figure 5-13: Expression of CFL1, HSPE1, PTGES3 and SH3BGRL3 in serum samples.

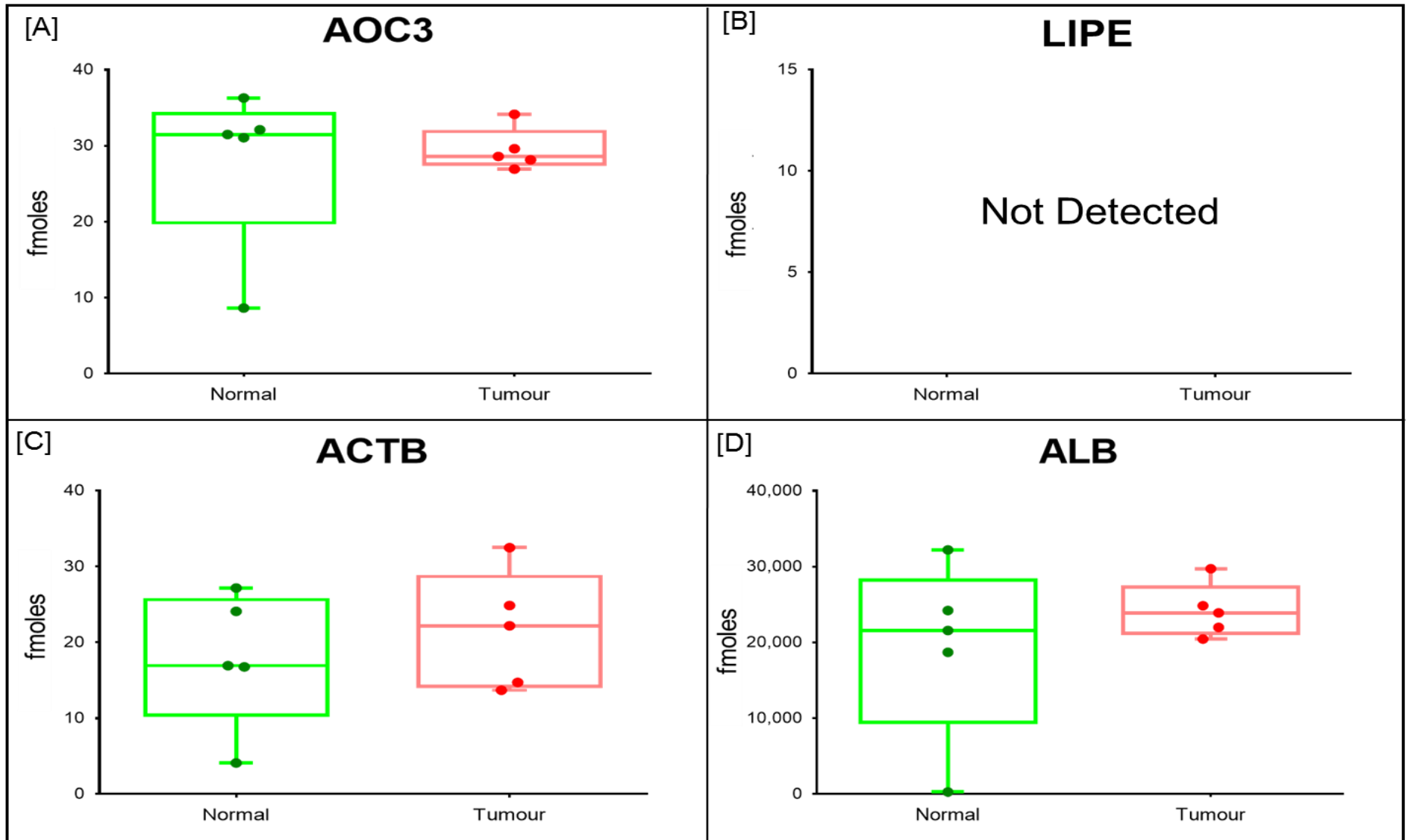


Figure 5-14: Expression of AOC3, LIPE, ACTB and ALB in serum samples.

5.4.4.4 Nipple aspirate fluid

Eight NAF samples were analysed, matched pairs collected from both breasts of 4 patients (Table 5-4). All samples were analysed in triplicate and the amount of target proteins, was calculated from calibration curves of standard peptides. Of a total of 18 peptides investigated 9 were detected above the LoQ, 13 were at the LOD and 5 were not detected. As a consequence, nine peptides representing 7 target proteins, were quantified (only AOC3 was omitted) in the NAF sample with MRM-MS assay (Supplementary table 5.4). The average amount of cofilin-1 was significantly high (p value 0.04) in breast with ductal carcinoma (PTN3) compared to normal breast (Figure 5-15, A). The average amount of HSPE1, was high (12.70 fmoles) in breast with ductal carcinoma; PTN3, as compare to normal breast (1.25 fmoles), with p value <0.0001 (Figure 5-15, B). The expression of PTGES3 was decreased significantly in disease breast of PTN3 (p value <0.0001) and PTN4 (p value 0.007) (Figure 5-15, C). SH3BGRL3-3, was detected and quantified in NAF, with decreased expression (p value 0.453) in breast with lobular carcinoma (PTN2) (Figure 5-15, D). The expression of serum albumin (ALB) was high (p value 0.002) in breast with fibrocystic changes (PTN1-L) as compared to breasts with cancer (Figure 5-16,D).

Highest amount (4.23 fmoles) of Hormone-sensitive lipase (LIPE), was observed in normal breast of PTN3, with p value <0.0001, compared to breast with tumour. (Figure 5-16,B). The average amount of actin, cytoplasmic 1 (ACTB) in NAF was from 14.27 fmoles to 123.77 fmoles in all NAF samples, with clear difference in left and right breasts. The significant difference was observed in PTN2 (pvalue 0.0196) and PTN3 (pvalue 0.0189) (Figure 5-16,C).

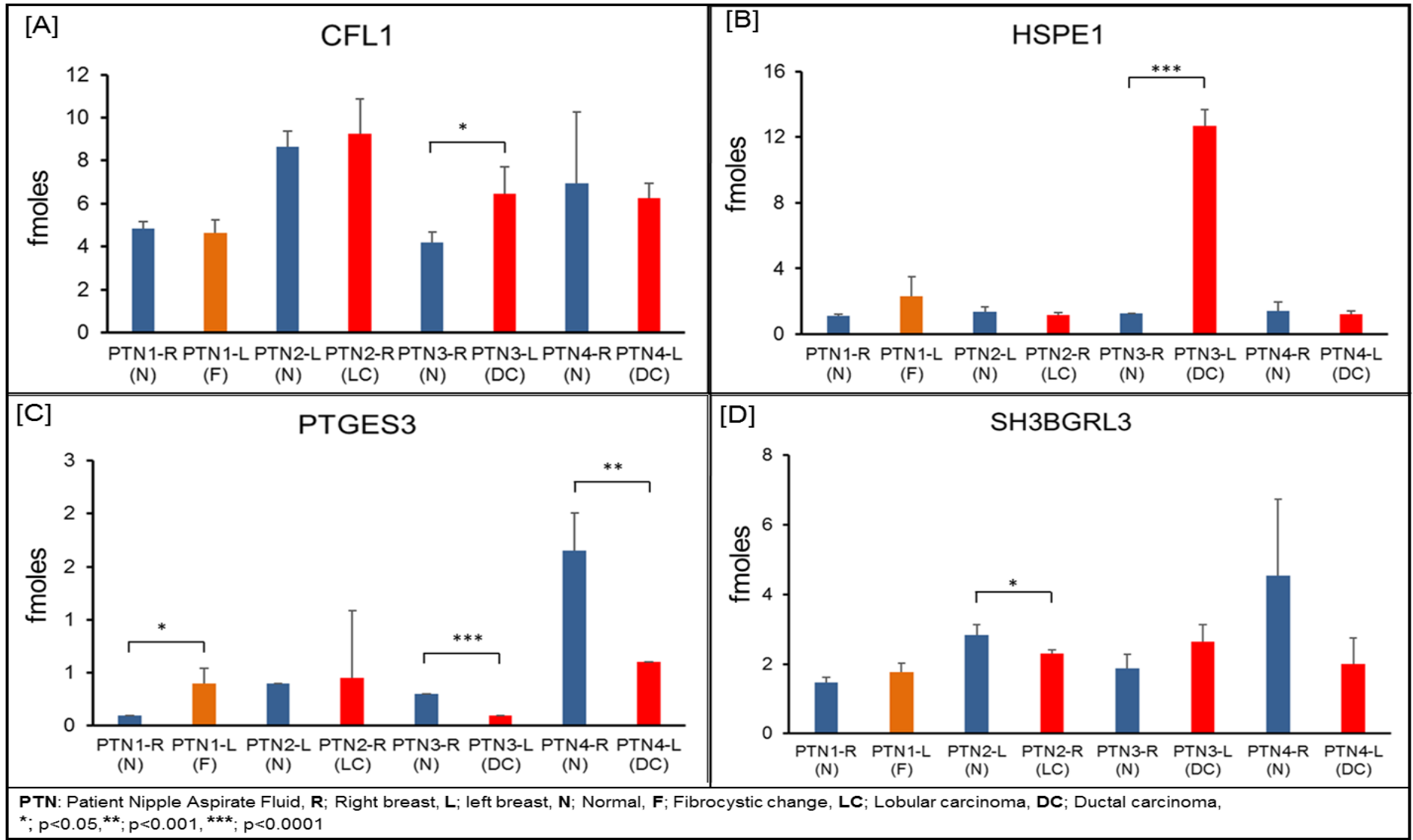


Figure 5-15: Expression of CFL1, HSPE1, PTGES3 and SH3BGRL3 in nipple aspirate fluid.

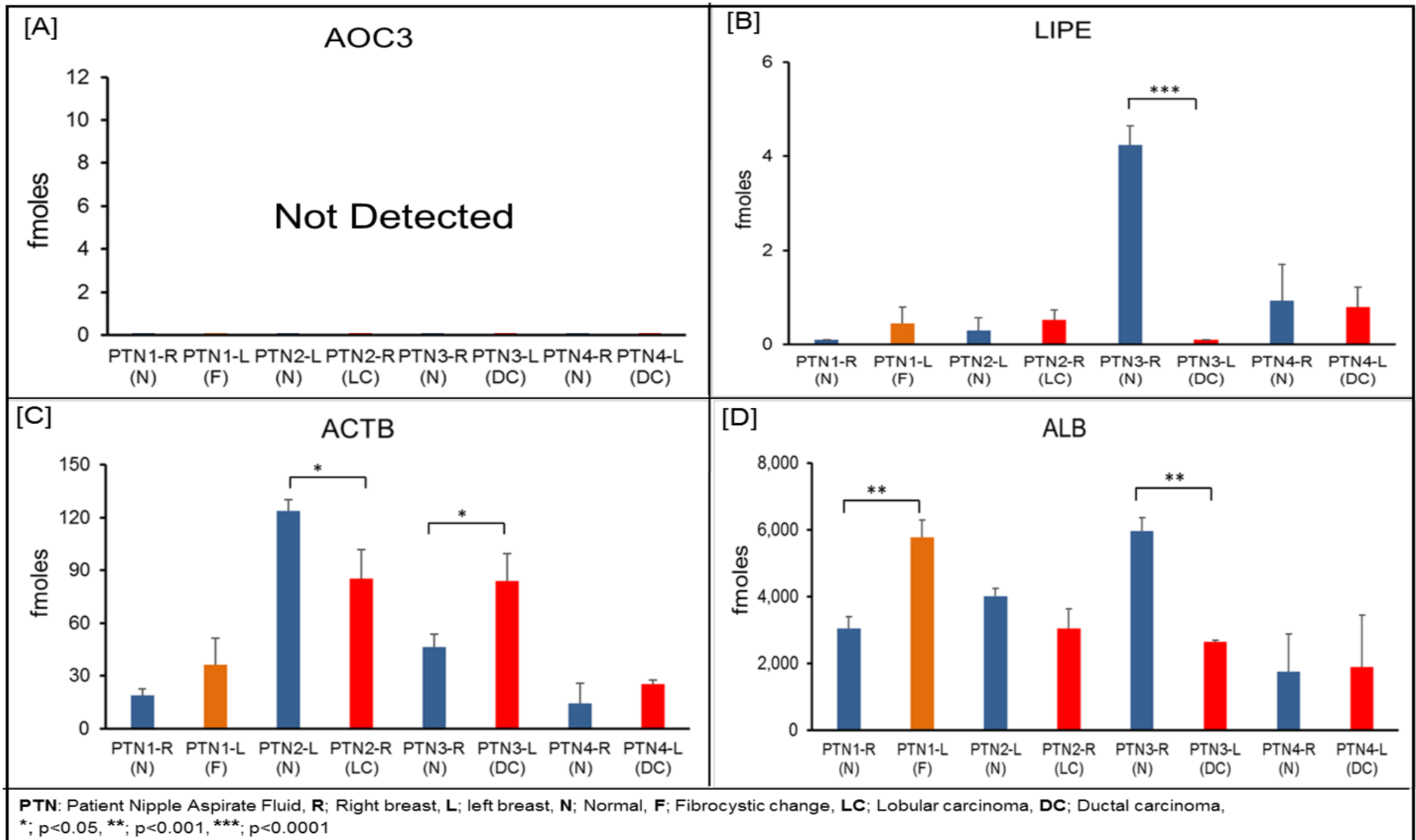


Figure 5-16: Expression of AOC3, LIPE, ACTB and ALB in nipple aspirate fluid.

5.4.4.5 MRM-MS assay coefficient of variation in biological samples

In breast cell lines, the CV (%) for target proteins was in dynamic range of 1% to 35% with median of 6% for all quantified proteins (Figure 5-17). The median CV for tissues was 10% but two proteins; AOC3 and LIPE, have higher CV of 55%, in one patient (PTT3). The dynamic range of CV of MRM-MS in serum samples, was from 3% to 74% but median for all target proteins was 17%. One peptide of LIPE had higher CV (%) for three NAF samples, which resulted 88%, 81% and 78% of CV for target protein (LIPE) in these samples. The average CV in NAF samples was 18% for all quantified proteins (Figure 5-17).

In this study, the multiplex MRM-MS assay for 8 target proteins, had LoD and LoQ in fmoles/1.25 μ g of lysate, with median CV% of 6% in breast cell lines and 10% in tissue biopsies. This sensitivity was achieved without depletion and enrichment of proteomics samples. A multiplex immuno-MRM assay developed by Schoenherr et al 2012, for quantification of estrogen receptor (ER) and human epidermal growth factor receptor 2 (HER2) levels in breast cancer cell lines and tissue biopsies, had LoD fmoles/mg of lysate and reproducibility of method was median coefficients of variation of approximately 10% in breast cell lines and 23% in tissue biopsies (Schoenherr et al., 2012)

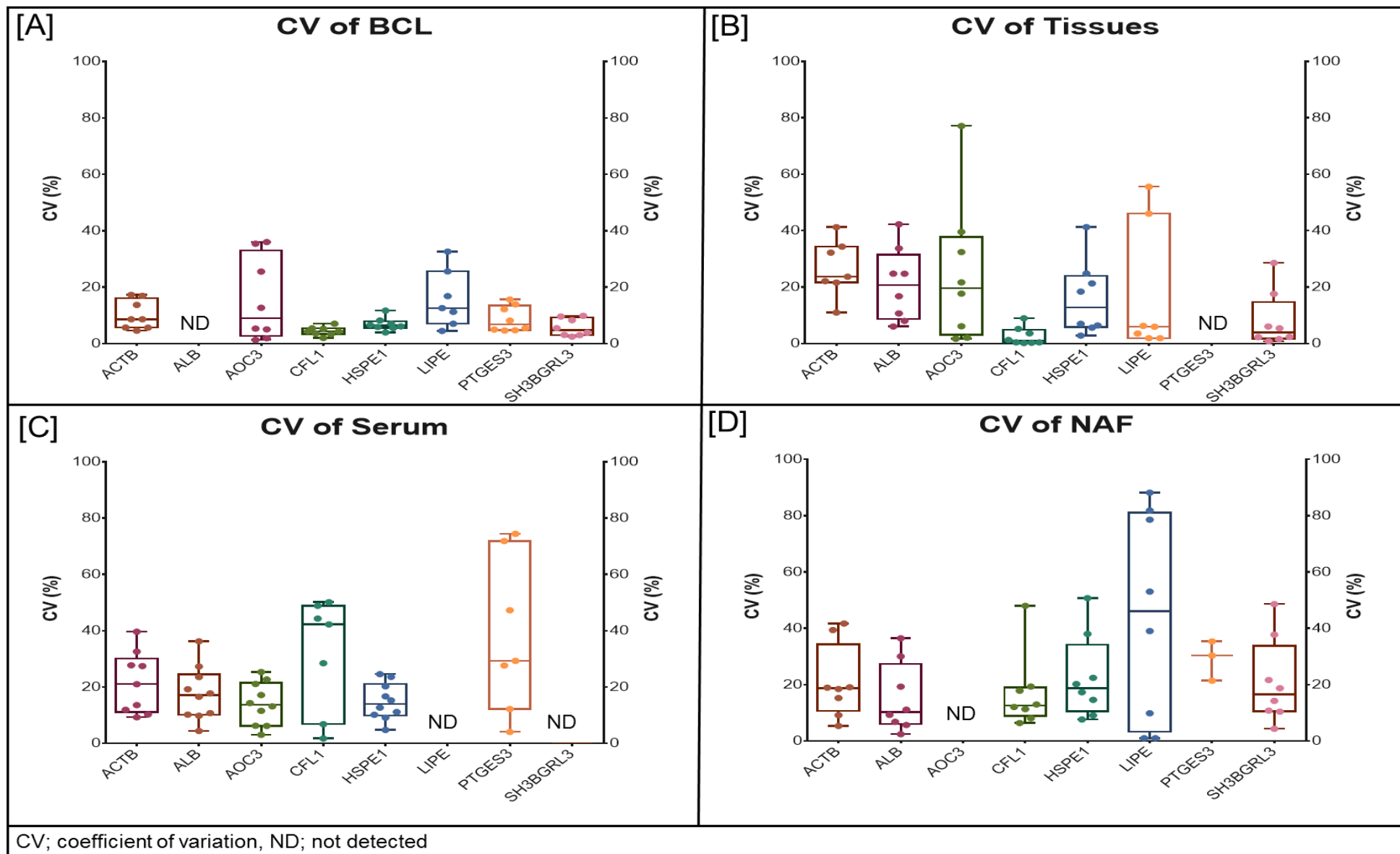


Figure 5-17: CV values for the multiplexed MRM-MS assays measured in breast cell lines (BCL), tissue biopsies, serum, and nipple aspirate fluid (NAF), consisting of targets (18 peptides, 8 proteins).

5.5 Discussion

Detection of secreted disease biomarkers in biological fluids by analytical methods is required to improve cancer diagnosis and prognosis in a clinical setting. These methods should be sensitive, robust, and rapid, as well as multiplexed assays. The current paradigm in protein biomarkers studies is that a panel of proteins may provide increased diagnostic accuracy compared to a single protein biomarker (Addona et al., 2011, Neagu et al., 2011). We, and others, have found that a multiplexed assay has a higher diagnostic and prognostic value than screening for individual protein biomarker (Domanski et al., 2012a, Chen et al., 2012, Shaheed et al., 2013a). Based on these findings, a targeted, multiplexed LC-MRM/MS approach was developed, with synthetic peptides for the quantitation of a panel of candidate breast cancer biomarker proteins in un-depleted and non-enriched human cell lines, tissue biopsies, serum and NAF.

The best method for comparison of peptide or protein in a biological system, is stable isotope (SI) dilution, in which a stable isotope-labelled synthetic peptide is used as an internal standard for each target peptide. SI-MRM assay provides the best analytical performance and characteristics for all targeted peptide, with typical linearity over 4–5 orders of magnitude, coefficient of variation (method reproducibility) <20%, and limits of detection (LoD) in cell and tissue lysates in the range of femtomoles per milligram of protein (Addona et al., 2009, Keshishian et al., 2009, Zhang et al., 2011). The main disadvantage of SI-MRM assays is the cost and high-purity isotope-labelled peptides for a multiplexed assay (Liebler and Zimmerman, 2013). QconCAT polypeptides provide a more economical way to generate heavy

isotope-labelled peptides in a multiplex assay (>100 target proteins) (Chen and Turko, 2014, Brownridge et al., 2012). An advantage of using QconCAT polypeptides is to minimise the sample processing variations, that can be introduced due to trypsin digestion, because QconCAT reference standards and biological samples are mixed together before digestion step (Brownridge et al., 2012). However, QconCAT technology is more complicated; (i) selection of target peptides, (ii) synthesis of genes for target peptides (iii), sub-cloning the genes in to an expression vector, (iv) expression of heavy isotope labelled polypeptide, and time consuming construction (2-3 months) (Meng and Veenstra, 2011, Chen and Turko, 2014). The need for a more cost-effective multiplex strategy for studies involving large number of peptides or proteins led to the development of a “light” peptide approach, in which calibration curves of standard peptides are prepared before and after measurement of trypsin-digested biological samples. In this latter case, the cost of synthetic peptides standards is considerably less for the high purity (>99%) products and in amounts suitable for assay development (1 to 4 mg) (Zhang et al., 2011, Liebler and Zimmerman, 2013). The light peptides based (LP) MRM assay is the best choice in such situation, as this assay can be optimised quickly without the cost and delay involved in obtaining the SI labelled peptides.

LP-MRM approach seems straight forward but precise quantification of target peptides can be challenging because samples and standards peptides are analysed separately on LC-MS. So, variations introduced due to LC-MS fluctuations (ionisation and solvents mixing) or background matrix of samples, result in ion suppression or enhancement of co-eluting analytes from one analysis to the other and also within one LC-MS analysis (Lange et al., 2008). Despite these limitations, LP-MRM quantification can be performed with high

accuracy, if the sample processing is well controlled and the samples are closely related in protein composition and background matrix (Wasinger et al., 2013, Liebler and Zimmerman, 2013).

To develop a multiplex MRM assay, the performance of the three Waters triple-quadrupole mass spectrometers; Xevo TQD, Premier XE and Quattro Ultima, for 18-target peptides was compared. All target peptides were analysed separately to optimise LC-MS conditions. To increase the multiplexity of MRM-MS assays, a timed acquisition mode, termed scheduled MRM acquisition, can be performed to increase the number of peptides analysed in a single chromatographic run, while maintaining the highest degree of sensitivity and MRM transition duty cycle (Stahl-Zeng et al., 2007). In scheduled MRM analysis, the transitions of target peptides were acquired only during a defined elution time window. Hence, the retention time (RT) of each peptide must be established and the LC conditions must be stable and reproducible. In optimised MRM-MS assay, all transitions of a target peptide were acquired within two minutes of elution time window (Figure 5-4).

Quattro Ultima is operated with MassLynx NT, which allowed only monitoring of 30 MRM transitions in a single method (WatersCorporation, 2017b), so I created two methods; 9 peptides or 27 transitions for each method. In that case, Quattro Ultima consumed double time and samples compare to Xevo TQD and Premier XE. The T- Wave (travelling Wave) feature in the Xevo-TQD and Premier XE instruments, improved the sensitivity, selectivity and speed of analysis on both instruments compared to the Quattro Ultima (WatersCorporation, 2017a). The Waters Xevo TQD also featured; IntelliStart Technology, digital detector with dynamic range of 4×10^6 , and scan speed up

to 10,000 Da/s, which provided the high speed and high quality data for qualitative and quantitative applications (WatersCorporation, 2017a). In the optimised MRM-MS assay developed here, the dynamic range of standard peptides at 20 pmoles/ μ L, was 1×10^6 for HSPE1-1 while for all other peptides was an average 5×10^5 on Xevo TQD. The average dynamic range of Quattro Ultima and Premier XE, was 1×10^4 and 1×10^3 respectively, for the same concentration of standard peptides. The calibration curves of target peptides generated on Xevo TQD, had better LoD and LoQ as compare to Premier XE and Quattro Ultima. A limitation of Quattro Ultima and Premier XE was the relatively low resolution of precursor m/z measurements, which may allow interference from background contaminants in complex mixtures. Both instruments were also limited by their duty cycle, the rate at which transitions can be acquired with an acceptable signal-to-noise ratio (WatersCorporation, 2017a, Liebler and Zimmerman, 2013).

A plasma study conducted by the National Cancer Institute (NCI), across multiple laboratories and instrument platforms, found that multiplexed SIMR-MS based assays were highly reproducible and transferable across laboratories, achieving CV's in the range of 10–25%. The LoD and LoQ values were in the high hundreds of nanograms per millilitre to low micrograms per millilitre concentration ranges in unfractionated plasma samples (Addona et al., 2009). In this study, the reproducibility (median CV) of LP-MRM-MS assay in serum and NAF sample, was 17% and 18% respectively. The reproducibility (CV%) of MRM assay can be improve by using stable isotope labelled peptides and increasing the LC-MS analysis time or decreasing the amount of sample being analysed to minimise the interference of background matrix (Doerr, 2013).

The Xevo TQD-optimised MRM-MS assay for breast cancer-associated biological samples was evaluated. In breast cell lines, total 7 proteins were quantified with 13 peptides, in all eight cell lines in triplicate analysis. The sensitivity of assay was highest in tissue biopsies; 14 peptides were above LoQ that represented 7 target proteins. In NAF samples more peptides and proteins were quantified compared to serum sample (Figure 5-18). NAF is secreted continuously from breast's ducts to maintain the normal physiology of breast (Sauter et al., 1997a). Thus, protein biomarkers secreted from the primary tumour have high concentration in ducts before travel in the lymphatic and/or blood systems (Gilbey et al., 2004), which mean NAF represent a true micro environment of breast physiology. In our MRM-MS multiplex assay, more proteins and peptides were detected in NAF than in serum (Figure 5-18), which make it an ideal biofluid, for clinical verification and validation of protein biomarkers in breast cancer patients.

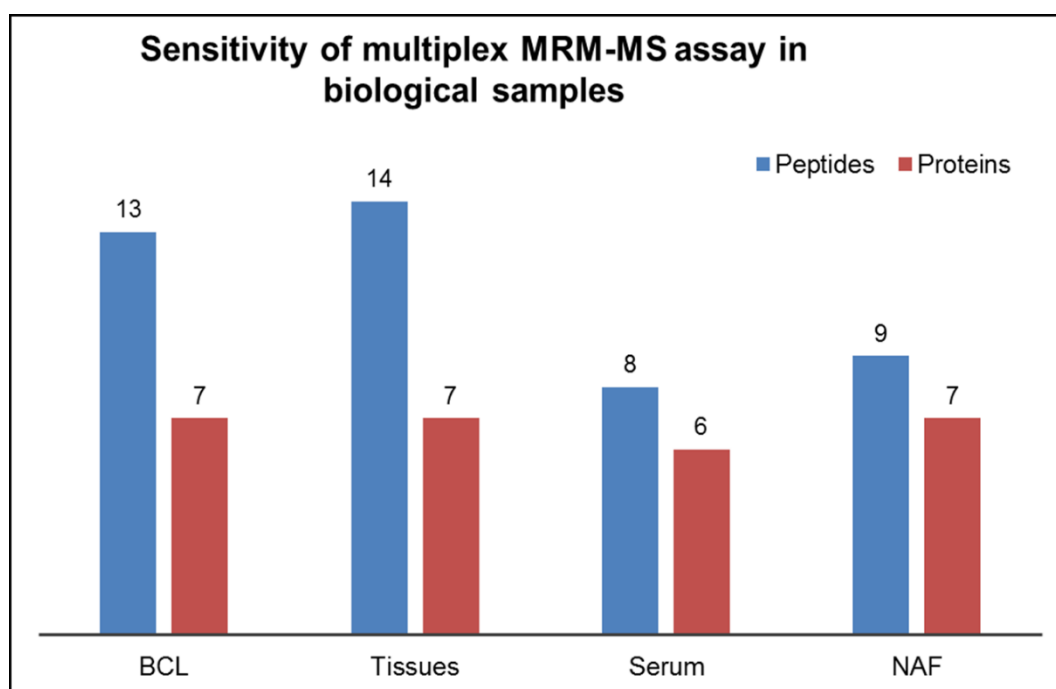


Figure 5-18: Sensitivity of multiplex MRM-MS assay in breast cell lines (BCL), tissues biopsies, serum, and nipple aspirate fluid (NAF).

One peptide of PTGES3; LTFSCCLGGSDNFK, selected because of the limited availability of unique peptides, but was less than ideal. As part of the Mudpit strategy, protein samples were reduced with dithiothreitol and treated with iodoacetamide to modify cysteine and prevent recombination of disulphide bridges of cysteine, (Weerapana et al., 2010). As cysteine is very reactive and prone to oxidation, it may have been modified prior to sample preparation, changing the peptide mass and preventing it from being detected in biological samples.

The sensitivity of the MRM-MS assay depended on several factors, including sample type (e.g., biofluids, tissue, cell line, and immune-precipitates) and whether any enrichment or fractionation techniques were incorporated prior to MRM analysis. In MRM-MS assay, the high abundant co-eluting peptides in complex mixture scans can suppress transitions of a targeted peptide. The fractionation of peptides with strong cation exchange chromatography and depletion of high abundance proteins by immunoaffinity, can increase the LoD and LoQ of a MRM-MS assay in more complex samples like tissues and serum but extra steps can lead to greater technical variation (Keshishian et al., 2009, Fortin et al., 2009).

Cofilin-1 plays an important role in breast cancer invasion and metastasis (Wang et al., 2007). The expression of cofilin-1 is increased in breast cancer tissue biopsies compared to normal (Wang et al., 2007, Shaheed et al., 2013a). A study conducted, on 30 lung cancer patients and 30 healthy volunteers, found high expression of CFL1 in serum collected from patients with stage III and IV lung cancer compared to stage II patients and healthy volunteers (Zheng et al., 2013). In this study, I found that the amount of cofilin-1 was high

in tumour tissues collected from invasive carcinoma patients, while no significance difference was observed in serum samples collected from breast cancer patients and healthy volunteers. One matched pair of NAF samples; PTN3, collected from ductal carcinoma patients, had high expression cofilin-1 in diseased breast as compare to normal breast. Cofilin-1 showed increased expression in diseased sample compared to normal and can used as a potential biomarker for early detection of breast cancer (Sato et al., 2017) and a large-scale study will provide more confidence about its sensitivity and specificity.

The co-chaperone PTGES3 regulates estrogen receptor (ER) and promotes tumour cell adhesion and invasion (Oxelmark et al., 2006). In our pervious study, along with other published research, we reported the increased expression of PTGES3 in breast cancer (Simpson et al., 2011, Shaheed et al., 2013a). PTGES3 was also detected with high concentration in all serum samples derived from breast cancer patients. But in nipple aspirate fluid, the amount of PTGES3 was decreased in breasts with ductal carcinoma as compare to healthy breast. The altered expression of PTGES3 in cell line, tissue, serum and NAF samples, suggested its status as biomarker may be specific for Luminal type breast cancer (Katsyv et al., 2016). Chaperone protein; HSPE1, was reported to decreased in basal type cancer as compare to luminal type breast cancer (Coumans et al., 2014). High expression of HSPE1 was also observed in diseased tissues and NAF, collected from two invasive carcinomas and one ductal carcinoma patients respectively. So, increased expression of HSPE1 could be potential biomarker for Luminal and claudin-low type breast cancers (Lianos et al., 2015).

The increased expression of SH3 domain binding glutamic acid-rich protein like 3 (SH3BGRL3) promotes the epithelial-mesenchymal transition, cell migration, and proliferation in EGFR+ bladder cancer (Michalak et al., 2016). In this study, we found higher expression of SH3BGRL3 in ER+ breast cancer cell lines (MCF-7) and no significant expression difference was observed in EGFR+ breast cell lines (MDA-MB-453, MDA-MB-468 and MDA-MB-453). The expression SH3BGRL was decreased in NAF samples collected from diseased breast compared to healthy.

The expression of AOC3 and LIPE had been reported to decrease in invasive lobular and ductal carcinomas of the breast (Zhao et al., 2004, Kim et al., 2015). Low expression was also observed in two NAF samples collected from patients lobular and ductal carcinomas, suggesting both proteins as potential biomarkers for breast cancer. The expression of albumin (ALB) was 6-fold less in NAF samples compared to serum and showed no relationship to breast cancer. This observation suggested, less plasma content in NAF samples, therefore enabling the omission of immunodepletion to see less abundant proteins, and making NAF a better option for verification and validation of protein biomarkers. The control for cellularity, ACTB (Guo et al., 2013), had highest amount in breast cell lines average of 1.2 pmoles in 1.25 μ g of cell lysate while in NAF was quantified as 40 fmoles indicating a lower level of cellular components in NAF samples.

5.6 Conclusions

In this study, a rapid and robust MRM-MS assay was established using synthetic standard peptides for the multiplexed quantitation of 8 candidate proteins in total extracts from human breast cell lines and un-depleted (without removal of the most abundant proteins such as serum albumin) tissue biopsies, and liquid biopsies serum and nipple aspirate fluid. The optimised protocol for MRM-MS of cofilin-1, PTGES3, HSPE1, SH3BGRL3, AOC3 and LIPE in breast cancer biopsies (cell lines, tissue, serum and NAF) described here provides an approach for a large-scale investigation of co-relation expression (i.e. biomarkers amount vs patient's characteristics; grade and histological type and/or tumour differentiation status, lymph node status, metastasis, etc.). Moreover, its implementation in a retrospective cohort and the results showed here are a significant step toward the validation process of these proteins as prognostic biomarkers. Despite the high-to-moderate abundance levels of these target proteins in nipple aspirate fluid, they remained "putative" biomarkers that need further verification and validation before clinical use. Addition of other diagnostic and prognostic biomarkers in to this multiplex assay from the NAF discovery work (Chapter 4), could help to improve the sensitivity and specificity of an assay for breast cancer detection. More proteins and peptides were detected and quantified in NAF compared to other samples. Nipple aspirate fluid was the preferred sample source for investigation of breast cancer since its collection is non-invasive and has the potential to provide a comprehensive picture of the physiological status of an individual through comprehensive quantitative proteomic approach.

CHAPTER 6. FINAL DISCUSSION AND FUTURE DIRECTIONS

6.1 Discussion

Proteomics has a potential value to study the structure and function of proteins in complex biological samples. Large scale quantitative proteomics analysis to understand cellular function is now possible due to developments in MS, computational algorithms and biochemical technologies. Advancements in protein fractionation and labelling techniques have also improved identification of less abundant proteins in biological samples. Studies characterising breast cancers prior to these investigations were limited by technological capabilities or only focused on the proteome of tissue biopsies and immortalised cell lines (Geiger et al., 2012, Lawrence et al., 2015) for diagnostic and prognostic biomarkers of breast cancer, not truly convenient for the clinical environment. Because of heterogeneity of breast cancer, it is challenging to differentiate between each subtype on the basis of a single gene or protein. Therefore, a group of markers may serve as a more specific and sensitive signature for diagnosing different types of breast cancer. The application of some of the latest advancements in MS proteomics (Orbitrap Fusion, MRM-MS), to carefully acquired breast cancer biopsies, has provided data-rich characterisation of the disease.

Before identifying phenotype-specific markers, one of the fundamental property of iTRAQ quantification, the dynamic range compression of reporter ions, was investigated. (**Chapter 3**). Orbitrap Fusion, using SPS mode for isobaric tagging experiment (MS^3 for reporter ion measurements), exhibited improved dynamic range due to reduced background chemical noise, compared

to the Ultraflex II. Furthermore, the Orbitrap Fusion acquired significantly more peptide species per unit time and with greater accuracy making it the instrument of choice for quantitative proteomic analysis.

We use the iTRAQ data to examine the functional networks of proteins within the established breast cancer subtypes. We reasoned that analysis at the protein level, rather than genes and transcripts, may more directly reflect microenvironment of breast cancer. There is a low correlation between the copy number of gene and relative expression at proteins because many genomic variations are not or only partially translated to the protein level (Geiger et al., 2010, Zhang et al., 2014a, Tyanova et al., 2016). The correlation between mRNA and protein level is also not clear, therefore only studying mRNA does not necessarily reflect active cellular processes and functions (Nagaraj et al., 2011, Schwanhäusser et al., 2011). From our proteomics results of phenotypically well-characterised breast cancer cell lines, 2467 proteins were identified on both instruments (Orbitrap Fusion and Ultraflex II), of which 1430 were common. We demonstrated that analysis at the proteomics level could differentiate coherent changes in cellular and molecular processes, and identify key molecular networks associated with each subtype of breast cancer. To that end, protein signatures were identified for five different phenotypes of breast cancer; luminal A (MCF-7, 34 proteins), luminal B (ZR-75, 42 proteins), HER2 (MDA-MB-453, 40 proteins), basal like (MDA-MB-468, 91 proteins) and the largest signature (499 proteins) associated with the stem cell-derived cancer cell line, MDA-MB-231 (Chapter 3, section 3.3.5). Previously, direct comparisons of cancer phenotypes and protein levels in large-scale proteomics projects have shown signature proteins in cancer cell lines (Nagaraj et al., 2011, Tyanova et al., 2016, Geiger et al., 2010). But these studies were more focused

on comparison within transfected cell lines. Our study, which was a comparison of HMEC against cancer cell lines, showed that that expression levels of proteins are not directly related to breast cancer. In particular, the proteomic data recapitulated the down-regulation of 56 proteins including ALDH1A3, ALDH1L2, KRT15, KRT16, KRT17, KRT6B, and up-regulation of 18 proteins include TPD52, CDK1, MCM2, MCM3, MCM5, in all breast cell lines, normal and tumorigenic, compared to HMEC. Hence, it is possible that expression of these proteins was not related to breast cancer, but rather to cell culture conditions or generic cell immortalisation processes. Therefore, selection of the right control was important to determine the proteomic signatures of breast cancer subtypes. Our iTRAQ based data constituted a system wide, quantitative view of the proteomes of the breast cancer cell lines, which served as the basis for further downstream biological interpretation and clinical validation using biopsy samples.

Breast secretions (nipple aspirate fluid [NAF]), as a potential diagnostic value in breast cancer were investigated (Chapter 4), as they are produced by epithelial cells lining breast ducts and lobules which are most commonly associated with the disease. Matched pairs of NAF were characterised for protein amount and SDS PAGE profile. SDS PAGE showed that most matched pairs have similar protein band patterns, which was corroborated by the high correlation of proteomics profiles. Four pairs from a healthy individual, and patients with benign phyllodes, ductal carcinoma in situ (DCIS) and invasive carcinoma, were analysed with an Orbitrap Fusion mass spectrometer using a label-free quantitation strategy (Chapter 4). We have identified more than double the number of proteins previously detected in NAF making this the most complete proteomic study to date. Within the NAF proteome, were 300 proteins

not found in plasma and 24% of the proteins currently part of the NCI Early Detection Research Network studying breast cancer (<https://ednr.nci.nih.gov>). The presence of drug metabolising enzymes in NAF, such as CYP1A1, could provide useful prognostic markers for the patient's responsiveness to chemotherapy.

The NAF database was explored for proteins that had previously been identified as members of the phenotype-specific clusters, to determine if NAF could provide indicators of cancer phenotype as well as stage (Chapter 3). A total 784 proteins were common in NAF and breast cancer cell lines, including 324 associated with our specific breast cancer phenotype signatures (Table 6-1). Four luminal A (FBP1, KRT18, KRT8, and S100A16) 10 luminal B, 7 HER2-type, 17 basal-like and 64 claudin-low signature proteins, were identified in all NAF samples.

A total 784 proteins were common in NAF and breast cancer cell lines, including 324 associated with our specific breast cancer phenotype signatures. In NAF, there was also 56 proteins which differentiate between luminal and basal type of breast cancer (Chapter 3, Section 3.3.5.1), including CD44 (already a developed biomarker for colon and breast cancer (Basakran, 2015)) and Von Hippel-Lindau disease tumour suppressor (VHL, associated with inherited familial cancer syndrome; retinal, cerebellar and spinal hemangioblastoma, as well as renal cell carcinoma and pancreatic tumours) (Stolle, 2016). Analysis of these potential phenotype markers would provide an excellent focus for further validation.

		Breast cancer clinical stages				Common Proteins	Gene Name
		HV	PB	PD	PI		
Breast cancer phenotypes	Luminal A	5	11	5	11	5	FBP1, KRT18, KRT8, S100A16
	Luminal B	14	17	12	15	10	ASAH1, AHSG, CSRP1, EEF1G, ENDOD1, HSP90AB1, LDLR, PPT1, PPIC, CANT1
	HER2	10	16	12	10	7	ARPC1B, CALML5, CTSB, LGALS3BP, GSN, HSPB1, ACSL1
	Basal like	22	26	22	21	17	DBI, HEXB, CASP14, CAT, SLC44A2, PHGDH, GBA, GNA13, KRT1, KRT5, P4HB, PACSIN3, S100A13, S100A8, S100A9, GPNMB, TACSTD2
	Claudin low	103	126	94	97	64	RPS3, RPLP1, PGD, HSPA5, ARPC4, APRT, ARF4, ENO1, ANXA5, PEA15, ACLY, B4GALT1, CALM1, CALR, IGF2R, CRABP2, CHMP4B, COL6A2, CSTB, ACO1, DNASE2, DDB1, RPN2, EHD4, EEF2, HSP90B1, EPS8L2, EIF4A1, EIF5A, CAPZA1, FASN, FABP3, LGALS1, GSTO1, GNB1, HEBP2, HYOU1, IDH1, F11R, KRT77, LTF, GLO1, LDHA, MYL6, MTPN, AHNAK, FAM129B, PGAM1, PFN1, P3H1, RAB11B, RRAS, ARHGDI, RHOC, PPP2R1A, SOD2, VAT1, TXN, TPM3, TPM4, MYO1C, MYO6, VPS35, VIM
	Luminal vs Basal	79	101	71	81	56	ALDH9A1, NT5E, ARF6, AKR1B1, GLA, RNPEP, ANXA1, ANXA11, ANXA2, ANXA3, B2M, CBR1, COMT, CD44, CD59, CLIC1, CLTC, CPNE3, CAND1, CTSC, DPP7, EHD1, STOM, BLVRB, ALDOA, GGCT, G6PD, GSTP1, GNG12, GNAS, HSPA13, KRT17, KRT7, MFGE8, LDHB, CAPG, MSN, MYL12B, SLC9A3R1, LCN2, NAMPT, FKBP1A, PGM1, PAFAH1B2, PODXL, CUTA, TGM2, NPEPPS, PTPRS, SELENBP1, VCP, TMED10, UBA1, VASP, NSF, VCL

Table represents the expression of breast cancer phenotype signature proteins in NAF samples collected from HV; healthy volunteer PB; patient benign, PD; patient ductal carcinoma in situ (DCIS), PI; patient invasive carcinoma. The phenotype signature proteins detected in all NAF samples is presented as common proteins with gene names.

Table 6-1: Comparison of phenotype signature with clinical stages of breast cancer.

We also identified 183 proteins that are differentially expressed in the healthy volunteer compared to the patients with malignancies (benign, DCIS and invasive carcinoma) NAF samples, which may correlate with phenotype-specific markers (Table 6-2). Of these, 45 proteins had significantly different levels (t-test $p < 0.05$) in the healthy volunteer (both paired samples) relative to the patients with disease (3 matched pairs). These may be predictive biomarkers for different phenotypes of breast cancer and could be the very earliest measures of tumourigenicity. The NAF from the diseased patients have 138 potential phenotype biomarkers that were significantly changed compared to the healthy volunteer (7 for luminal A, 9 for luminal B, 11 for HER2, 14 for basal-like and 52 for claudin-low type). Therefore, these panels of phenotype-associated proteins could provide extra value in disease stratification, also in terms of disease stage. This provides an excellent focus for screening a larger series of NAF samples to define their role in breast cancer early detection, diagnosis and prognosis.

Breast cancer phenotypes	Healthy NAF		Cancerous NAF	
	Signature	Potential biomarkers	Signature	Potential biomarkers
	Luminal A	-	-	7
Luminal B	4	ENDOD1, NME3, PDCD4, CANT1	9	CSRP1, ASAH1, CORO1B, MGST1, ACP2, EEF1G, IDH2, CRIP2, VDAC3
HER2	1	GSN	11	HSPB1, ACSL1, LGALS3BP, CALML5, CLIC4, PDLIM1, GSTM3, ALDH4A1, CRAT, SORD, ECH1
Basal like	3	PYCARD, CASP14, P4HB	14	S100A8, S100A9, DBI, RPL18, S100A13, PAPSS2, TUBB4A, PRSS1, YES1, MME, TUBB4B, GNA13, SLC44A2, PADI2
Claudin low	22	RPLP2, HSPA5, APRT, ENO1, DHX9, B4GALT1, CALR, CRABP2, EEF1B2, FABP3, GLRX, HLA-A, LTF, MYL6, MTPN, YBX1, PLIN3, PTTG1IP, RRBP1, VAT1, TMEM263, MYO1C	52	NACA, PSMA7, GSR, AK1, RPLP1, RALB, IGF2R, EEF2, ARF4, RAB5A, ACO1, CRK, RPS3, ERP29, FASN, ARHGDI, HSPE1, ALM1, EIF5A, GFPT1, EPS8L1, ACSS2, FARSB, MYO6, SURF4, SEC22B, IDH1, SLC25A5, RHOC, PFKL, PAICS, SDHA, USP5, ACLY, COPA, EIF4A1, OSTC, S100A10, ACTR1A, RPS7, SSBP1, GLO1, AHNAK, DYNC1H1, KPNB1, ACOX1, RAB11B, CYFIP1, AACS, SLC44A1, HSPH1, VPS35
Luminal vs Basal	15	RPS21, RPS28, ALDH9A1, ARF6, RNPEP, CALU, CD44, CLIC1, DPYSL2, ALDOA, KRT17, MFGE8, RCN1, SELENBP1, SUMF2	45	KRT7, G6PD, ANXA3, CBR1, SERPINB1, RPL22, PGM1, NAMPT, VCP, LCN2, NSF, NQO1, AP1G1, CPNE3, GGCT, MTHFD1, APEH, SLC25A6, AKR1B1, COMT, UBA1, BLVRB, GDI1, ATIC, GALNS, SSR1, RPL6, ALDH1A3, SLC25A1, NPEPPS, SLC25A3, CLTC, AP1B1, DCTN2, CUL3, TKFC, SRI, GPD1L, GBF1, PDLIM5, EHD1, MCCC2, DNPEP, STARD10, AP1M2

Table represents the comparison of phenotype signature proteins against healthy and disease specific signature in NAF samples collected from HV; healthy volunteer PB; patient benign, PD; patient ductal carcinoma in situ (DCIS), PI; patient invasive carcinoma.

Table 6-2: The expression of phenotype specific proteins in healthy and cancerous NAF.

The collection of NAF is non-invasive compared to tissue biopsies and circumvents the limitations of plasma (e.g. 10 orders of magnitude in protein abundance). So NAF is a valuable source for biomarker discovery associated with origin of the disease and provides important information about cancer progression and specific to different phenotype of breast cancer.

The collection of NAF sample from the same donor can be challenging in expression and collection. This process can be improved to 90% of cases with the aid of oxytocin nasal spray (Suijkerbuijk et al., 2010), supported by application of manual or mechanical pumps normally used for milk expression by mothers with preterm infants (Jones and Spencer, 2007). Oxytocin is a peptide hormone

plays a role in child birth and with stimulation of the nipples from breastfeeding (Chiras, 2013).

NAF contained many breast cancer markers, growth factors and receptors that have been detected in plasma, but required a number of independent assays (immunoassays and mass spectrometry). We evaluated the effectiveness of three LC/MRM-MS platforms (Quattro Premier, Quattro Ultima and Xevo TQD) for the development of a single multiplexed assay, using a preliminary panel of 8 biomarkers (Chapter 5). Furthermore, we compared the MRM and iTRAQ results of selected putative biomarkers (CFN1, PTGES3, HSPE1) with Western blotting using breast cancer cell line extracts (Figure 6.1). In all three cases, the MRM and iTRAQ results exhibited very encouraging correlation in expression for each cell line relative to HMEC controls, but poor correlation with Western blotting. As Western blotting is based on completely different methodology this is not unexpected. The average of CV of MRM-MS multiplex assay was; 9% in cell lines, 17 % in tissue biopsies, 22% in serum samples and 24% in nipple aspirate fluid. Previous research has shown that light peptides-based (LP) MRM assays easily exceeded the performance of typical Western blotting, yielding CVs ranging from 10-40% for the former compared to 0-100% for the latter, for target peptides and proteins (Fortin et al., 2009, Picotti et al., 2013, Liebler and Zimmerman, 2013). To date this is the first study to develop a multiplexed MRM assay for phenotype and stage-specific breast cancer detection.

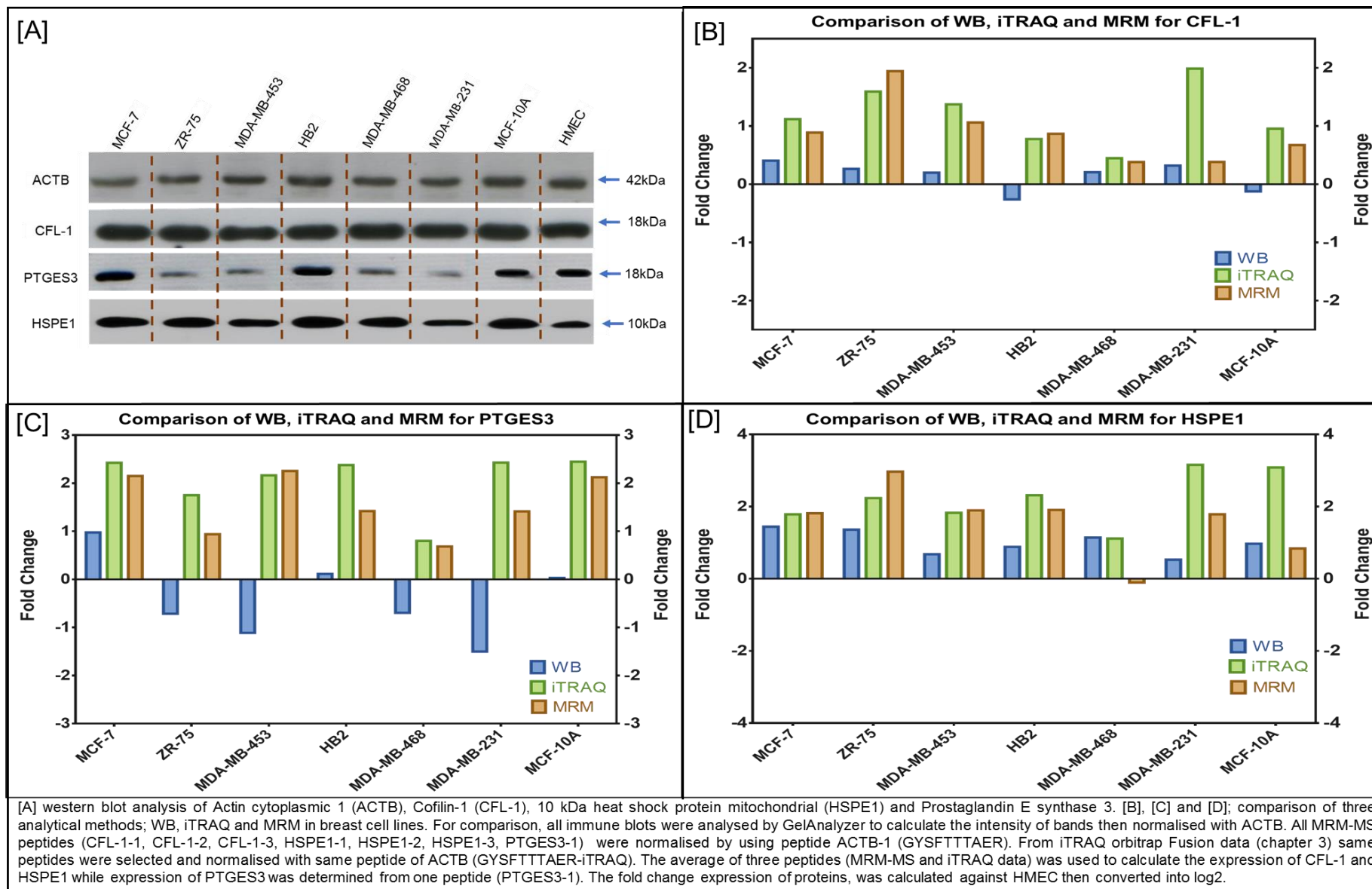


Figure 6-1: (A) ACTB, CFL-1, HSPE1 and PTGES3 Western blot analysis of phenotypic cell lines. Comparison of three analytical methods; Western blotting, iTRAQ and MRM-MS, for expression of (B) CFL-1, (C) HSPE1 and (D) PTGES3 in breast cell lines.

MRM-MS approach is more precise and accurate compared to traditional immune assays, for quantitative analysis of disease related biomarkers in complex biomaterials, such as blood plasma and tissues. MRM-MS based targeted protein assay, offers two main advantages over immune-assays; (i) ability to develop a specific assay for any protein or post-translationally modified protein form, without the requirement for an antibody, (ii) multiplex assay (analysis of many peptides representative of many proteins in a single assay). Targeted peptide-based protein assays provide high throughput, quantitative proteomic data and have common application in the systematic development of assays for protein biomarker verification and validation studies (Liebler and Zimmerman, 2013).

6.2 Future directions

We have here shown for the first time that global profiling of phenotype-specific breast cancer cell lines compared to primary cells, is possible with high quantification accuracy at Orbitrap Fusion, to define protein panels for each breast cancer phenotypes. The liquid biopsy approach using NAF provides a breakthrough in oncoproteomics to reveal novel diagnostic and prognostic applications. We have the potential to develop a multiplex MRM-MS assay using the Orbitrap Fusion to measure phenotype and stage-specific biomarkers using different clinical materials. The high degree of analytical and quantitative reproducibility which provides a solid basis for the extension of this analytical method to validate the findings presented in the thesis; -

- Use meta-analysis to refine the panel of phenotype- and stage-specific markers for multiplexed MRM-MS assay development
- Use tissue microarray of selected targets to correlation expression with phenotype, stage and outcome
- Incorporate stable-isotope-labelled synthetic peptide standards into the MRM assay to enable absolute quantification and improve assay reproducibility
- Investigate NAF samples for additional molecular (mRNA, miRNA [15] and biological (ductal cells, exosomes, microbe [16] signatures to complement proteomics profiling
- For the creation of a substantial NAF biofluid bank, various methods of collection should be explored with clinical, health and community partners and volunteers, such as manual or mechanical pumps used for milk and by oxytocin nasal spray (Suijkerbuijk et al., 2010, Jones and Spencer, 2007).

Finally, by combining the proteome profiling, miRNA analysis (Canto et al., 2016) and characterization of the microbiome (Chan et al., 2016a) in NAF, will improve our knowledge about the microenvironment of breast cancer, which will result in development of novel breast cancer detection and treatment strategies.

The use of specific fractionation schemes, increase the identified and quantified peptides and proteins. A systematic and efficient analysis of vast genomic and proteomic data sets is a major challenge for researchers today. To overcome limitations of current proteomics strategies with regard to the dynamic range of peptides, cost and time, alternative mass spectrometry-based approaches are being explored. Targeted strategies exemplified by DIA and MRM-MS detect, quantify, and possibly collect a product ion spectrum to confirm the identity of a peptide with much greater sensitivity because the precursor ion is not detected in the full mass spectrum (Doerr, 2015). A systematic and efficient evaluation of proteomics results requires (1) automatic retrieval of user defined customised database; (2) an intuitive graphical platform to display and analyse experimental data (3) efficient bioinformatics software tools for data interpretation, prediction of function, and modelling; (4) scalability and reconstruction of data base and bioinformatics tools (Chandramouli and Qian, 2009). Further technological developments like DIA and MRM-MS will enable a larger proportion of the proteome to be visualized and will further enhance our ability to characterize biological systems. As such, these advances in proteomics will impact diagnostic research and development.

CHAPTER 7. REFERENCES

- ADDONA, T. A., ABBATIELLO, S. E., SCHILLING, B., SKATES, S. J., MANI, D., BUNK, D. M., SPIEGELMAN, C. H., ZIMMERMAN, L. J., HAM, A.-J. L. & KESHISHIAN, H. 2009. Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nature biotechnology*, 27, 633-641.
- ADDONA, T. A., SHI, X., KESHISHIAN, H., MANI, D., BURGESS, M., GILLETTE, M. A., CLAUSER, K. R., SHEN, D., LEWIS, G. D. & FARRELL, L. A. 2011. A pipeline that integrates the discovery and verification of plasma protein biomarkers reveals candidate markers for cardiovascular disease. *Nature biotechnology*, 29, 635-643.
- ADRIANCE, M. C., INMAN, J. L., PETERSEN, O. W. & BISSELL, M. J. 2005. Myoepithelial cells: good fences make good neighbors. *Breast Cancer Res*, 7, 190-197.
- AEBERSOLD, R. & MANN, M. 2016. Mass-spectrometric exploration of proteome structure and function. *Nature*, 537, 347-355.
- AGGARWAL, K., CHOE, L. H. & LEE, K. H. 2006. Shotgun proteomics using the iTRAQ isobaric tags. *Briefings in functional genomics & proteomics*, 5, 112-120.
- AGGARWAL, N. & SLOANE, B. F. 2014. Cathepsin B: multiple roles in cancer. *PROTEOMICS-Clinical Applications*, 8, 427-437.
- ALEXANDER, H., STEGNER, A. L., WAGNER-MANN, C., DU BOIS, G. C., ALEXANDER, S. & SAUTER, E. R. 2004. Proteomic analysis to identify breast cancer biomarkers in nipple aspirate fluid. *Clinical Cancer Research*, 10, 7500-7510.
- ALLURI, P. & NEWMAN, L. A. 2014. Basal-like and triple-negative breast cancers: searching for positives among many negatives. *Surgical oncology clinics of North America*, 23, 567-577.
- ALORAIFI, F., MCDEVITT, T., MARTINIANO, R., MCGREEVY, J., MCLAUGHLIN, R., EGAN, C. M., CODY, N., MEANY, M., KENNY, E., GREEN, A. J., BRADLEY, D. G., GERAGHTY, J. G. & BRACKEN, A. P. 2015. Detection of novel germline mutations for breast cancer in non-BRCA1/2 families. *FEBS J*.
- ALTELAAR, A. M., MUNOZ, J. & HECK, A. J. 2013. Next-generation proteomics: towards an integrative view of proteome dynamics. *Nature Reviews Genetics*, 14, 35-48.
- ANDERSON, N. L., MATHESON, A. D. & STEINER, S. 2000. Proteomics: applications in basic and applied biology. *Current Opinion in Biotechnology*, 11, 408-412.
- ARIAZI, E. A., ARIAZI, J. L., CORDERA, F. & JORDAN, V. C. 2006. Estrogen receptors as therapeutic targets in breast cancer. *Current topics in medicinal chemistry*, 6, 181-202.
- ARONSON, J. 2005. Biomarkers and surrogate endpoints. *British journal of clinical pharmacology*, 59, 491-494.
- AVRIL, N., ROSE, C. A., SCHELLING, M., DOSE, J., KUHN, W., BENSE, S., WEBER, W., ZIEGLER, S., GRAEFF, H. & SCHWAIGER, M. 2000. Breast imaging with positron emission tomography and fluorine-18 fluorodeoxyglucose: use and limitations. *J Clin Oncol*, 18, 3495-502.
- AZAD, N. S., RASOOL, N., ANNUNZIATA, C. M., MINASIAN, L., WHITELEY, G. & KOHN, E. C. 2006. Proteomics in clinical trials and practice present uses and future promise. *Molecular & Cellular Proteomics*, 5, 1819-1829.
- AZIM, H. A., JR. & PARTRIDGE, A. H. 2014. Biology of breast cancer in young women. *Breast Cancer Res*, 16, 427.
- AZOULAY, S., LAÉ, M., FRÉNEAUX, P., MERLE, S., AL GHUZLAN, A., CHNECKER, C., ROSTY, C., KLIJANIENKO, J., SIGAL-ZAFRANI, B. & SALMON, R. 2005. KIT is highly expressed in adenoid cystic carcinoma of the breast, a basal-like carcinoma associated with a favorable outcome. *Modern pathology*, 18, 1623-1631.
- BAI, F., LIU, S. & WITZMANN, F. A. 2005. A "de-streaking" method for two-dimensional electrophoresis using the reducing agent tris (2-carboxyethyl)-phosphine hydrochloride and alkylating agent vinylpyridine. *Proteomics*, 5, 2043-2047.

- BALTZELL, K. A., WRENSCH, M. & SISON, J. D. 2006. A descriptive study of variables associated with obtaining nipple aspirate fluid in a cohort of non-lactating women. *BMC Womens Health*, 6, 15.
- BANTSCHIEFF, M., BOESCHE, M., EBERHARD, D., MATTHIESON, T., SWEETMAN, G. & KUSTER, B. 2008. Robust and sensitive iTRAQ quantification on an LTQ Orbitrap mass spectrometer. *Molecular & Cellular Proteomics*, 7, 1702-1713.
- BANTSCHIEFF, M., LEMEER, S., SAVITSKI, M. M. & KUSTER, B. 2012. Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Analytical and bioanalytical chemistry*, 404, 939-965.
- BANTSCHIEFF, M., SCHIRLE, M., SWEETMAN, G., RICK, J. & KUSTER, B. 2007. Quantitative mass spectrometry in proteomics: a critical review. *Analytical and bioanalytical chemistry*, 389, 1017-1031.
- BANYS, M., HARTKOPF, A. D., KRAWCZYK, N., KAISER, T., MEIER-STIEGEN, F., FEHM, T. & NEUBAUER, H. 2012. Dormancy in breast cancer. *Breast Cancer (Dove Med Press)*, 4, 183-91.
- BARRETINA, J., CAPONIGRO, G., STRANSKY, N., VENKATESAN, K., MARGOLIN, A. A., KIM, S., WILSON, C. J., LEHÁR, J., KRYUKOV, G. V. & SONKIN, D. 2012. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, 483, 603-607.
- BASAKRAN, N. S. 2015. CD44 as a potential diagnostic tumor marker. *Saudi medical journal*, 36, 273-279.
- BECK, K. L., WEBER, D., PHINNEY, B. S., SMILOWITZ, J. T., HINDE, K., LONNERDAL, B., KORF, I. & LEMAY, D. G. 2015. Comparative Proteomics of Human and Macaque Milk Reveals Species-Specific Nutrition during Postnatal Development. *J Proteome Res*, 14, 2143-57.
- BERG, W. A., WEINBERG, I. N., NARAYANAN, D., LOBRANO, M. E., ROSS, E., AMODEI, L., TAFRA, L., ADLER, L. P., UDDO, J. & STEIN, W. 2006. High-Resolution Fluorodeoxyglucose Positron Emission Tomography with Compression ("Positron Emission Mammography") is Highly Accurate in Depicting Primary Breast Cancer. *The breast journal*, 12, 309-323.
- BERGNER, A., KELLNER, J., TUFMAN, A. & HUBER, R. M. 2009. Journal of Experimental & Clinical Cancer Research. *Journal of Experimental & Clinical Cancer Research*, 28, 25.
- BIANCHI, S., BENDINELLI, B., SALADINO, V., VEZZOSI, V., BRANCATO, B., NORI, J. & PALLI, D. 2015. Non-Malignant Breast Papillary Lesions-B3 Diagnosed on Ultrasound-Guided 14-Gauge Needle Core Biopsy: Analysis of 114 Cases from a Single Institution and Review of the Literature. *Pathology & Oncology Research*, 1-12.
- BLEICHER, R. J. & MORROW, M. 2007. MRI and breast cancer: role in detection, diagnosis, and staging. *Oncology*, 21, 1521-8.
- BODZON-KULAKOWSKA, A., BIERCZYNSKA-KRZYSIK, A., DYLAG, T., DRABIK, A., SUDER, P., NOGA, M., JARZEBINSKA, J. & SILBERRING, J. 2007. Methods for samples preparation in proteomic research. *Journal of Chromatography B*, 849, 1-31.
- BOECKMANN, B., BAIROCH, A., APWEILER, R., BLATTER, M.-C., ESTREICHER, A., GASTEIGER, E., MARTIN, M. J., MICHOU, K., O'DONOVAN, C. & PHAN, I. 2003. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic acids research*, 31, 365-370.
- BÖHM, D., KELLER, K., PIETER, J., BOEHM, N., WOLTERS, D., SIGGELKOW, W., LEBRECHT, A., SCHMIDT, M., KÖLBL, H. & PFEIFFER, N. 2012. Comparison of tear protein levels in breast cancer patients and healthy controls using a de novo proteomic approach. *Oncology reports*, 28, 429-438.
- BONDARENKO, P. V., CHELIUS, D. & SHALER, T. A. 2002. Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry. *Analytical chemistry*, 74, 4741-4749.

- BOUCHAL, P., ROUMELIOTIS, T., HRSTKA, R., NENUTIL, R., VOJTESEK, B. & GARBIS, S. D. 2008. Biomarker discovery in low-grade breast cancer using isobaric stable isotope tags and two-dimensional liquid chromatography-tandem mass spectrometry (iTRAQ-2DLC-MS/MS) based quantitative proteomic analysis. *Journal of proteome research*, 8, 362-373.
- BOUTET, E., LIEBERHERR, D., TOGNOLLI, M., SCHNEIDER, M., BANSAL, P., BRIDGE, A. J., POUX, S., BOUGUELERET, L. & XENARIOS, I. 2016. UniProtKB/Swiss-Prot, the manually annotated section of the UniProt KnowledgeBase: how to use the entry view. *Plant Bioinformatics: Methods and Protocols*, 23-54.
- BRADFORD, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry*, 72, 248-254.
- BRIERLEY, J. D. 2017. *TNM classification of malignant tumours*, John Wiley & Sons.
- BROOKS, M. 2009. Breast cancer screening and biomarkers. *Methods Mol Biol*, 472, 307-21.
- BROSCH, M., YU, L., HUBBARD, T. & CHOUDHARY, J. 2009. Accurate and sensitive peptide identification with Mascot Percolator. *Journal of proteome research*, 8, 3176-3181.
- BROWNRIDGE, P. J., HARMAN, V. M., SIMPSON, D. M. & BEYNON, R. J. 2012. Absolute multiplexed protein quantification using QconCAT technology. *Quantitative Methods in Proteomics*, 267-293.
- BRUINS, A. P., COVEY, T. R. & HENION, J. D. 1987. Ion spray interface for combined liquid chromatography/atmospheric pressure ionization mass spectrometry. *Analytical Chemistry*, 59, 2642-2646.
- BRUNORO, G. V., CARVALHO, P. C., FERREIRA, A. T., PERALES, J., VALENTE, R. H., DE MOURA GALLO, C. V., PAGNONCELLI, D. & NEVES-FERREIRA, A. G. 2015. Proteomic profiling of nipple aspirate fluid (NAF): Exploring the complementarity of different peptide fractionation strategies. *J Proteomics*, 117, 86-94.
- CADDOO, K., FORNIER, M. & MORRIS, P. 2013. Biological subtypes of breast cancer: current concepts and implications for recurrence patterns. *The quarterly journal of nuclear medicine and molecular imaging: official publication of the Italian Association of Nuclear Medicine (AIMN)[and] the International Association of Radiopharmacology (IAR),[and] Section of the Society of..* 57, 312-321.
- CALDERÓN-GONZÁLEZ, K. G., RUSTARAZO, M. L. V., LABRA-BARRIOS, M. L., BAZÁN-MÉNDEZ, C. I., TAVERA-TAPIA, A., HERRERA-AGUIRRE, M. E., DEL PINO, M. M. S., GALLEGOS-PÉREZ, J. L., GONZÁLEZ-MÁRQUEZ, H. & HERNÁNDEZ-HERNÁNDEZ, J. M. 2015. Determination of the protein expression profiles of breast cancer cell lines by quantitative proteomics using iTRAQ labelling and tandem mass spectrometry. *Journal of proteomics*, 124, 50-78.
- CALVO, S. E., CLAUSER, K. R. & MOOTHA, V. K. 2016. MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. *Nucleic Acids Res*, 44, D1251-7.
- CANCER, A. J. C. O. 2002. *Breast. AJCC cancer staging manual*. Springer.
- CANCERRESEARCHUK. 2016a. *Breast Cancer* [Online]. Cancer Research UK. Available: <http://www.cancerresearchuk.org/> [Accessed].
- CANCERRESEARCHUK. 2016b. *Breast cancer statistics* [Online]. Cancer Research UK. Available: <http://www.cancerresearchuk.org/health-professional/cancer-statistics> [Accessed 05/12 2016].
- CANTO, D., MATOS, L., MARIAN, C., WILLEY, S., SIDAWY, M., DA CUNHA, P. A., RONE, J. D., LI, X., GUSEV, Y. & HADDAD, B. R. 2016. MicroRNA analysis of breast ductal fluid in breast cancer patients. *International journal of oncology*, 48, 2071-2078.
- CAPPELLO, F., CZARNECKA, A. M., LA ROCCA, G., DI STEFANO, A., ZUMMO, G. & MACARIO, A. J. 2007. Hsp60 and Hsp10 as antitumour molecular agents. *Cancer biology & therapy*, 6, 487-489.

- CARGILE, B. J., SEVINSKY, J. R., ESSADER, A. S., STEPHENSON JR, J. L. & BUNDY, J. L. 2005. Immobilized pH gradient isoelectric focusing as a first-dimension separation in shotgun proteomics. *Journal of biomolecular techniques: JBT*, 16, 181.
- CASTEDO, M., PERFETTINI, J., ROUMIER, T. & KROEMER, G. 2002. Cyclin-dependent kinase-1: linking apoptosis to cell cycle and mitotic catastrophe. *Cell death and differentiation*, 9, 1287-1293.
- CELIS, J. E., GROMOVA, I., GROMOV, P., MOREIRA, J., CABEZÓN, T., FRIIS, E. & RANK, F. 2006. Molecular pathology of breast apocrine carcinomas: a protein expression signature specific for benign apocrine metaplasia. *FEBS letters*, 580, 2935-2944.
- CHAN, A. A., BASHIR, M., RIVAS, M. N., DUVALL, K., SIELING, P. A., PIEBER, T. R., VAISHAMPAYAN, P. A., LOVE, S. M. & LEE, D. J. 2016a. Characterization of the microbiome of nipple aspirate fluid of breast cancer survivors. *Scientific reports*, 6, 28061.
- CHAN, A. A., BASHIR, M., RIVAS, M. N., DUVALL, K., SIELING, P. A., PIEBER, T. R., VAISHAMPAYAN, P. A., LOVE, S. M. & LEE, D. J. 2016b. Characterization of the microbiome of nipple aspirate fluid of breast cancer survivors. *Scientific Reports*, 6.
- CHANDRAMOULI, K. & QIAN, P.-Y. 2009. Proteomics: challenges, techniques and possibilities to overcome biological sample complexity. *Human genomics and proteomics: HGP*, 2009.
- CHANG, Y.-F., HUNG, S.-H., SU, L.-C., CHEN, R.-C. & CHOU, C. 2014. Association of HER2 ECD and t-PSA serum levels for possible breast cancer diagnosis. *Sensing and Bio-Sensing Research*, 2, 1-7.
- CHARAFE-JAUFFRET, E., GINESTIER, C., MONVILLE, F., FINETTI, P., ADELAIDE, J., CERVERA, N., FEKAIRI, S., XERRI, L., JACQUEMIER, J. & BIRNBAUM, D. 2005. Gene expression profiling of breast cell lines identifies potential new basal markers. *Oncogene*, 25, 2273-2284.
- CHATTERTON, R. T., JR., KHAN, S. A., HEINZ, R., IVANCIC, D. & LEE, O. 2010. Patterns of sex steroid hormones in nipple aspirate fluid during the menstrual cycle and after menopause in relation to serum concentrations. *Cancer Epidemiol Biomarkers Prev*, 19, 275-9.
- CHEANG, M. C., CHIA, S. K., VODUC, D., GAO, D., LEUNG, S., SNIDER, J., WATSON, M., DAVIES, S., BERNARD, P. S. & PARKER, J. S. 2009. Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. *Journal of the National Cancer Institute*.
- CHEN, C.-N., CHANG, C.-C., SU, T.-E., HSU, W.-M., JENG, Y.-M., HO, M.-C., HSIEH, F.-J., LEE, P.-H., KUO, M.-L. & LEE, H. 2009. Identification of calreticulin as a prognosis marker and angiogenic regulator in human gastric cancer. *Annals of surgical oncology*, 16, 524-533.
- CHEN, H., XU, C., JIN, Q. E. & LIU, Z. 2014. S100 protein family in human cancer. *Am J Cancer Res*, 4, 89-115.
- CHEN, J. & TURKO, I. V. 2014. Trends in QconCATs for targeted proteomics. *TrAC Trends in Analytical Chemistry*, 57, 1-5.
- CHEN, Y.-T., CHEN, H.-W., DOMANSKI, D., SMITH, D. S., LIANG, K.-H., WU, C.-C., CHEN, C.-L., CHUNG, T., CHEN, M.-C. & CHANG, Y.-S. 2012. Multiplexed quantification of 63 proteins in human urine by multiple reaction monitoring-based mass spectrometry for discovery of potential bladder cancer biomarkers. *Journal of proteomics*, 75, 3529-3545.
- CHEN, Z., YUAN, Y.-C., WANG, Y., LIU, Z., CHAN, H. J. & CHEN, S. 2015. Down-regulation of programmed cell death 4 (PDCD4) is associated with aromatase inhibitor resistance and a poor prognosis in estrogen receptor-positive breast cancer. *Breast cancer research and treatment*, 152, 29-39.
- CHIRAS, D. D. 2013. *Human biology*, Jones & Bartlett Publishers.
- CHOE, L., D'ASCENZO, M., RELKIN, N. R., PAPPIN, D., ROSS, P., WILLIAMSON, B., GUERTIN, S., PRIBIL, P. & LEE, K. H. 2007. 8-Plex quantitation of changes in cerebrospinal fluid

- protein expression in subjects undergoing intravenous immunoglobulin treatment for Alzheimer's disease. *Proteomics*, 7, 3651-3660.
- CHOI, S., CHO, K., KIM, J., YEA, K., PARK, G., LEE, J., RYU, S. H., KIM, J. & KIM, Y. H. 2009. Comparative proteome analysis using amine-reactive isobaric tagging reagents coupled with 2D LC/MS/MS in 3T3-L1 adipocytes following hypoxia or normoxia. *Biochemical and biophysical research communications*, 383, 135-140.
- CIFANI, P., KIRIK, U., WALDEMARSON, S. & JAMES, P. 2015. Molecular portrait of breast-cancer-derived cell lines reveals poor similarity with tumors. *Journal of proteome research*, 14, 2819-2827.
- CITY, O. 2012. AGR2 as a potential biomarker of human lung adenocarcinoma. *Osaka City Med. J*, 58, 13-24.
- COMISAROW, M. B. & MARSHALL, A. G. 1974. Fourier transform ion cyclotron resonance spectroscopy. *Chemical physics letters*, 25, 282-283.
- COPPOLINO, M. G., WOODSIDE, M. J., DEMAUREX, N., GRINSTEIN, S., ST-ARNAUD, R. & DEDHAR, S. 1997. Calreticulin is essential for integrin-mediated calcium signalling and cell adhesion.
- COPSON, E., ECCLES, B., MAISHMAN, T., GERTY, S., STANTON, L., CUTRESS, R. I., ALTMAN, D. G., DURCAN, L., SIMMONDS, P., LAWRENCE, G., JONES, L., BLISS, J., ECCLES, D. & GROUP, P. S. S. 2013. Prospective observational study of breast cancer treatment outcomes for UK women aged 18-40 years at diagnosis: the POSH study. *J Natl Cancer Inst*, 105, 978-88.
- CORRAO, S., CAMPANELLA, C., ANZALONE, R., FARINA, F., ZUMMO, G., DE MACARIO, E. C., MACARIO, A. J., CAPPELLO, F. & LA ROCCA, G. 2010. Human Hsp10 and Early Pregnancy Factor (EPF) and their relationship and involvement in cancer and immunity: current knowledge and perspectives. *Life sciences*, 86, 145-152.
- COUMANS, J. V., GAU, D., POLJAK, A., WASINGER, V., ROY, P. & MOENS, P. D. 2014. Profilin-1 overexpression in MDA-MB-231 breast cancer cells is associated with alterations in proteomics biomarkers of cell proliferation, survival, and motility as revealed by global proteomics analyses. *Omics: a journal of integrative biology*, 18, 778-791.
- COX, J. & MANN, M. 2012. 1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data. *BMC bioinformatics*, 13, S12.
- CRICK, F. 1970. Central dogma of molecular biology. *Nature*, 227, 561-563.
- CUTILLAS, P. R. & TIMMS, J. F. 2010. LC-MS/MS in Proteomics. *Methods in Molecular Biology*, 658.
- DANFORTH, D. N., WARNER, A. C., WANGSA, D., RIED, T., DUELLI, D., FILIE, A. C. & PRINDIVILLE, S. A. 2015. An Improved Breast Epithelial Sampling Method for Molecular Profiling and Biomarker Analysis in Women at Risk for Breast Cancer. *Breast Cancer (Auckl)*, 9, 31-40.
- DAVIES, K. R., CANTOR, S. B. & BREWSTER, A. M. 2015. Better contralateral breast cancer risk estimation and alternative options to contralateral prophylactic mastectomy. *Int J Womens Health*, 7, 181-7.
- DE GODOY, L. M., OLSEN, J. V., COX, J., NIELSEN, M. L., HUBNER, N. C., FRÖHLICH, F., WALTHER, T. C. & MANN, M. 2008. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature*, 455, 1251-1254.
- DE GROOT, J., MOELANS, C., ELIAS, S., HENNINK, A., VEROLME, B., SUIJKERBUIJK, K., JAGER, A., SEYNAEVE, C., BOS, P. & WITKAMP, A. 2015. Repeated nipple fluid aspiration: compliance and feasibility results from a prospective multicenter study. *PLoS one*, 10, e0127895.
- DE MARCHI, T., KUHN, E., DEKKER, L. J., STINGL, C., BRAAKMAN, R. B., OPDAM, M., LINN, S. C., SWEEP, F. C., SPAN, P. N. & LUIDER, T. M. 2016. Targeted MS assay predicting tamoxifen resistance in estrogen-receptor-positive breast cancer tissues and sera. *Journal of proteome research*, 15, 1230-1242.

- DESANTIS, C., MA, J., BRYAN, L. & JEMAL, A. 2014. Breast cancer statistics, 2013. *CA: a cancer journal for clinicians*, 64, 52-62.
- DESANTIS, C., SIEGEL, R., BANDI, P. & JEMAL, A. 2011. Breast cancer statistics, 2011. *CA Cancer J Clin*, 61, 409-18.
- DESIDERIO, D. M. 2013. *Mass spectrometry: clinical and biomedical applications*, Springer Science & Business Media.
- DESTA, Z., WARD, B. A., SOUKHOVA, N. V. & FLOCKHART, D. A. 2004. Comprehensive evaluation of tamoxifen sequential biotransformation by the human cytochrome P450 system in vitro: prominent roles for CYP3A and CYP2D6. *J Pharmacol Exp Ther*, 310, 1062-75.
- DEUTSCHER, S. L., DICKERSON, M., GUI, G., NEWTON, J., HOLM, J. E., VOGELTANZ-HOLM, N., KLIETHERMES, B., HEWETT, J. E., KUMAR, S. R. & QUINN, T. P. 2010a. Carbohydrate antigens in nipple aspirate fluid predict the presence of atypia and cancer in women requiring diagnostic breast biopsy. *BMC cancer*, 10, 519.
- DEUTSCHER, S. L., DICKERSON, M., GUI, G., NEWTON, J., HOLM, J. E., VOGELTANZ-HOLM, N., KLIETHERMES, B., HEWETT, J. E., KUMAR, S. R., QUINN, T. P. & SAUTER, E. R. 2010b. Carbohydrate antigens in nipple aspirate fluid predict the presence of atypia and cancer in women requiring diagnostic breast biopsy. *BMC Cancer*, 10, 519.
- DI PALMA, S., HENNRICH, M. L., HECK, A. J. & MOHAMMED, S. 2012. Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis. *Journal of proteomics*, 75, 3791-3813.
- DIAMENT, B. J. & NOBLE, W. S. 2011. Faster SEQUEST searching for peptide identification from tandem mass spectra. *Journal of proteome research*, 10, 3871-3879.
- DIAS, K., DVORKIN-GHEVA, A., HALLETT, R. M., WU, Y., HASSELL, J., POND, G. R., LEVINE, M., WHELAN, T. & BANE, A. L. 2017. Claudin-Low Breast Cancer; Clinical & Pathological Characteristics. *PloS one*, 12, e0168669.
- DIETZ, J. R., CROWE, J. P., GRUNDFEST, S., ARRIGAIN, S. & KIM, J. A. 2002. Directed duct excision by using mammary ductoscopy in patients with pathologic nipple discharge. *Surgery*, 132, 582-588.
- DJURIC, Z., CHEN, G., REN, J., VENKATRAMANAMOORTHY, R., COVINGTON, C. Y., KUCUK, O. & HEILBRUN, L. K. 2007. Effects of high fruit-vegetable and/or low-fat intervention on breast nipple aspirate fluid micronutrient levels. *Cancer Epidemiol Biomarkers Prev*, 16, 1393-9.
- DJURIC, Z., REN, J., MEKHOVICH, O., VENKATRANAMOORTHY, R. & HEILBRUN, L. K. 2006. Effects of high fruit-vegetable and/or low-fat intervention on plasma micronutrient levels. *Journal of the American College of Nutrition*, 25, 178-187.
- DJURIC, Z., VISSCHER, D. W., HEILBRUN, L. K., CHEN, G., ATKINS, M. & COVINGTON, C. Y. 2005. Influence of lactation history on breast nipple aspirate fluid yields and fluid composition. *The breast journal*, 11, 92-99.
- DOERR, A. 2013. Mass spectrometry-based targeted proteomics. *Nature methods*, 10, 23.
- DOERR, A. 2015. DIA mass spectrometry. *Nature Methods*, 12, 35-35.
- DOMANSKI, D., FREUE, G. V. C., SOJO, L., KUZYSK, M. A., RATKAY, L., PARKER, C. E., GOLDBERG, Y. P. & BORCHERS, C. H. 2012a. The use of multiplexed MRM for the discovery of biomarkers to differentiate iron-deficiency anemia from anemia of inflammation. *Journal of proteomics*, 75, 3514-3528.
- DOMANSKI, D., PERCY, A. J., YANG, J., CHAMBERS, A. G., HILL, J. S., FREUE, G. V. C. & BORCHERS, C. H. 2012b. MRM-based multiplexed quantitation of 67 putative cardiovascular disease biomarkers in human plasma. *Proteomics*, 12, 1222-1243.
- DOMON, B. & AEBERSOLD, R. 2010. Options and considerations when selecting a quantitative proteomics strategy. *Nature biotechnology*, 28, 710-721.
- DOOLEY, W. C., LJUNG, B.-M., VERONESI, U., CAZZANIGA, M., ELLEDGE, R. M., O'SHAUGHNESSY, J. A., KUERER, H. M., HUNG, D. T., KHAN, S. A. & PHILLIPS, R. F. 2001.

- Ductal lavage for detection of cellular atypia in women at high risk for breast cancer. *Journal of the national cancer institute*, 93, 1624-1632.
- DORFER, V., PICHLER, P., STRANZL, T., STADLMANN, J., TAUS, T., WINKLER, S. & MECHTLER, K. 2014. MS Amanda, a universal identification algorithm optimized for high accuracy tandem mass spectra. *Journal of proteome research*, 13, 3679-3684.
- DRAKE, R. R., CAZARES, L. H., JONES, E. E., FULLER, T. W., SEMMES, O. J. & LARONGA, C. 2011. Challenges to developing proteomic-based breast cancer diagnostics.
- DRUKTEINIS, J. S., MOONEY, B. P., FLOWERS, C. I. & GATENBY, R. A. 2013. Beyond mammography: new frontiers in breast cancer screening. *Am J Med*, 126, 472-9.
- DUNCAN, M. W., AEBERSOLD, R. & CAPRIOLI, R. M. 2010. The pros and cons of peptide-centric proteomics [AU: OK?]. *Nature biotechnology*, 28, 1.
- ECCLES, S. A., ABOAGYE, E. O., ALI, S., ANDERSON, A. S., ARMES, J., BERDITCHEVSKI, F., BLAYDES, J. P., BRENNAN, K., BROWN, N. J., BRYANT, H. E., BUNDRED, N. J., BURCHELL, J. M., CAMPBELL, A. M., CARROLL, J. S., CLARKE, R. B., COLES, C. E., COOK, G. J., COX, A., CURTIN, N. J., DEKKER, L. V., SILVA IDOS, S., DUFFY, S. W., EASTON, D. F., ECCLES, D. M., EDWARDS, D. R., EDWARDS, J., EVANS, D., FENLON, D. F., FLANAGAN, J. M., FOSTER, C., GALLAGHER, W. M., GARCIA-CLOSAS, M., GEE, J. M., GESCHER, A. J., GOH, V., GROVES, A. M., HARVEY, A. J., HARVIE, M., HENNESSY, B. T., HISCOX, S., HOLEN, I., HOWELL, S. J., HOWELL, A., HUBBARD, G., HULBERT-WILLIAMS, N., HUNTER, M. S., JASANI, B., JONES, L. J., KEY, T. J., KIRWAN, C. C., KONG, A., KUNKLER, I. H., LANGDON, S. P., LEACH, M. O., MANN, D. J., MARSHALL, J. F., MARTIN, L., MARTIN, S. G., MACDOUGALL, J. E., MILES, D. W., MILLER, W. R., MORRIS, J. R., MOSS, S. M., MULLAN, P., NATRAJAN, R., O'CONNOR, J. P., O'CONNOR, R., PALMIERI, C., PHAROAH, P. D., RAKHA, E. A., REED, E., ROBINSON, S. P., SAHAI, E., SAXTON, J. M., SCHMID, P., SMALLEY, M. J., SPEIRS, V., STEIN, R., STINGL, J., STREULI, C. H., TUTT, A. N., VELIKOVA, G., WALKER, R. A., WATSON, C. J., WILLIAMS, K. J., YOUNG, L. S. & THOMPSON, A. M. 2013. Critical research gaps and translational priorities for the successful prevention and treatment of breast cancer. *Breast Cancer Res*, 15, R92.
- EDDY, J. A., SUNG, J., GEMAN, D. & PRICE, N. D. 2010. Relative expression analysis for molecular cancer diagnosis and prognosis. *Technology in cancer research & treatment*, 9, 149-159.
- ELIAS, J. E. & GYGI, S. P. 2007. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature methods*, 4, 207-214.
- ELIASSEN, A. H., MISSMER, S. A., TWOROGER, S. S., SPIEGELMAN, D., BARBIERI, R. L., DOWSETT, M. & HANKINSON, S. E. 2006. Endogenous steroid hormone concentrations and risk of breast cancer among premenopausal women. *Journal of the National Cancer Institute*, 98, 1406-1415.
- ELLIOTT, M. H., SMITH, D. S., PARKER, C. E. & BORCHERS, C. 2009. Current trends in quantitative proteomics. *Journal of Mass Spectrometry*, 44, 1637-1660.
- ENG, J. K., JAHAN, T. A. & HOOPMANN, M. R. 2013. Comet: An open-source MS/MS sequence database search tool. *Proteomics*, 13, 22-24.
- ENGELMAN, J. A. & CANTLEY, L. C. 2008. A sweet new role for EGFR in cancer. *Cancer cell*, 13, 375-376.
- ESPINA, V. & LIOTTA, L. A. 2011. What is the malignant nature of human ductal carcinoma in situ? *Nat Rev Cancer*, 11, 68-75.
- ESSADER, A. S., CARGILE, B. J., BUNDY, J. L. & STEPHENSON, J. L. 2005. A comparison of immobilized pH gradient isoelectric focusing and strong-cation-exchange chromatography as a first dimension in shotgun proteomics. *Proteomics*, 5, 24-34.
- FABIAN, C. J., KIMLER, B. F., BRADY, D. A., MAYO, M. S., CHANG, C. J., FERRARO, J. A., ZALLES, C. M., STANTON, A. L., MASOOD, S. & GRIZZLE, W. E. 2002. A Phase II Breast Cancer Chemoprevention Trial of Oral α -Difluoromethylornithine Breast Tissue, Imaging, and Serum and Urine Biomarkers. *Clinical cancer research*, 8, 3105-3117.

- FABIAN, C. J., KIMLER, B. F., ZALLES, C. M., KLEMP, J. R., KAMEL, S., ZEIGER, S. & MAYO, M. S. 2000. Short-term breast cancer prediction by random periareolar fine-needle aspiration cytology and the Gail risk model. *Journal of the National Cancer Institute*, 92, 1217-1227.
- FACKLER, M. J., MALONE, K., ZHANG, Z., SCHILLING, E., GARRETT-MAYER, E., SWIFT-SCANLAN, T., LANGE, J., NAYAR, R., DAVIDSON, N. E. & KHAN, S. A. 2006. Quantitative multiplex methylation-specific PCR analysis doubles detection of tumor cells in breast ductal fluid. *Clinical Cancer Research*, 12, 3306-3310.
- FEIST, P. & HUMMON, A. B. 2015. Proteomic challenges: sample preparation techniques for microgram-quantity protein analysis from biological samples. *International journal of molecular sciences*, 16, 3537-3563.
- FENN, J. B., MANN, M., MENG, C. K., WONG, S. F. & WHITEHOUSE, C. M. 1990. Electrospray ionization—principles and practice. *Mass Spectrometry Reviews*, 9, 37-70.
- FERLAY J, SOERJOMATARAM I, ERVIK M, D., R, E. S., MATHERS C, REBELO M, PARKIN DM, FORMAN D & BRAY, F. 2013. *Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11* [Online]. Lyon, France: GLOBOCAN. Available: <http://globocan.iarc.fr> [Accessed 04/08/2014].
- FINK, J. L., KARUNARATNE, S., MITTAL, A., GARDINER, D. M., HAMILTON, N., MAHONY, D., KAI, C., SUZUKI, H., HAYASHIZAKI, Y. & TEASDALE, R. D. 2008. Towards defining the nuclear proteome. *Genome Biol*, 9, R15.
- FORTIN, T., SALVADOR, A., CHARRIER, J. P., LENZ, C., LACOUX, X., MORLA, A., CHOQUET-KASTYLEVSKY, G. & LEMOINE, J. 2009. Clinical quantitation of prostate-specific antigen biomarker in the low nanogram/milliliter range by conventional bore liquid chromatography-tandem mass spectrometry (multiple reaction monitoring) coupling and correlation with ELISA tests. *Molecular & Cellular Proteomics*, 8, 1006-1015.
- FRANCIS, A., THOMAS, J., FALLOWFIELD, L., WALLIS, M., BARTLETT, J. M., BROOKES, C., ROBERTS, T., PIRRIE, S., GAUNT, C., YOUNG, J., BILLINGHAM, L., DODWELL, D., HANBY, A., PINDER, S. E., EVANS, A., REED, M., JENKINS, V., MATTHEWS, L., WILCOX, M., FAIRBROTHER, P., BOWDEN, S. & REA, D. 2015. Addressing overtreatment of screen detected DCIS; the LORIS trial. *Eur J Cancer*, 51, 2296-303.
- FRANKEL, L. B., CHRISTOFFERSEN, N. R., JACOBSEN, A., LINDOW, M., KROGH, A. & LUND, A. H. 2008. Programmed cell death 4 (PDCD4) is an important functional target of the microRNA miR-21 in breast cancer cells. *Journal of Biological Chemistry*, 283, 1026-1033.
- FRITZSCHE, F. R., DAHL, E., PAHL, S., BURKHARDT, M., LUO, J., MAYORDOMO, E., GANSUKH, T., DANKOF, A., KNUECHEL, R. & DENKERT, C. 2006. Prognostic relevance of AGR2 expression in breast cancer. *Clinical Cancer Research*, 12, 1728-1734.
- FULFORD, L., EASTON, D., REIS-FILHO, J., SOFRONIS, A., GILLETT, C., LAKHANI, S. & HANBY, A. 2006. Specific morphological features predictive for the basal phenotype in grade 3 invasive ductal carcinoma of breast. *Histopathology*, 49, 22-34.
- FUSARO, V. A., MANI, D., MESIROV, J. P. & CARR, S. A. 2009. Prediction of high-responding peptides for targeted protein assays by mass spectrometry. *Nature biotechnology*, 27, 190-198.
- GALVÃO, E., MARTINS, L., IBIAPINA, J., ANDRADE, H. & MONTE, S. 2011. Breast cancer proteomics: a review for clinicians. *Journal of cancer research and clinical oncology*, 137, 915-925.
- GAN, C. S., CHONG, P. K., PHAM, T. K. & WRIGHT, P. C. 2007. Technical, experimental, and biological variations in isobaric tags for relative and absolute quantitation (iTRAQ). *Journal of proteome research*, 6, 821-827.
- GARBIS, S. D., TYRITZIS, S. I., ROUMELIOTIS, T., ZEREFOS, P., GIANNOPOULOU, E. G., VLAHOU, A., KOSSIDA, S., DIAZ, J., VOUREKAS, S. & TAMVAKOPOULOS, C. 2008. Search for potential markers for prostate cancer diagnosis, prognosis and treatment in clinical tissue specimens using amine-specific isobaric tagging (iTRAQ) with two-dimensional

- liquid chromatography and tandem mass spectrometry. *The Journal of Proteome Research*, 7, 3146-3158.
- GARDAI, S. J., MCPHILLIPS, K. A., FRASCH, S. C., JANSSEN, W. J., STAREFELDT, A., MURPHY-ULLRICH, J. E., BRATTON, D. L., OLDENBORG, P.-A., MICHALAK, M. & HENSON, P. M. 2005. Cell-surface calreticulin initiates clearance of viable or apoptotic cells through trans-activation of LRP on the phagocyte. *Cell*, 123, 321-334.
- GAST, M.-C. W., SCHELLENS, J. H. & BEIJNEN, J. H. 2009. Clinical proteomics in breast cancer: a review. *Breast cancer research and treatment*, 116, 17-29.
- GATZA, M. L., KUNG, H.-N., BLACKWELL, K. L., DEWHIRST, M. W., MARKS, J. R. & CHI, J.-T. 2011. Analysis of tumor environmental response and oncogenic pathway activation identifies distinct basal and luminal features in HER2-related breast tumor subtypes. *Breast Cancer Res*, 13, R62.
- GEER, L. Y., MARKEY, S. P., KOWALAK, J. A., WAGNER, L., XU, M., MAYNARD, D. M., YANG, X., SHI, W. & BRYANT, S. H. 2004. Open mass spectrometry search algorithm. *Journal of proteome research*, 3, 958-964.
- GEIGER, T., COX, J. & MANN, M. 2010. Proteomic changes resulting from gene copy number variations in cancer cells. *PLoS genetics*, 6, e1001090.
- GEIGER, T., MADDEN, S. F., GALLAGHER, W. M., COX, J. & MANN, M. 2012. Proteomic portrait of human breast cancer progression identifies novel prognostic markers. *Cancer research*, 72, 2428-2439.
- GILBEY, A., BURNETT, D., COLEMAN, R. & HOLEN, I. 2004. The detection of circulating breast cancer cells in blood. *Journal of clinical pathology*, 57, 903-911.
- GIULIANO, V. & GIULIANO, C. 2013. Improved breast cancer detection in asymptomatic women using 3D-automated breast ultrasound in mammographically dense breasts. *Clin Imaging*, 37, 480-6.
- GLASER, J., NEUMANN, M., MEI, Q., BETZ, B., SEIER, N., BEYER, I., FEHM, T., NEUBAUER, H., NIEDERACHER, D. & FLEISCH, M. 2014. Macrophage capping protein CapG is a putative oncogene involved in migration and invasiveness in ovarian carcinoma. *BioMed research international*, 2014.
- GLEN, A., GAN, C. S., HAMDY, F. C., EATON, C. L., CROSS, S. S., CATTO, J. W., WRIGHT, P. C. & REHMAN, I. 2008. iTRAQ-facilitated proteomic analysis of human prostate cancer cells identifies proteins associated with progression. *Journal of proteome research*, 7, 897-907.
- GLENTIS, A., GURCHENKOV, V. & MATIC VIGNJEVIC, D. 2014. Assembly, heterogeneity, and breaching of the basement membranes. *Cell Adh Migr*, 8, 236-45.
- GOLDHIRSCH, A., WOOD, W., COATES, A., GELBER, R., THÜRLIMANN, B. & SENN, H.-J. 2011. Strategies for subtypes—dealing with the diversity of breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Annals of oncology*, mdr304.
- GOTZSCHE, P. C. & JORGENSEN, K. J. 2013. Screening for breast cancer with mammography. *Cochrane Database Syst Rev*, 6, CD001877.
- GPM. 2016. *The Global Proteome Machine* [Online]. Available: <http://gpmdb.thegpm.org/index.html> [Accessed 06/06 2016].
- GRIFFIN, T. J., XIE, H., BANDHAKAVI, S., POPKO, J., MOHAN, A., CARLIS, J. V. & HIGGINS, L. 2007. iTRAQ reagent-based quantitative proteomic analysis on a linear ion trap mass spectrometer. *Journal of proteome research*, 6, 4200-4209.
- GRY, M., RIMINI, R., STRÖMBERG, S., ASPLUND, A., PONTÉN, F., UHLÉN, M. & NILSSON, P. 2009. Correlations between RNA and protein expression profiles in 23 human cell lines. *BMC genomics*, 10, 365.
- GSTAIGER, M. & AEBERSOLD, R. 2009. Applying mass spectrometry-based proteomics to genetics, genomics and network biology. *Nature Reviews Genetics*, 10, 617-627.

- GUJRAL, T. S., KARP, R. L., FINSKI, A., CHAN, M., SCHWARTZ, P. E., MACBEATH, G. & SORGER, P. 2012. Profiling phospho-signaling networks in breast cancer using reverse-phase protein arrays. *Oncogene*, 32, 3470-3476.
- GUNDRY, R. L., WHITE, M. Y., MURRAY, C. I., KANE, L. A., FU, Q., STANLEY, B. A. & VAN EYK, J. E. 2009. Preparation of Proteins and Peptides for Mass Spectrometry Analysis in a Bottom-Up Proteomics Workflow. *Current protocols in molecular biology*, 10.25. 1-10.25. 23.
- GUO, C., LIU, S., WANG, J., SUN, M.-Z. & GREENAWAY, F. T. 2013. ACTB in cancer. *Clinica Chimica Acta*, 417, 39-44.
- GÜTH, U., HUANG, D. J., HUBER, M., SCHÖTZAU, A., WRUK, D., HOLZGREVE, W., WIGHT, E. & ZANETTI-DÄLLENBACH, R. 2008. Tumor size and detection in breast cancer: Self-examination and clinical breast examination are at their limit. *Cancer detection and prevention*, 32, 224-228.
- HAAS, B. M., KALRA, V., GEISEL, J., RAGHU, M., DURAND, M. & PHILPOTTS, L. E. 2013. Comparison of tomosynthesis plus digital mammography and digital mammography alone for breast cancer screening. *Radiology*, 269, 694-700.
- HACIHASANOG'LU, R. & GÖZÜM, S. 2008. The effect of training on the knowledge levels and beliefs regarding breast self-examination on women attending a public education centre. *European Journal of Oncology Nursing*, 12, 58-64.
- HAN, X., ASLANIAN, A. & YATES, J. R. 2008. Mass spectrometry for proteomics. *Current opinion in chemical biology*, 12, 483-490.
- HANAHAH, D. & WEINBERG, R. A. 2000. The hallmarks of cancer. *cell*, 100, 57-70.
- HANASH, S. M., PITTERI, S. J. & FACA, V. M. 2008. Mining the plasma proteome for cancer biomarkers. *Nature*, 452, 571-579.
- HARDT, M., WITKOWSKA, H. E., WEBB, S., THOMAS, L. R., DIXON, S. E., HALL, S. C. & FISHER, S. J. 2005. Assessing the effects of diurnal variation on the composition of human parotid saliva: quantitative analysis of native peptides using iTRAQ reagents. *Analytical chemistry*, 77, 4947-4954.
- HARLAN, R. & ZHANG, H. 2014. Targeted proteomics: a bridge between discovery and validation. *Expert review of proteomics*, 11, 657-661.
- HARRIS, L., FRITSCH, H., MENNEL, R., NORTON, L., RAVDIN, P., TAUBE, S., SOMERFIELD, M. R., HAYES, D. F. & BAST, R. C., JR. 2007. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *J Clin Oncol*, 25, 5287-312.
- HARTMAN, A. R., DANIEL, B. L., KURIAN, A. W., MILLS, M. A., NOWELS, K. W., DIRBAS, F. M., KINGHAM, K. E., CHUN, N. M., HERFKENS, R. J. & FORD, J. M. 2004. Breast magnetic resonance image screening and ductal lavage in women at high genetic risk for breast carcinoma. *Cancer*, 100, 479-489.
- HAWKE, M. 2002. Update on cerumen and ceruminolytics. *Ear, nose, & throat journal*, 81, 23-24.
- HEGDE, P. S., WHITE, I. R. & DEBOUCK, C. 2003. Interplay of transcriptomics and proteomics. *Current opinion in biotechnology*, 14, 647-651.
- HERSCHKOWITZ, J. I., SIMIN, K., WEIGMAN, V. J., MIKAELIAN, I., USARY, J., HU, Z., RASMUSSEN, K. E., JONES, L. P., ASSEFNIA, S. & CHANDRASEKHARAN, S. 2007. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome biology*, 8, R76.
- HO, C., LAM, C., CHAN, M., CHEUNG, R., LAW, L., LIT, L., NG, K., SUEN, M. & TAI, H. 2003. Electrospray ionisation mass spectrometry: principles and clinical applications. *Clinical Biochemist Reviews*, 24, 3-12.
- HO, S. K., THIKE, A. A., CHEOK, P. Y., TSE, G. M. & TAN, P. H. 2013. Phyllodes tumours of the breast: the role of CD34, vascular endothelial growth factor and beta-catenin in histological grading and clinical outcome. *Histopathology*, 63, 393-406.

- HOLLE, A., KOSTER, C. & FRANZEN, J. 1997. Mass resolution in time-of-flight mass spectrometers with reflectors. Google Patents.
- HOLLIDAY, D. L. & SPEIRS, V. 2011. Choosing the right cell line for breast cancer research. *Breast Cancer Res*, 13, 215.
- HONDERMARCK, H., TASTET, C., EL YAZIDI-BELKOURA, I., TOILLON, R.-A. & LE BOURHIS, X. 2008. Proteomics of breast cancer: the quest for markers and therapeutic targets. *Journal of proteome research*, 7, 1403-1411.
- HOPE, T. A. & ILES, S. E. 2004. Technology review: the use of electrical impedance scanning in the detection of breast cancer. *Breast Cancer Res*, 6, 69-74.
- HÖRTH, P., MILLER, C. A., PRECKEL, T. & WENZ, C. 2006. Efficient fractionation and improved protein identification by peptide OFFGEL electrophoresis. *Molecular & cellular proteomics*, 5, 1968-1974.
- HOUSSAMI, N. & TURNER, R. M. 2014. Staging the axilla in women with breast cancer: the utility of preoperative ultrasound-guided needle biopsy. *Cancer biology & medicine*, 11, 69.
- HSU, W.-M., HSIEH, F.-J., JENG, Y., KUO, M.-L., CHEN, C.-N., LAI, D.-M., HSIEH, L.-J., WANG, B.-T., TSAO, P.-N. & LEE, H. 2005. Calreticulin expression in neuroblastoma—a novel independent prognostic factor. *Annals of oncology*, 16, 314-321.
- HU, J., QIAN, J., BORISOV, O., PAN, S., LI, Y., LIU, T., DENG, L., WANNEMACHER, K., KURNELLAS, M. & PATTERSON, C. 2006a. Optimized proteomic analysis of a mouse model of cerebellar dysfunction using amine-specific isobaric tags. *Proteomics*, 6, 4321-4334.
- HU, M. & POLYAK, K. 2008. Molecular characterisation of the tumour microenvironment in breast cancer. *European journal of cancer*, 44, 2760-2765.
- HU, S., LOO, J. A. & WONG, D. T. 2006b. Human body fluid proteome analysis. *Proteomics*, 6, 6326-6353.
- HUANG, Y., ANDERSON, K. E., NAGAMANI, M., GRADY, J. J. & LU, L.-J. W. 2008. Dietary intake of lactose as a strong predictor for secretor status of nipple aspirate fluid in healthy premenopausal nonlactating women. *Clinical Cancer Research*, 14, 1386-1392.
- HULTIN-ROSENBERG, L., FORSHED, J., BRANCA, R. M., LEHTIÖ, J. & JOHANSSON, H. J. 2013. Defining, comparing, and improving iTRAQ quantification in mass spectrometry proteomics data. *Molecular & Cellular Proteomics*, 12, 2021-2031.
- HUNT, D. F., YATES, J. R., SHABANOWITZ, J., WINSTON, S. & HAUER, C. R. 1986. Protein sequencing by tandem mass spectrometry. *Proceedings of the National Academy of Sciences*, 83, 6233-6237.
- HÜTTENHAIN, R., SOSTE, M., SELEVSEK, N., RÖST, H., SETHI, A., CARAPITO, C., FARRAH, T., DEUTSCH, E. W., KUSEBAUCH, U. & MORITZ, R. L. 2012. Reproducible quantification of cancer-associated proteins in body fluids using targeted proteomics. *Science translational medicine*, 4, 142ra94-142ra94.
- IBARRA-DRENDALL, C., TROCH, M. M., BARRY, W. T., BROADWATER, G., PETRICOIN III, E. F., WULFKUHLE, J., LIOTTA, L. A., LEM, S., BAKER JR, J. C. & FORD, A. C. 2012. Pilot and feasibility study: prospective proteomic profiling of mammary epithelial cells from high-risk women provides evidence of activation of pro-survival pathways. *Breast cancer research and treatment*, 132, 487-498.
- JEGO, G., HAZOUMÉ, A., SEIGNEURIC, R. & GARRIDO, C. 2013. Targeting heat shock proteins in cancer. *Cancer letters*, 332, 275-285.
- JOHNS, L. E., MOSS, S. M. & AGE TRIAL MANAGEMENT, G. 2010. False-positive results in the randomized controlled trial of mammographic screening from age 40 ("Age" trial). *Cancer Epidemiol Biomarkers Prev*, 19, 2758-64.
- JONES, E. & SPENCER, S. A. 2007. Optimising the provision of human milk for preterm infants. *Arch Dis Child Fetal Neonatal Ed*, 92, F236-8.
- JUHASZ, P., ROSKEY, M. T., SMIRNOV, I. P., HAFF, L. A., VESTAL, M. L. & MARTIN, S. A. 1996. Applications of delayed extraction matrix-assisted laser desorption ionization time-of-

- flight mass spectrometry to oligonucleotide analysis. *Analytical chemistry*, 68, 941-946.
- JURINKE, C., OETH, P. & VAN DEN BOOM, D. 2004. MALDI-TOF mass spectrometry. *Molecular biotechnology*, 26, 147-163.
- KAGEYAMA, S., ISONO, T., MATSUDA, S., USHIO, Y., SATOMURA, S., TERAJ, A., ARAI, Y., KAWAKITA, M., OKADA, Y. & YOSHIKI, T. 2009. Urinary calreticulin in the diagnosis of bladder urothelial carcinoma. *International journal of urology*, 16, 481-486.
- KÄLL, L., CANTERBURY, J. D., WESTON, J., NOBLE, W. S. & MACCOSS, M. J. 2007. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nature methods*, 4, 923-925.
- KARAS, M. & HILLENKAMP, F. 1988. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical chemistry*, 60, 2299-2301.
- KARP, N. A., HUBER, W., SADOWSKI, P. G., CHARLES, P. D., HESTER, S. V. & LILLEY, K. S. 2010. Addressing accuracy and precision issues in iTRAQ quantitation. *Molecular & Cellular Proteomics*, 9, 1885-1897.
- KATSYV, I., WANG, M., SONG, W. M., ZHOU, X., ZHAO, Y., PARK, S., ZHU, J., ZHANG, B. & IRIE, H. Y. 2016. EPRS is a critical regulator of cell proliferation and estrogen signaling in ER+ breast cancer. *Oncotarget*, 7, 69592-69605.
- KENNEDY, J. J., ABBATIELLO, S. E., KIM, K., YAN, P., WHITEAKER, J. R., LIN, C., KIM, J. S., ZHANG, Y., WANG, X. & IVEY, R. G. 2014. Demonstrating the feasibility of large-scale development of standardized assays to quantify human proteins. *Nature methods*, 11, 149-155.
- KERSEY, P. J., DUARTE, J., WILLIAMS, A., KARAVIDOPOULOU, Y., BIRNEY, E. & APWEILER, R. 2004. The International Protein Index: an integrated database for proteomics experiments. *Proteomics*, 4, 1985-1988.
- KESHAMOUNI, V. G., MICHAILIDIS, G., GRASSO, C. S., ANTHWAL, S., STRAHLER, J. R., WALKER, A., ARENBERG, D. A., REDDY, R. C., AKULAPALLI, S. & THANNICKAL, V. J. 2006a. Differential protein expression profiling by iTRAQ-2DLC-MS/MS of lung cancer cells undergoing epithelial-mesenchymal transition reveals a migratory/invasive phenotype. *Journal of proteome research*, 5, 1143-1154.
- KESHAMOUNI, V. G., MICHAILIDIS, G., GRASSO, C. S., ANTHWAL, S., STRAHLER, J. R., WALKER, A., ARENBERG, D. A., REDDY, R. C., AKULAPALLI, S. & THANNICKAL, V. J. 2006b. Differential protein expression profiling by iTRAQ- 2DLC- MS/MS of lung cancer cells undergoing epithelial-mesenchymal transition reveals a migratory/invasive phenotype. *Journal of proteome research*, 5, 1143-1154.
- KESHISHIAN, H., ADDONA, T., BURGESS, M., MANI, D., SHI, X., KUHN, E., SABATINE, M. S., GERSZTEN, R. E. & CARR, S. A. 2009. Quantification of cardiovascular biomarkers in patient plasma by targeted mass spectrometry and stable isotope dilution. *Molecular & cellular proteomics*, 8, 2339-2349.
- KHAN, S. A., LANKES, H. A., PATIL, D. B., BRYK, M., HOU, N., IVANCIC, D., NAYAR, R., MASOOD, S. & RADEMAKER, A. 2009. Ductal lavage is an inefficient method of biomarker measurement in high-risk women. *Cancer Prev Res (Phila)*, 2, 265-73.
- KHAN, S. A., WILEY, E. L., RODRIGUEZ, N., BAIRD, C., RAMAKRISHNAN, R., NAYAR, R., BRYK, M., BETHKE, K. B., STARADUB, V. L. & WOLFMAN, J. 2004. Ductal lavage findings in women with known breast cancer undergoing mastectomy. *Journal of the National Cancer Institute*, 96, 1510-1517.
- KIM, M.-S., PINTO, S. M., GETNET, D., NIRUJOGI, R. S., MANDA, S. S., CHAERKADY, R., MADUGUNDU, A. K., KELKAR, D. S., ISSERLIN, R. & JAIN, S. 2014. A draft map of the human proteome. *Nature*, 509, 575-581.
- KIM, S., LEE, Y. & KOO, J. S. 2015. Differential expression of lipid metabolism-related proteins in different breast cancer subtypes. *PloS one*, 10, e0119473.
- KIM, T. K. & EBERWINE, J. H. 2010. Mammalian cell transfection: the present and the future. *Analytical and bioanalytical chemistry*, 397, 3173-3178.

- KIMURA, H., FUMOTO, K., SHOJIMA, K., NOJIMA, S., OSUGI, Y., TOMIHARA, H., EGUCHI, H., SHINTANI, Y., ENDO, H. & INOUE, M. 2016. CKAP4 is a Dickkopf1 receptor and is involved in tumor progression. *The Journal of clinical investigation*, 126, 2689-2705.
- KING, M.-C., MARKS, J. H. & MANDELL, J. B. 2003. Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. *Science*, 302, 643-646.
- KING, T. B., COLBY, S. M. & REILLY, J. P. 1995. High resolution MALDI-TOF mass spectra of three proteins obtained using space—velocity correlation focusing. *International journal of mass spectrometry and ion processes*, 145, L1-L7.
- KITTERINGHAM, N. R., JENKINS, R. E., LANE, C. S., ELLIOTT, V. L. & PARK, B. K. 2009. Multiple reaction monitoring for quantitative biomarker analysis in proteomics and metabolomics. *Journal of Chromatography B*, 877, 1229-1239.
- KÖCHER, T., PICHLER, P., SCHUTZBIER, M., STINGL, C., KAUL, A., TEUCHER, N., HASENFUSS, G., PENNINGER, J. M. & MECHTLER, K. 2009. High precision quantitative proteomics using iTRAQ on an LTQ Orbitrap: a new mass spectrometric method combining the benefits of all. *Journal of proteome research*, 8, 4743-4752.
- KOLB, T. M., LICHY, J. & NEWHOUSE, J. H. 2002. Comparison of the performance of screening mammography, physical examination, and breast US and evaluation of factors that influence them: an analysis of 27,825 patient evaluations. *Radiology*, 225, 165-75.
- KONG, D., LI, Y., WANG, Z. & SARKAR, F. H. 2011. Cancer stem cells and epithelial-to-mesenchymal transition (EMT)-phenotypic cells: are they cousins or twins? *Cancers*, 3, 716-729.
- KOZAK, M. 2007. Some thoughts about translational regulation: forward and backward glances. *Journal of cellular biochemistry*, 102, 280-290.
- KROGH, A., LARSSON, B., VON HEIJNE, G. & SONNHAMMER, E. L. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology*, 305, 567-580.
- KROUSKOP, T. A., WHEELER, T. M., KALLEL, F., GARRA, B. S. & HALL, T. 1998. Elastic moduli of breast and prostate tissues under compression. *Ultrason Imaging*, 20, 260-74.
- KRUTCHINSKY, A. N., KALKUM, M. & CHAIT, B. T. 2001. Automatic identification of proteins with a MALDI-quadrupole ion trap mass spectrometer. *Analytical chemistry*, 73, 5066-5077.
- KUMAR, J. K., TABOR, S. & RICHARDSON, C. C. 2004. Proteomic analysis of thioredoxin-targeted proteins in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 3759-3764.
- KUMAR, S. R., SAUTER, E. R., QUINN, T. P. & DEUTSCHER, S. L. 2005. Thomsen-Friedenreich and Tn antigens in nipple fluid: carbohydrate biomarkers for breast cancer detection. *Clinical cancer research*, 11, 6868-6871.
- KURONO, S., KANEKO, Y., MATSUURA, N., OISHI, H., NOGUCHI, S., KIM, S. J., TAMAKI, Y., AIKAWA, T., KOTSUMA, Y., INAJI, H. & MATSUURA, S. 2016. Identification of potential breast cancer markers in nipple discharge by protein profile analysis using two-dimensional nano-liquid chromatography/nanoelectrospray ionization-mass spectrometry. *Proteomics Clin Appl*, 10, 605-13.
- LACROIX-TRIKI, M., GEYER, F. C., LAMBROS, M. B., SAVAGE, K., ELLIS, I. O., LEE, A. H. & REIS-FILHO, J. S. 2010. beta-catenin/Wnt signalling pathway in fibromatosis, metaplastic carcinomas and phyllodes tumours of the breast. *Mod Pathol*, 23, 1438-48.
- LAÉ, M., FRÉNEAUX, P., SASTRE-GARAU, X., CHOUCANE, O., SIGAL-ZAFRANI, B. & VINCENT-SALOMON, A. 2009. Secretory breast carcinomas with ETV6-NTRK3 fusion gene belong to the basal-like carcinoma spectrum. *Modern Pathology*, 22, 291-298.
- LAMOND, A. I., UHLEN, M., HORNING, S., MAKAROV, A., ROBINSON, C. V., SERRANO, L., HARTL, F. U., BAUMEISTER, W., WERENSKIOLD, A. K. & ANDERSEN, J. S. 2012. Advancing cell biology through proteomics in space and time (PROSPECTS). *Molecular & Cellular Proteomics*, 11, O112. 017731.

- LANGE, V., PICOTTI, P., DOMON, B. & AEBERSOLD, R. 2008. Selected reaction monitoring for quantitative proteomics: a tutorial. *Molecular systems biology*, 4, 222.
- LAWRENCE, R. T., PEREZ, E. M., HERNÁNDEZ, D., MILLER, C. P., HAAS, K. M., IRIE, H. Y., LEE, S.-I., BLAU, C. A. & VILLÉN, J. 2015. The proteomic landscape of triple-negative breast cancer. *Cell reports*, 11, 630-644.
- LECHNER, L., DE NOOIJER, J. & DE VRIES, H. 2004. Breast self-examination: longitudinal predictors of intention and subsequent behaviour. *European Journal of Cancer Prevention*, 13, 369-376.
- LEE, M. M., WRENSCH, M. R., MIIKE, R. & PETRAKIS, N. 1992. The association of dietary fat with ability to obtain breast fluid by nipple aspiration. *Cancer Epidemiology Biomarkers & Prevention*, 1, 277-280.
- LEI, J., YANG, P., ZHANG, L., WANG, Y. & YANG, K. 2014. Diagnostic accuracy of digital breast tomosynthesis versus digital mammography for benign and malignant lesions in breasts: a meta-analysis. *Eur Radiol*, 24, 595-602.
- LEIDY, J., KHAN, A. & KANDIL, D. 2014. Basal-like breast cancer: update on clinicopathologic, immunohistochemical, and molecular features. *Archives of Pathology and Laboratory Medicine*, 138, 37-43.
- LEROY, B., ANDERSON, M. & SOUSSI, T. 2014. TP53 mutations in human cancer: database reassessment and prospects for the next decade. *Human mutation*, 35, 672-688.
- LEROY, B., FOURNIER, J. L., ISHIOKA, C., MONTI, P., INGA, A., FRONZA, G. & SOUSSI, T. 2013. The TP53 website: an integrative resource centre for the TP53 mutation database and TP53 mutant analysis. *Nucleic acids research*, 41, D962-D969.
- LESUR, A. & DOMON, B. 2015. Advances in high-resolution accurate mass spectrometry application to targeted proteomics. *Proteomics*, 15, 880-890.
- LI, S.-X., LIU, L.-J., DONG, L.-W., SHI, H.-G., PAN, Y.-F., TAN, Y.-X., ZHANG, J., ZHANG, B., DING, Z.-W. & JIANG, T.-Y. 2014a. CKAP4 inhibited growth and metastasis of hepatocellular carcinoma through regulating EGFR signaling. *Tumor Biology*, 35, 7999-8005.
- LI, S. X., TANG, G. S., ZHOU, D. X., PAN, Y. F., TAN, Y. X., ZHANG, J., ZHANG, B., DING, Z. W., LIU, L. J. & JIANG, T. Y. 2014b. Prognostic significance of cytoskeleton-associated membrane protein 4 and its palmitoyl acyltransferase DHHC2 in hepatocellular carcinoma. *Cancer*, 120, 1520-1531.
- LI, W., WU, C., YAO, Y., DONG, B., WEI, Z., LV, X., ZHANG, J. & XU, Y. 2014c. MUC4 modulates human glioblastoma cell proliferation and invasion by upregulating EGFR expression. *Neuroscience letters*, 566, 82-87.
- LI, Z., ADAMS, R. M., CHOUREY, K., HURST, G. B., HETTICH, R. L. & PAN, C. 2012. Systematic comparison of label-free, metabolic labeling, and isobaric chemical labeling for quantitative proteomics on LTQ Orbitrap Velos. *Journal of proteome research*, 11, 1582-1590.
- LIANOS, G. D., ALEXIOU, G. A., MANGANO, A., MANGANO, A., RAUSEI, S., BONI, L., DIONIGI, G. & ROUKOS, D. H. 2015. The role of heat shock proteins in cancer. *Cancer letters*, 360, 114-118.
- LIEBLER, D. 2001. *Introduction to proteomics: tools for the new biology*, Springer Science & Business Media.
- LIEBLER, D. C. & ZIMMERMAN, L. J. 2013. Targeted quantitation of proteins by mass spectrometry. *Biochemistry*, 52, 3797-3806.
- LIEN, H. C., WANG, C. C., HUANG, C. S., YANG, Y. W., KUO, W. H. & YAO, Y. T. 2013. Ubiquitin carboxy-terminal hydrolase L1 may be involved in the development of mammary phyllodes tumors. *Virchows Arch*, 462, 155-61.
- LIN, W.-T., HUNG, W.-N., YIAN, Y.-H., WU, K.-P., HAN, C.-L., CHEN, Y.-R., CHEN, Y.-J., SUNG, T.-Y. & HSU, W.-L. 2006. Multi-Q: a fully automated tool for multiplexed protein quantitation. *Journal of proteome research*, 5, 2328-2338.

- LITHGOW, D., NYAMATHI, A., ELASHOFF, D., MARTINEZ-MAZA, O. & COVINGTON, C. 2007. C-reactive protein in nipple aspirate fluid associated with Gail model factors. *Biological research for nursing*, 9, 108-116.
- LIVASY, C. A., KARACA, G., NANDA, R., TRETIAKOVA, M. S., OLOPADE, O. I., MOORE, D. T. & PEROU, C. M. 2006. Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma. *Modern pathology*, 19, 264-271.
- LOBERG, M., LOUSDAL, M. L., BRETTHAUER, M. & KALAGER, M. 2015. Benefits and harms of mammography screening. *Breast Cancer Res*, 17, 63.
- LOBODA, A., KRUTCHINSKY, A., BROMIRSKI, M., ENS, W. & STANDING, K. 2000. A tandem quadrupole/time-of-flight mass spectrometer with a matrix-assisted laser desorption/ionization source: Design and performance. *Rapid Communications in Mass Spectrometry*, 14, 1047-1057.
- LOO, J., YAN, W., RAMACHANDRAN, P. & WONG, D. 2010. Comparative human salivary and plasma proteomes. *Journal of dental research*, 89, 1016-1023.
- LOPEZ, J. 2007. Two-dimensional electrophoresis in proteome expression analysis. *Journal of chromatography B*, 849, 190-202.
- LOUDERBOUGH, J. M. & SCHROEDER, J. A. 2011. Understanding the dual nature of CD44 in breast cancer progression. *Molecular Cancer Research*, 9, 1573-1586.
- LU, Y.-C., WENG, W.-C. & LEE, H. 2015. Functional roles of calreticulin in cancer biology. *BioMed research international*, 2015.
- MACKAY, A., WEIGELT, B., GRIGORIADIS, A., KREIKE, B., NATRAJAN, R., A'HERN, R., TAN, D. S., DOWSETT, M., ASHWORTH, A. & REIS-FILHO, J. S. 2011. Microarray-based class discovery for molecular classification of breast cancer: analysis of interobserver agreement. *Journal of the National Cancer Institute*, 103, 662-673.
- MACLEAN, B., TOMAZELA, D. M., SHULMAN, N., CHAMBERS, M., FINNEY, G. L., FREWEN, B., KERN, R., TABB, D. L., LIEBLER, D. C. & MACCOSS, M. J. 2010. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics*, 26, 966-968.
- MAHER, C. A., KUMAR-SINHA, C., CAO, X., KALYANA-SUNDARAM, S., HAN, B., JING, X., SAM, L., BARRETTE, T., PALANISAMY, N. & CHINNAIYAN, A. M. 2009. Transcriptome sequencing to detect gene fusions in cancer. *Nature*, 458, 97-101.
- MAHONEY, D. W., THERNEAU, T. M., HEPPELMANN, C. J., HIGGINS, L., BENSON, L. M., ZENKA, R. M., JAGTAP, P., NELSESTUEN, G. L., BERGEN III, H. R. & OBERG, A. L. 2011. Relative quantification: characterization of bias, variability and fold changes in mass spectrometry data from iTRAQ-labeled peptides. *Journal of proteome research*, 10, 4325-4333.
- MAIER, T., GÜELL, M. & SERRANO, L. 2009. Correlation of mRNA and protein in complex biological samples. *FEBS letters*, 583, 3966-3973.
- MAKAROV, A., DENISOV, E., KHOLOMEEV, A., BALSCHUN, W., LANGE, O., STRUPAT, K. & HORNING, S. 2006. Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Analytical chemistry*, 78, 2113-2120.
- MAKAWITA, S., DIMITROMANOLAKIS, A., SOOSAIPILLAI, A., SOLEAS, I., CHAN, A., GALLINGER, S., HAUN, R. S., BLASUTIG, I. M. & DIAMANDIS, E. P. 2013. Validation of four candidate pancreatic cancer serological biomarkers that improve the performance of CA19. 9. *BMC cancer*, 13, 1.
- MAKKI, J., MYINT, O., WYNN, A. A., SAMSUDIN, A. T. & JOHN, D. V. 2015. Expression distribution of cancer stem cells, epithelial to mesenchymal transition, and telomerase activity in breast cancer and their association with clinicopathologic characteristics. *Clinical medicine insights. Pathology*, 8, 1.
- MANADAS, B., MENDES, V. M., ENGLISH, J. & DUNN, M. J. 2010. Peptide fractionation in proteomics approaches. *Expert review of proteomics*, 7, 655-663.

- MANDELSON, M. T., OESTREICHER, N., PORTER, P. L., WHITE, D., FINDER, C. A., TAPLIN, S. H. & WHITE, E. 2000. Breast density as a predictor of mammographic detection: comparison of interval- and screen-detected cancers. *J Natl Cancer Inst*, 92, 1081-7.
- MANNELLO, F. & LIGI, D. 2013. Resolving breast cancer heterogeneity by searching reliable protein cancer biomarkers in the breast fluid secretome. *BMC Cancer*, 13, 344.
- MANNELLO, F., LIGI, D. & CANALE, M. 2013. Aluminium, carbonyls and cytokines in human nipple aspirate fluids: Possible relationship between inflammation, oxidative stress and breast cancer microenvironment. *Journal of inorganic biochemistry*, 128, 250-256.
- MANNELLO, F., TONTI, G. A. & DARBRE, P. D. 2009. Concentration of aluminium in breast cyst fluids collected from women affected by gross cystic breast disease. *Journal of Applied Toxicology*, 29, 1-6.
- MANNELLO, F., TONTI, G. A., MEDDA, V., SIMONE, P. & DARBRE, P. D. 2011. Analysis of aluminium content and iron homeostasis in nipple aspirate fluids from healthy women and breast cancer-affected patients. *Journal of Applied Toxicology*, 31, 262-269.
- MANNELLO, F., TONTI, G. A., PEDERZOLI, A., SIMONE, P., SMANIOTTO, A. & MEDDA, V. 2010. Detection of Superoxide Dismutase-1 in Nipple Aspirate Fluids: A Reactive Oxygen Species—Regulating Enzyme in the Breast Cancer Microenvironment. *Clinical breast cancer*, 10, 238-245.
- MANNELLO, F., TONTI, G. A. M. & PAPA, S. 2006. Human gross cyst breast disease and cystic fluid: bio-molecular, morphological, and clinical studies. *Breast cancer research and treatment*, 97, 115-129.
- MARMET, P. 1971. Quadrupole Mass Analyzers. *Journal of Vacuum Science & Technology*, 8, 262-262.
- MARTIN, L. J. & BOYD, N. F. 2008. Mammographic density. Potential mechanisms of breast cancer risk associated with mammographic density: hypotheses based on epidemiological evidence. *Breast Cancer Res*, 10, 201.
- MASKARINEC, G., HEBISHI, S., CUSTER, L. & FRANKE, A. A. 2008. The relation of soy intake and isoflavone levels in nipple aspirate fluid. *European Journal of Cancer Prevention*, 17, 67-70.
- MASKARINEC, G., OLLBERDING, N. J., CONROY, S. M., MORIMOTO, Y., PAGANO, I. S., FRANKE, A. A., GENTZSCHEIN, E. & STANCZYK, F. Z. 2011. Estrogen levels in nipple aspirate fluid and serum during a randomized soy trial. *Cancer Epidemiology Biomarkers & Prevention*, 20, 1815-1821.
- MATRIX-SCIENCE. 2016. *Mascot Server* [Online]. Matrix-Science. Available: http://www.matrixscience.com/search_form_select.html [Accessed 12/09 2016].
- MATT, P., FU, Z., FU, Q. & VAN EYK, J. E. 2008. Biomarker discovery: proteome fractionation and separation in biological samples. *Physiological genomics*, 33, 12-17.
- MAYER, I. A., ABRAMSON, V. G., LEHMANN, B. D. & PIETENPOL, J. A. 2014. New strategies for triple-negative breast cancer—deciphering the heterogeneity. *Clinical cancer research*, 20, 782-790.
- MCALISTER, G. C., NUSINOW, D. P., JEDRYCHOWSKI, M. P., WÜHR, M., HUTTLIN, E. L., ERICKSON, B. K., RAD, R., HAAS, W. & GYGI, S. P. 2014. MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. *Analytical chemistry*, 86, 7150-7158.
- MEDZIHRADESKY, K. F., CAMPBELL, J. M., BALDWIN, M. A., FALICK, A. M., JUHASZ, P., VESTAL, M. L. & BURLINGAME, A. L. 2000. The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. *Analytical chemistry*, 72, 552-558.
- MENG, Z. & VEENSTRA, T. D. 2011. Targeted mass spectrometry approaches for protein biomarker verification. *Journal of proteomics*, 74, 2650-2659.
- METZKER, M. L. 2010. Sequencing technologies—the next generation. *Nature reviews genetics*, 11, 31-46.

- MICHALAK, M., WARNKEN, U., ANDRÉ, S., SCHNÖLZER, M., GABIUS, H.-J. & KOPITZ, J. 2016. Detection of proteome changes in human colon cancer induced by cell surface binding of growth-inhibitory human galectin-4 using quantitative SILAC-based proteomics. *Journal of Proteome Research*, 15, 4412-4422.
- MICHALSKI, A., DAMOC, E., LANGE, O., DENISOV, E., NOLTING, D., MÜLLER, M., VINER, R., SCHWARTZ, J., REMES, P. & BELFORD, M. 2012. Ultra high resolution linear ion trap Orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes. *Molecular & Cellular Proteomics*, 11, O111. 013698.
- MINER, J. H. & PATTON, B. L. 1999. Laminin-11. *Int J Biochem Cell Biol*, 31, 811-6.
- MITCHELL, G., ANTILL, Y. C., MURRAY, W., KIRK, J., SALISBURY, E., LINDEMAN, G. J., DI IULIO, J., MILNER, A. D., DEVEREAUX, L. & PHILLIPS, K.-A. 2005. Nipple aspiration and ductal lavage in women with a germline BRCA1 or BRCA2 mutation. *Breast Cancer Res*, 7, 1122-31.
- MIYOSHI, Y., TAGUCHI, T., KIM, S. J., TAMAKI, Y. & NOGUCHI, S. 2005. Prediction of response to docetaxel by immunohistochemical analysis of CYP3A4 expression in human breast cancers. *Breast Cancer*, 12, 11-5.
- MOORE, M. 2016. *Activating senescence in p16-positive Basal-like breast cancer*. Activating senescence in p16-positive Basal-like breast cancer.
- MOORE, S. M., HESS, S. M. & JORGENSON, J. W. 2016. Extraction, enrichment, solubilization, and digestion techniques for membrane proteomics. *Journal of proteome research*, 15, 1243-1252.
- MORIMOTO, R. I. 1993. Cells in stress: transcriptional activation of heat shock genes. *SCIENCE-NEW YORK THEN WASHINGTON-*, 259, 1409-1409.
- MOSS, S. M., CUCKLE, H., EVANS, A., JOHNS, L., WALLER, M., BOBROW, L. & TRIAL MANAGEMENT, G. 2006. Effect of mammographic screening from age 40 years on breast cancer mortality at 10 years' follow-up: a randomised controlled trial. *Lancet*, 368, 2053-60.
- MUKOHARA, T. 2015. PI3K mutations in breast cancer: prognostic and therapeutic implications. *Breast Cancer: Targets and Therapy*, 7, 111.
- MUSHLIN, A. I., KOUIDES, R. W. & SHAPIRO, D. E. 1998. Estimating the accuracy of screening mammography: a meta-analysis. *Am J Prev Med*, 14, 143-53.
- NAGARAJ, N., WISNIEWSKI, J. R., GEIGER, T., COX, J., KIRCHER, M., KELSO, J., PÄÄBO, S. & MANN, M. 2011. Deep proteome and transcriptome mapping of a human cancer cell line. *Molecular systems biology*, 7, 548.
- NCBI. 2017. *PubMed; Biomedical literature from MEDLINE, life science journals, and online books* [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/?term=> [Accessed 25/02 2017].
- NEAGU, M., CONSTANTIN, C., TANASE, C. & BODA, D. 2011. Patented biomarker panels in early detection of cancer. *Recent patents on biomarkers*, 1, 10-24.
- NELSON, H. D., TYNE, K., NAIK, A., BOUGATSOS, C., CHAN, B. K., HUMPHREY, L. & FORCE, U. S. P. S. T. 2009. Screening for breast cancer: an update for the U.S. Preventive Services Task Force. *Ann Intern Med*, 151, 727-37, W237-42.
- NESVIZHSKII, A. I., KELLER, A., KOLKER, E. & AEBERSOLD, R. 2003. A statistical model for identifying proteins by tandem mass spectrometry. *Analytical chemistry*, 75, 4646-4658.
- NEUMANN, M., NEVES, R., SCHULZ, S., NEUBAUER, H., FLEISCH, M., FEHM, T. & NIEDERACHER, D. 2014. Macrophage-capping protein (CapG) as a putative oncogene is overexpressed in invasive breast carcinoma cells. *Oncology Research and Treatment*, 37, 20.
- NICHOLSON, R., GEE, J. & HARPER, M. 2001. EGFR and cancer prognosis. *European journal of cancer*, 37, 9-15.
- NIELSEN, T. O., HSU, F. D., JENSEN, K., CHEANG, M., KARACA, G., HU, Z., HERNANDEZ-BOUSSARD, T., LIVASY, C., COWAN, D. & DRESSLER, L. 2004. Immunohistochemical and

- clinical characterization of the basal-like subtype of invasive breast carcinoma. *Clinical Cancer Research*, 10, 5367-5374.
- NIKOLENKO, G. N., STENGELIN, M. K., SARDESAI, L., GLEZER, E. N. & WOHLSTADTER, J. N. 2015. Accurate measurement of free and complexed PSA concentrations in serum of women using a novel technology with fg/mL sensitivity. AACR.
- O'NEILL, F., MADDEN, S. F., AHERNE, S. T., CLYNES, M., CROWN, J., DOOLAN, P. & O'CONNOR, R. 2012. Gene expression changes as markers of early lapatinib response in a panel of breast cancer cell lines. *Mol Cancer*, 11, 41.
- O'DONNELL, R., HOLLAND, J., DEETH, H. & ALEWOOD, P. 2004. Milk proteomics. *International Dairy Journal*, 14, 1013-1023.
- OBEID, M., TESNIERE, A., GHIRINGHELLI, F., FIMIA, G. M., APETOH, L., PERFETTINI, J.-L., CASTEDO, M., MIGNOT, G., PANARETAKIS, T. & CASARES, N. 2007. Calreticulin exposure dictates the immunogenicity of cancer cell death. *Nature medicine*, 13, 54-61.
- OBENAUER, S., LUFTNER-NAGEL, S., VON HEYDEN, D., MUNZEL, U., BAUM, F. & GRABBE, E. 2002. Screen film vs full-field digital mammography: image quality, detectability and characterization of lesions. *Eur Radiol*, 12, 1697-702.
- ODA, M., MAKITA, M., IWAYA, K., AKIYAMA, F., KOHNO, N., TSUCHIYA, B., IWASE, T. & MATSUBARA, O. 2012. High levels of DJ-1 protein in nipple fluid of patients with breast cancer. *Cancer science*, 103, 1172-1176.
- OLD, W. M., MEYER-ARENDR, K., AVELINE-WOLF, L., PIERCE, K. G., MENDOZA, A., SEVINSKY, J. R., RESING, K. A. & AHN, N. G. 2005. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Molecular & cellular proteomics*, 4, 1487-1502.
- OLIVIER, M., LANGER, A., CARRIERI, P., BERGH, J., KLAAR, S., EYFJORD, J., THEILLET, C., RODRIGUEZ, C., LIDEREAU, R. & BI, I. 2006. The clinical value of somatic TP53 gene mutations in 1,794 patients with breast cancer. *Clinical cancer research*, 12, 1157-1167.
- OLSEN, J. V., MACEK, B., LANGE, O., MAKAROV, A., HORNING, S. & MANN, M. 2007. Higher-energy C-trap dissociation for peptide modification analysis. *Nature methods*, 4, 709-712.
- OLSEN, J. V., ONG, S.-E. & MANN, M. 2004. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Molecular & Cellular Proteomics*, 3, 608-614.
- OMENN, G. S., STATES, D. J., ADAMSKI, M., BLACKWELL, T. W., MENON, R., HERMJAKOB, H., APWEILER, R., HAAB, B. B., SIMPSON, R. J., EDDER, J. S., KAPP, E. A., MORITZ, R. L., CHAN, D. W., RAI, A. J., ADMON, A., AEBERSOLD, R., ENG, J., HANCOCK, W. S., HEFTA, S. A., MEYER, H., PAIK, Y. K., YOO, J. S., PING, P., POUNDS, J., ADKINS, J., QIAN, X., WANG, R., WASINGER, V., WU, C. Y., ZHAO, X., ZENG, R., ARCHAKOV, A., TSUGITA, A., BEER, I., PANDEY, A., PISANO, M., ANDREWS, P., TAMMEN, H., SPEICHER, D. W. & HANASH, S. M. 2005. Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *Proteomics*, 5, 3226-45.
- ONG, S.-E. 2012. The expanding field of SILAC. *Analytical and bioanalytical chemistry*, 404, 967-976.
- ONG, S.-E., BLAGOEV, B., KRATCHMAROVA, I., KRISTENSEN, D. B., STEEN, H., PANDEY, A. & MANN, M. 2002. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & cellular proteomics*, 1, 376-386.
- ONG, S.-E. & MANN, M. 2005. Mass spectrometry-based proteomics turns quantitative. *Nature chemical biology*, 1, 252-262.

- OW, S. Y., SALIM, M., NOIREL, J., EVANS, C., REHMAN, I. & WRIGHT, P. C. 2009. iTRAQ underestimation in simple and complex mixtures: "the good, the bad and the ugly". *Journal of proteome research*, 8, 5347-5355.
- OW, S. Y., SALIM, M., NOIREL, J., EVANS, C. & WRIGHT, P. 2011. Minimising iTRAQ ratio compression through understanding LC-MS elution dependence and high-resolution HILIC fractionation. *Proteomics*, 11, 2341-2346.
- OXELMARK, E., ROTH, J. M., BROOKS, P. C., BRAUNSTEIN, S. E., SCHNEIDER, R. J. & GARABEDIAN, M. J. 2006. The cochaperone p23 differentially regulates estrogen receptor target genes and promotes tumor cell adhesion and invasion. *Molecular and cellular biology*, 26, 5205-5213.
- PALMER, D. J., KELLY, V. C., SMIT, A. M., KUY, S., KNIGHT, C. G. & COOPER, G. J. 2006. Human colostrum: identification of minor proteins in the aqueous phase by proteomics. *Proteomics*, 6, 2208-2216.
- PAN, S. & AEBERSOLD, R. 2007. Quantitative proteomics by stable isotope labeling and mass spectrometry. *Mass Spectrometry Data Analysis in Proteomics*, 209-218.
- PARK, J.-S., KIM, H.-Y., KIM, H.-W., CHAE, G.-N., OH, H.-T., PARK, J.-Y., SHIM, H., SEO, M., SHIN, E.-Y. & KIM, E.-G. 2005. Increased caveolin-1, a cause for the declined adipogenic potential of senescent human mesenchymal stem cells. *Mechanisms of ageing and development*, 126, 551-559.
- PARK, S.-W., ZHEN, G., VERHAEGHE, C., NAKAGAMI, Y., NGUYENVU, L. T., BARCZAK, A. J., KILLEEN, N. & ERLE, D. J. 2009. The protein disulfide isomerase AGR2 is essential for production of intestinal mucus. *Proceedings of the National Academy of Sciences*, 106, 6950-6955.
- PAUL, D., KUMAR, A., GAJBHIYE, A., SANTRA, M. K. & SRIKANTH, R. 2013. Mass spectrometry-based proteomics in molecular diagnostics: discovery of cancer biomarkers using tissue culture. *BioMed research international*, 2013.
- PAUL, M. K. & MUKHOPADHYAY, A. K. 2004. Tyrosine kinase—Role and significance in Cancer. *International journal of medical sciences*, 1, 101.
- PAVLIDES, S., TSIRIGOS, A., MIGNECO, G., WHITAKER-MENEZES, D., CHIAVARINA, B., FLOMENBERG, N., FRANK, P. G., CASIMIRO, M. C., WANG, C. & PESTELL, R. G. 2010. The autophagic tumor stroma model of cancer: Role of oxidative stress and ketone production in fueling tumor cell metabolism. *Cell Cycle*, 9, 3485.
- PAVLOU, M. P., KULASINGAM, V., SAUTER, E. R., KLIETHERMES, B. & DIAMANDIS, E. P. 2010. Nipple aspirate fluid proteome of healthy females and patients with breast cancer. *Clinical chemistry*, 56, 848-855.
- PAWELETZ, C. P., TROCK, B., PENNANEN, M., TSANGARIS, T., MAGNANT, C., LIOTTA, L. A. & PETRICOIN III, E. F. 2001. Proteomic patterns of nipple aspirate fluids obtained by SELDI-TOF: potential for new biomarkers to aid in the diagnosis of breast cancer. *Disease markers*, 17, 301-307.
- PAWLIK, T. M., HAWKE, D. H., LIU, Y., KRISHNAMURTHY, S., FRITSCH, H., HUNT, K. K. & KUERER, H. M. 2006. Proteomic analysis of nipple aspirate fluid from women with early-stage breast cancer using isotope-coded affinity tags and tandem mass spectrometry reveals differential expression of vitamin D binding protein. *BMC cancer*, 6, 68.
- PEKAR SECOND, T., BLETHROW, J. D., SCHWARTZ, J. C., MERRIHEW, G. E., MACCOSS, M. J., SWANEY, D. L., RUSSELL, J. D., COON, J. J. & ZABROUSKOV, V. 2009. Dual-pressure linear ion trap mass spectrometer improving the analysis of complex protein mixtures. *Analytical chemistry*, 81, 7757-7765.
- PEPTIDEATLAS. 2017. *PeptideAtlas; tandem mass spectrometry proteomics experiment data base* [Online]. Available: <http://www.peptideatlas.org/> [Accessed 18/01 2017].
- PERCY, A. J., CHAMBERS, A. G., YANG, J. & BORCHERS, C. H. 2013. Multiplexed MRM-based quantitation of candidate cancer biomarker proteins in undepleted and non-enriched human plasma. *Proteomics*, 13, 2202-2215.

- PERCY, A. J., CHAMBERS, A. G., YANG, J., HARDIE, D. B. & BORCHERS, C. H. 2014. Advances in multiplexed MRM-based protein biomarker quantitation toward clinical utility. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1844, 917-926.
- PERNEMALM, M. & LEHTIÖ, J. 2012. A novel prefractionation method combining protein and peptide isoelectric focusing in immobilized pH gradient strips. *Journal of proteome research*, 12, 1014-1019.
- PEROU, C. M. 2011. Molecular stratification of triple-negative breast cancers. *The oncologist*, 16, 61-70.
- PEROU, C. M., SØRLIE, T., EISEN, M. B., VAN DE RIJN, M., JEFFREY, S. S., REES, C. A., POLLACK, J. R., ROSS, D. T., JOHNSEN, H. & AKSLEN, L. A. 2000. Molecular portraits of human breast tumours. *Nature*, 406, 747-752.
- PETRAKIS, N., LEE, R., MIIKE, R., DUPUY, M. & MORRIS, M. 1988. Coloration of breast fluid related to concentration of cholesterol, cholesterol epoxides, estrogen, and lipid peroxides. *American journal of clinical pathology*, 89, 117-120.
- PETRAKIS, N. L. 1986. Physiologic, biochemical, and cytologic aspects of nipple aspirate fluid. *Breast cancer research and treatment*, 8, 7-19.
- PETRAKIS, N. L. 1993. Nipple aspirate fluid in epidemiologic studies of breast disease. *Epidemiologic reviews*, 15, 188-195.
- PETRUCCELLI, N., DALY, M. & FELDMAN, G. 1993. BRCA1 and BRCA2 hereditary breast and ovarian cancer.
- PHYLLIS, R. 1997. HPLC and CE Principles and Practice. Academic Press, London.
- PICHLER, P., KÖCHER, T., HOLZMANN, J., MAZANEK, M., TAUS, T., AMMERER, G. & MECHTLER, K. 2010. Peptide labeling with isobaric tags yields higher identification rates using iTRAQ 4-plex compared to TMT 6-plex and iTRAQ 8-plex on LTQ Orbitrap. *Analytical chemistry*, 82, 6549-6558.
- PICOTTI, P. & AEBERSOLD, R. 2012. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nature methods*, 9, 555-566.
- PICOTTI, P., CLÉMENT-ZIZA, M., LAM, H., CAMPBELL, D. S., SCHMIDT, A., DEUTSCH, E. W., RÖST, H., SUN, Z., RINNER, O. & REITER, L. 2013. A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. *Nature*, 494, 266-270.
- PINHO, S. S. & REIS, C. A. 2015. Glycosylation in cancer: mechanisms and clinical implications. *Nature Reviews Cancer*, 15, 540-555.
- PIPALIYA, N., SOLANKE, D., RATHI, C., PATEL, R., INGLE, M. & SAWANT, P. 2016. Esomeprazole induced galactorrhea: a novel side effect. *Clinical journal of gastroenterology*, 9, 13-16.
- PISANO, E. D., GATSONIS, C., HENDRICK, E., YAFFE, M., BAUM, J. K., ACHARYYA, S., CONANT, E. F., FAJARDO, L. L., BASSETT, L., D'ORSI, C., JONG, R., REBNER, M. & DIGITAL MAMMOGRAPHIC IMAGING SCREENING TRIAL INVESTIGATORS, G. 2005. Diagnostic performance of digital versus film mammography for breast-cancer screening. *N Engl J Med*, 353, 1773-83.
- POZNIAK, Y., BALINT-LAHAT, N., RUDOLPH, J. D., LINDSKOG, C., KATZIR, R., AVIVI, C., PONTÉN, F., RUPPIN, E., BARSHACK, I. & GEIGER, T. 2016. System-wide clinical proteomics of breast cancer reveals global remodeling of tissue homeostasis. *Cell systems*, 2, 172-184.
- PRAT, A., PARKER, J. S., KARGINOVA, O., FAN, C., LIVASY, C., HERSCHKOWITZ, J. I., HE, X. & PEROU, C. M. 2010. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res*, 12, R68.
- PRAT, A. & PEROU, C. M. 2009. Mammary development meets cancer genomics. *Nature medicine*, 15, 842-844.
- PRAT, A. & PEROU, C. M. 2011. Deconstructing the molecular portraits of breast cancer. *Molecular oncology*, 5, 5-23.

- PRUITT, K. D., TATUSOVA, T. & MAGLOTT, D. R. 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic acids research*, 35, D61-D65.
- QIN, W., GUI, G., ZHANG, K., TWELVES, D., KLIETHERMES, B. & SAUTER, E. R. 2012. Proteins and carbohydrates in nipple aspirate fluid predict the presence of atypia and cancer in women requiring diagnostic breast biopsy. *BMC cancer*, 12, 52.
- QIU, Y., PU, T., LI, L., CHENG, F., LU, C., SUN, L., TENG, X., YE, F. & BU, H. 2014. The expression of aldehyde dehydrogenase family in breast cancer. *Journal of breast cancer*, 17, 54-60.
- RAUNIYAR, N. & YATES III, J. R. 2014. Isobaric labeling-based relative quantification in shotgun proteomics. *Journal of proteome research*, 13, 5293-5309.
- RAYNAUD, C., MALLORY, A. C., LATRASSE, D., JÉGU, T., BRUGGEMAN, Q., DELARUE, M., BERGOUNIOUX, C. & BENHAMED, M. 2014. Chromatin meets the cell cycle. *Journal of experimental botany*, ert433.
- REDONDO, C. M., GAGO-DOMÍNGUEZ, M., PONTE, S. M., CASTELO, M. E., JIANG, X., GARCÍA, A. A., FERNANDEZ, M. P., TOMÉ, M. A., FRAGA, M. & GUDE, F. 2012. Breast feeding, parity and breast cancer subtypes in a Spanish cohort. *PloS one*, 7, e40543.
- REIS-FILHO, J. S. & LAKHANI, S. R. 2003. Genetic alterations in pre-invasive lesions. *Breast Cancer Research*, 5, 313-319.
- REMUS, D. 2016. The Role of Mcm2–7 in Replication Initiation. *The Initiation of DNA Replication in Eukaryotes*. Springer.
- RIAL-OTERO, R., CARREIRA, R., CORDEIRO, F., MORO, A., SANTOS, H., VALE, G., MOURA, I. & CAPELO, J. 2007. Ultrasonic assisted protein enzymatic digestion for fast protein identification by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry: sonoreactor versus ultrasonic probe. *Journal of Chromatography A*, 1166, 101-107.
- RIEDER, V., SALAMA, M., GLOCKNER, L., MUHR, D., BERGER, A., TEA, M. K., PFEILER, G., RAPPAPORT-FUERHAUSER, C., GSCHWANTLER-KAULICH, D., WEINGARTSHOFER, S. & SINGER, C. F. 2016. Effect of lifestyle and reproductive factors on the onset of breast cancer in female BRCA 1 and 2 mutation carriers. *Mol Genet Genomic Med*, 4, 172-7.
- RIGHETTI, P. G. 2006. Real and imaginary artefacts in proteome analysis via two-dimensional maps. *Journal of Chromatography B*, 841, 14-22.
- RODENHISER, D. & MANN, M. 2006. Epigenetics and human disease: translating basic biology into clinical applications. *Canadian Medical Association Journal*, 174, 341-348.
- RODRIGUEZ-TORRES, M. & ALLAN, A. L. 2016. Aldehyde dehydrogenase as a marker and functional mediator of metastasis in solid tumors. *Clinical & experimental metastasis*, 33, 97-113.
- RONCADA, P., STIPETIC, L. H., BONIZZI, L., BURCHMORE, R. J. & KENNEDY, M. W. 2013. Proteomics as a tool to explore human milk in health and disease. *J Proteomics*, 88, 47-57.
- RONSEIN, G. E., PAMIR, N., VON HALLER, P. D., KIM, D. S., ODA, M. N., JARVIK, G. P., VAISAR, T. & HEINECKE, J. W. 2015. Parallel reaction monitoring (PRM) and selected reaction monitoring (SRM) exhibit comparable linearity, dynamic range and precision for targeted quantitative HDL proteomics. *Journal of proteomics*, 113, 388-399.
- ROSS, P. L., HUANG, Y. N., MARCHESE, J. N., WILLIAMSON, B., PARKER, K., HATTAN, S., KHAINOVSKI, N., PILLAI, S., DEY, S. & DANIELS, S. 2004. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Molecular & cellular proteomics*, 3, 1154-1169.
- SACHELARIE, I., KERR, K., GHESANI, M. & BLUM, R. H. 2005. Integrated PET-CT: evidence-based review of oncology indications. *Oncology (Williston Park, NY)*, 19, 481-90; discussion 490-2, 495-6.

- SAKURAI, K., ENOMOTO, K., MATSUO, S., AMANO, S. & SHIONO, M. 2011. CYP3A4 expression to predict treatment response to docetaxel for metastasis and recurrence of primary breast cancer. *Surg Today*, 41, 674-9.
- SALHAB, M., KEITH, L. G., LAGUENS, M., REEVES, W. & MOKBEL, K. 2006. The potential role of dynamic thermal analysis in breast cancer detection. *Int Semin Surg Oncol*, 3, 8.
- SARTORIUS, O. 1973. Breast fluid cells help in early cancer detection. *Jama*, 224, 823-827.
- SATELLI, A. & LI, S. 2011. Vimentin in cancer and its potential as a molecular target for cancer therapy. *Cellular and molecular life sciences*, 68, 3033-3046.
- SATO, M., TAKANO, S., SOGAWA, K., NODA, K., YOSHITOMI, H., ISHIBASHI, M., MOGUSHI, K., TAKIZAWA, H., OTSUKA, M. & SHIMIZU, H. 2017. Immune-complex level of cofilin-1 in sera is associated with cancer progression and poor prognosis in pancreatic cancer. *Cancer Science*.
- SAUTER, E., EHYA, H., BABB, J., DIAMANDIS, E., DALY, M., KLEIN-SZANTO, A., SIGURDSON, E., HOFFMAN, J., MALICK, J. & ENGSTROM, P. 1999. Biologic markers of risk in nipple aspirate fluid are associated with residual cancer and tumour size. *British journal of cancer*, 81, 1222.
- SAUTER, E., ROSS, E., DALY, M., KLEIN-SZANTO, A., ENGSTROM, P., SORLING, A., MALICK, J. & EHYA, H. 1997a. Nipple aspirate fluid: a promising non-invasive method to identify cellular markers of breast cancer risk. *British journal of cancer*, 76, 494.
- SAUTER, E. R., DALY, M., LINAHAN, K., EHYA, H., ENGSTROM, P. F., BONNEY, G., ROSS, E. A., YU, H. & DIAMANDIS, E. 1996. Prostate-specific antigen levels in nipple aspirate fluid correlate with breast cancer risk. *Cancer Epidemiology Biomarkers & Prevention*, 5, 967-970.
- SAUTER, E. R., KLEIN, G., WAGNER-MANN, C. & DIAMANDIS, E. P. 2004a. Prostate-specific antigen expression in nipple aspirate fluid is associated with advanced breast cancer. *Cancer detection and prevention*, 28, 27-31.
- SAUTER, E. R., LININGER, J., MAGKLARA, A., HEWETT, J. E. & DIAMANDIS, E. P. 2004b. Association of kallikrein expression in nipple aspirate fluid with breast cancer risk. *Int J Cancer*, 108, 588-91.
- SAUTER, E. R., ROSS, E., DALY, M., KLEIN-SZANTO, A., ENGSTROM, P. F., SORLING, A., MALICK, J. & EHYA, H. 1997b. Nipple aspirate fluid: a promising non-invasive method to identify cellular markers of breast cancer risk. *Br J Cancer*, 76, 494-501.
- SAUTER, E. R., SHAN, S., HEWETT, J. E., SPECKMAN, P. & DU BOIS, G. C. 2005. Proteomic analysis of nipple aspirate fluid using SELDI-TOF-MS. *International journal of cancer*, 114, 791-796.
- SAUTER, E. R., TICHANSKY, D. S., CHERVONEVA, I. & DIAMANDIS, E. P. 2002. Circulating testosterone and prostate-specific antigen in nipple aspirate fluid and tissue are associated with breast cancer. *Environmental health perspectives*, 110, 241.
- SAUTER, E. R., WINN, J. N., DALE, P. S. & WAGNER-MANN, C. 2006. Nipple aspirate fluid color is associated with breast cancer. *Cancer detection and prevention*, 30, 322-328.
- SAVITSKI, M. M., MATHIESON, T., ZINN, N., SWEETMAN, G., DOCE, C., BECHER, I., PACHL, F., KUSTER, B. & BANTSCHOFF, M. 2013. Measuring and managing ratio compression for accurate iTRAQ/TMT quantification. *Journal of proteome research*, 12, 3586-3598.
- SAVITSKI, M. M., SWEETMAN, G., ASKENAZI, M., MARTO, J. A., LANG, M., ZINN, N. & BANTSCHOFF, M. 2011. Delayed fragmentation and optimized isolation width settings for improvement of protein identification and accuracy of isobaric mass tag quantification on Orbitrap-type mass spectrometers. *Analytical chemistry*, 83, 8959-8967.
- SAWYER, E. J., POULSOM, R., HUNT, F. T., JEFFERY, R., ELIA, G., ELLIS, I. O., ELLIS, P., TOMLINSON, I. P. & HANBY, A. M. 2003. Malignant phyllodes tumours show stromal overexpression of c-myc and c-kit. *J Pathol*, 200, 59-64.

- SCHAAB, C., GEIGER, T., STOEHR, G., COX, J. & MANN, M. 2012. Analysis of high accuracy, quantitative proteomics data in the MaxQB database. *Molecular & Cellular Proteomics*, 11, M111. 014068.
- SCHIRRMESTER, H., KUHN, T., GUHLMANN, A., SANTJOHANSER, C., HORSTER, T., NUSSLE, K., KORETZ, K., GLATTING, G., RIEBER, A., KREIENBERG, R., BUCK, A. C. & RESKE, S. N. 2001. Fluorine-18 2-deoxy-2-fluoro-D-glucose PET in the preoperative staging of breast cancer: comparison with the standard staging procedures. *Eur J Nucl Med*, 28, 351-8.
- SCHMIDT, R., BAUMANN, F., KNUPFER, H., BRAUCKHOFF, M., HORN, L. C., SCHONFELDER, M., KOHLER, U. & PREISS, R. 2004. CYP3A4, CYP2C9 and CYP2B6 expression and ifosfamide turnover in breast cancer tissue microsomes. *Br J Cancer*, 90, 911-6.
- SCHNEIDER, S. S., ASLEBAGH, R., NGOUNOU WETIE, A. G., STURGEON, S. R., DARIE, C. C. & ARCARO, K. F. 2014. Using breast milk to assess breast cancer risk: the role of mass spectrometry-based proteomics. *Adv Exp Med Biol*, 806, 399-408.
- SCHOENHERR, R. M., WHITEAKER, J. R., ZHAO, L., IVEY, R. G., TRUTE, M., KENNEDY, J., VOYTOVICH, U. J., YAN, P., LIN, C. & PAULOVICH, A. G. 2012. Multiplexed quantification of estrogen receptor and HER2/Neu in tissue and cell lysates by peptide immunoaffinity enrichment mass spectrometry. *Proteomics*, 12, 1253-1260.
- SCHUERENBERG, M., LUEBBERT, C., EICKHOFF, H., KALKUM, M., LEHRACH, H. & NORDHOFF, E. 2000. Prestructured MALDI-MS sample supports. *Analytical chemistry*, 72, 3436-3442.
- SCHÜRENBERG, M. & FRANZEN, J. 2001. Sample support plates for Maldi mass spectrometry including methods for manufacture of plates and application of sample. Google Patents.
- SCHWANHÄUSSER, B., BUSSE, D., LI, N., DITTMAR, G., SCHUCHHARDT, J., WOLF, J., CHEN, W. & SELBACH, M. 2011. Global quantification of mammalian gene expression control. *Nature*, 473, 337.
- SCIEX, A. 2017. *iTRAQ® Reagents* [Online]. AB Sciex. Available: <https://sciex.com/products/standards-and-reagents/itraq-reagent> [Accessed 02/02 2016].
- SEAL, M., SPEERS, C., O'REILLY, S., GELMON, K., ELLARD, S. & CHIA, S. 2012. Outcomes of women with early-stage breast cancer receiving adjuvant trastuzumab. *Current Oncology*, 19, 197.
- SEGUIN, L., DESGROSELLIER, J. S., WEIS, S. M. & CHERESH, D. A. 2015. Integrins and cancer: regulators of cancer stemness, metastasis, and drug resistance. *Trends in cell biology*, 25, 234-240.
- SEITZ, H. K. & STICKEL, F. 2010. Acetaldehyde as an underestimated risk factor for cancer development: role of genetics in ethanol metabolism. *Genes & nutrition*, 5, 121-128.
- SENKO, M. W. 2015. Multinotch isolation for ms3 mass analysis. Google Patents.
- SENKO, M. W., REMES, P. M., CANTERBURY, J. D., MATHUR, R., SONG, Q., ELIUK, S. M., MULLEN, C., EARLEY, L., HARDMAN, M. & BLETHROW, J. D. 2013. Novel parallelized quadrupole/linear ion trap/Orbitrap tribrid mass spectrometer improving proteome coverage and peptide identification rates. *Analytical chemistry*, 85, 11710-11714.
- SHAHEED, S.-U., HADJISAVVAS, A., SOKRATOUS, K., LOADMAN, P., SUTTON, C. & KYRIACOU, K. 2015. Analysis of HSP10 as a putative biomarker of breast cancer. *Cancer Research*, 75, 1576-1576.
- SHAHEED, S.-U., RUSTOGI, N., SCALLY, A., WILSON, J., THYGESEN, H., LOIZIDOU, M. A., HADJISAVVAS, A., HANBY, A., SPEIRS, V. & LOADMAN, P. 2013a. Identification of stage-specific breast markers using quantitative proteomics. *Journal of proteome research*, 12, 5696-5708.
- SHAHEED, S. U., RUSTOGI, N., SCALLY, A., WILSON, J., THYGESEN, H., LOIZIDOU, M. A., HADJISAVVAS, A., HANBY, A., SPEIRS, V., LOADMAN, P., LINFORTH, R., KYRIACOU, K. & SUTTON, C. W. 2013b. Identification of stage-specific breast markers using quantitative proteomics. *J Proteome Res*, 12, 5696-708.

- SHAO, M.-M., CHAN, S. K., ALEX, M., LAM, C. C., TSANG, J. Y., LUI, P. C., LAW, B. K., TAN, P.-H. & GARY, M. T. 2012. Keratin expression in breast cancers. *Virchows Archiv*, 461, 313-322.
- SHEVCHENKO, A., WILM, M., VORM, O. & MANN, M. 1996. Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Analytical chemistry*, 68, 850-858.
- SHI, F., SHANG, L., PAN, B.-Q., WANG, X.-M., JIANG, Y.-Y., HAO, J.-J., ZHANG, Y., CAI, Y., XU, X. & ZHAN, Q.-M. 2014. Calreticulin Promotes Migration and Invasion of Esophageal Cancer Cells by Upregulating Neuropilin-1 Expression via STAT5A. *Clinical Cancer Research*, 20, 6153-6162.
- SHIN, B. K., WANG, H., YIM, A. M., LE NAOUR, F., BRICHORY, F., JANG, J. H., ZHAO, R., PURAVS, E., TRA, J. & MICHAEL, C. W. 2003. Global profiling of the cell surface proteome of cancer cells uncovers an abundance of proteins with chaperone function. *Journal of Biological Chemistry*, 278, 7607-7616.
- SIMPSON, N. E., GERTZ, J., IMBERG, K., MYERS, R. M. & GARABEDIAN, M. J. 2011. Research resource: enhanced genome-wide occupancy of estrogen receptor α by the cochaperone p23 in breast cancer cells. *Molecular Endocrinology*, 26, 194-202.
- SINGLETERY, S. E., ALLRED, C., ASHLEY, P., BASSETT, L. W., BERRY, D., BLAND, K. I., BORGES, P. I., CLARK, G., EDGE, S. B. & HAYES, D. F. 2002. Revision of the American Joint Committee on Cancer staging system for breast cancer. *Journal of clinical oncology*, 20, 3628-3636.
- SINHA, R., SINGH, R., MEHROTRA, S. & SINGH, R. 2009. Implications of free radicals and antioxidant levels in carcinoma of the breast: a never-ending battle for survival. *Indian journal of cancer*, 46.
- SJOBERG, J. 2014. 62ND ASMS Conference on Mass Spectrometry and Allied Topics. *Journal of The American Society for Mass Spectrometry*, 25, 1-263.
- SLOAN, E. K., STANLEY, K. L. & ANDERSON, R. L. 2004. Caveolin-1 inhibits breast cancer growth and metastasis. *Oncogene*, 23, 7893-7897.
- SMITH, R. A., DUFFY, S. W. & TABÁR, L. 2012. Breast cancer screening: the evolving evidence. *Oncology*, 26, 471-486.
- SMITH, S. M. & CAI, L. 2012. Cell specific CD44 expression in breast cancer requires the interaction of AP-1 and NF κ B with a novel cis-element. *PLoS one*, 7, e50867.
- SNYDER, L. R., KIRKLAND, J. J. & DOLAN, J. W. 2011. *Introduction to modern liquid chromatography*, John Wiley & Sons.
- SOBIN, L. H. & FLEMING, I. D. 1997. TNM classification of malignant tumors, (1997). *Cancer*, 80, 1803-1804.
- SOMASIRI, A., NIELSEN, J. S., MAKRETISOV, N., MCCOY, M. L., PRENTICE, L., GILKS, C. B., CHIA, S. K., GELMON, K. A., KERSHAW, D. B. & HUNTSMAN, D. G. 2004. Overexpression of the anti-adhesin podocalyxin is an independent predictor of breast cancer progression. *Cancer research*, 64, 5068-5073.
- SØRLIE, T., PEROU, C. M., TIBSHIRANI, R., AAS, T., GEISLER, S., JOHNSEN, H., HASTIE, T., EISEN, M. B., VAN DE RIJN, M. & JEFFREY, S. S. 2001. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences*, 98, 10869-10874.
- SPECHT, J. M. & MANKOFF, D. A. 2012. Advances in molecular imaging for breast cancer detection and characterization. *Breast Cancer Res*, 14, 206.
- STAHL-ZENG, J., LANGE, V., OSSOLA, R., ECKHARDT, K., KREK, W., AEBERSOLD, R. & DOMON, B. 2007. High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites. *Molecular & Cellular Proteomics*, 6, 1809-1817.
- STEWART, A., NIXON, D., ZAMCHECK, N. & AISENBERG, A. 1974. Carcinoembryonic antigen in breast cancer patients: serum levels and disease progress. *Cancer*, 33, 1246-1252.
- STOLLE, C. A. 2016. von Hippel-Lindau Disease. *Molecular Pathology in Clinical Practice*. Springer.

- STOUT, N. K., LEE, S. J., SCHECHTER, C. B., KERLIKOWSKA, K., ALAGOZ, O., BERRY, D., BUIST, D. S., CEVIK, M., CHISHOLM, G., DE KONING, H. J., HUANG, H., HUBBARD, R. A., MIGLIORETTI, D. L., MUNSELL, M. F., TRENTHAM-DIETZ, A., VAN RAVESTEYN, N. T., TOSTESON, A. N. & MANDELBLATT, J. S. 2014. Benefits, harms, and costs for breast cancer screening after US implementation of digital mammography. *J Natl Cancer Inst*, 106, dju092.
- SUCKAU, D., RESEMANN, A., SCHUERENBERG, M., HUFNAGEL, P., FRANZEN, J. & HOLLE, A. 2003. A novel MALDI LIFT-TOF/TOF mass spectrometer for proteomics. *Analytical and bioanalytical chemistry*, 376, 952-965.
- SUIJKERBUIJK, K. P., VAN DER WALL, E., MEIJRINK, H., PAN, X., BOREL RINKES, I. H., AUSEMS, M. G. & VAN DIEST, P. J. 2010. Successful oxytocin-assisted nipple aspiration in women at increased risk for breast cancer. *Fam Cancer*, 9, 321-5.
- SULLIVAN, W., STENSGARD, B., CAUCUTT, G., BARTHA, B., MCMAHON, N., ALNEMRI, E. S., LITWACK, G. & TOFT, D. 1997. Nucleotides and two functional states of hsp90. *Journal of Biological Chemistry*, 272, 8007-8012.
- SUTTON, C. W., RUSTOGI, N., GURKAN, C., SCALLY, A., LOIZIDOU, M. A., HADJISAVVAS, A. & KYRIACOU, K. 2010. Quantitative proteomic profiling of matched normal and tumor breast tissues. *Journal of proteome research*, 9, 3891-3902.
- TANIOKA, T., NAKATANI, Y., KOBAYASHI, T., TSUJIMOTO, M., OH-ISHI, S., MURAKAMI, M. & KUDO, I. 2003. Regulation of cytosolic prostaglandin E 2 synthase by 90-kDa heat shock protein. *Biochemical and biophysical research communications*, 303, 1018-1023.
- TELLI, M. L. 2016. Triple-Negative Breast Cancer. *Molecular Pathology of Breast Cancer*. Springer.
- THERMOFISHER, T. 2016. *TMTsixplex™ Isobaric Label Reagent Set*, [Online]. Available: <https://www.thermofisher.com/order/catalog/product/90061> [Accessed 02 2016].
- THOMPSON, A., SCHÄFER, J., KUHN, K., KIENLE, S., SCHWARZ, J., SCHMIDT, G., NEUMANN, T. & HAMON, C. 2003. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Analytical chemistry*, 75, 1895-1904.
- THOMPSON, C. C., ASHCROFT, F. J., PATEL, S., SARAGA, G., VIMALACHANDRAN, D., PRIME, W., CAMPBELL, F., DODSON, A., JENKINS, R. E. & LEMOINE, N. R. 2007. Pancreatic cancer cells overexpress gelsolin family-capping proteins, which contribute to their cell motility. *Gut*, 56, 95-106.
- THOMPSON, E. W., PAIK, S., BRÜNNER, N., SOMMERS, C. L., ZUGMAIER, G., CLARKE, R., SHIMA, T. B., TORRI, J., DONAHUE, S. & LIPPMAN, M. E. 1992. Association of increased basement membrane invasiveness with absence of estrogen receptor and expression of vimentin in human breast cancer cell lines. *Journal of cellular physiology*, 150, 534-544.
- TIFFANY, J. 2003. Tears in health and disease. *Eye*, 17, 923-926.
- TING, L., RAD, R., GYGI, S. P. & HAAS, W. 2011. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nature methods*, 8, 937-940.
- TOOGUN, O. A., DEZWAAN, D. C. & FREEMAN, B. C. 2008. The hsp90 molecular chaperone modulates multiple telomerase activities. *Molecular and cellular biology*, 28, 457-467.
- TROTTER, M. W., SADOWSKI, P. G., DUNKLEY, T. P., GROEN, A. J. & LILLEY, K. S. 2010. Improved sub-cellular resolution via simultaneous analysis of organelle proteomics data across varied experimental conditions. *Proteomics*, 10, 4213-4219.
- TSANG, J. Y., MENDOZA, P., LAM, C. C., YU, A. M., PUTTI, T. C., KARIM, R. Z., SCOLYER, R. A., LEE, C. S., TAN, P. H. & TSE, G. M. 2012. Involvement of alpha- and beta-catenins and E-cadherin in the development of mammary phyllodes tumours. *Histopathology*, 61, 667-74.
- TSE, G. M., LUI, P. C., LEE, C. S., KUNG, F. Y., SCOLYER, R. A., LAW, B. K., LAU, T. S., KARIM, R. & PUTTI, T. C. 2004. Stromal expression of vascular endothelial growth factor correlates with tumor grade and microvessel density in mammary phyllodes tumors: a multicenter study of 185 cases. *Hum Pathol*, 35, 1053-7.

- TSE, G. M., LUI, P. C., VONG, J. S., LAU, K. M., PUTTI, T. C., KARIM, R., SCOLYER, R. A., LEE, C. S., YU, A. M., NG, D. C., TSE, A. K. & TAN, P. H. 2009. Increased epidermal growth factor receptor (EGFR) expression in malignant mammary phyllodes tumors. *Breast Cancer Res Treat*, 114, 441-8.
- TYANOVA, S., ALBRECHTSEN, R., KRONQVIST, P., COX, J., MANN, M. & GEIGER, T. 2016. Proteomic maps of breast cancer subtypes. *Nature communications*, 7.
- UNIPROT.ORG. 2015. *The Universal Protein Resource (UniProt)* [Online]. Uniprot.org. Available: <http://www.uniprot.org/> [Accessed 05/03 2015].
- VACHON, C. M., BRANDT, K. R., GHOSH, K., SCOTT, C. G., MALONEY, S. D., CARSTON, M. J., PANKRATZ, V. S. & SELLERS, T. A. 2007. Mammographic breast density as a general marker of breast cancer risk. *Cancer Epidemiol Biomarkers Prev*, 16, 43-9.
- VAN DIJK, E. L., AUGER, H., JASZCZYSZYN, Y. & THERMES, C. 2014. Ten years of next-generation sequencing technology. *Trends in genetics*, 30, 418-426.
- VAN HOOFF, D., PINKSE, M. W., WARD-VAN OOSTWAARD, D., MUMMERY, C. L., HECK, A. J. & KRIJGSVELD, J. 2007. An experimental correction for arginine-to-proline conversion artifacts in SILAC-based quantitative proteomics. *Nature methods*, 4, 677-678.
- VARNUM, S. M., COVINGTON, C. C., WOODBURY, R. L., PETRITIS, K., KANGAS, L. J., ABDULLAH, M. S., POUNDS, J. G., SMITH, R. D. & ZANGAR, R. C. 2003. Proteomic characterization of nipple aspirate fluid: identification of potential biomarkers of breast cancer. *Breast cancer research and treatment*, 80, 87-97.
- VIKESAA, J., HANSEN, T. V., JØNSEN, L., BORUP, R., WEWER, U. M., CHRISTIANSEN, J. & NIELSEN, F. C. 2006. RNA-binding IMPs promote cell adhesion and invadopodia formation. *The EMBO journal*, 25, 1456-1468.
- VILELA, M. H., DE ALMEIDA, F. M., DE PAULA, G. M., RIBEIRO, N. B., CIRQUEIRA, M. B., SILVA, A. L. & MOREIRA, M. A. 2014. Utility of Ki-67, CD10, CD34, p53, CD117, and mast cell content in the differential diagnosis of cellular fibroadenomas and in the classification of phyllodes tumors of the breast. *Int J Surg Pathol*, 22, 485-91.
- VILLA FLOR BRUNORO, G., TEIXEIRA DA SILVA FERREIRA, A., RAMOS DE OLIVEIRA TRUGILHO, M., SOUSA DE OLIVEIRA, T., CLAUDIO BELO AMENDOLA, L., PERALES, J., HEMMI VALENTE, R., VITORIA DE MOURA GALLO, C., PAGNONCELLI, D. & GISELE DA COSTA NEVES-FERREIRA, A. 2014. Potential Correlation between Tumor Aggressiveness and Protein Expression Patterns of Nipple Aspirate Fluid (NAF) Revealed by Gel-Based Proteomic Analysis. *Current topics in medicinal chemistry*, 14, 359-368.
- VIZCAÍNO, J. A., CÔTÉ, R. G., CSORDAS, A., DIANES, J. A., FABREGAT, A., FOSTER, J. M., GRISS, J., ALPI, E., BIRIM, M. & CONTELL, J. 2013. The PRoteomics IDentifications (PRIDE) database and associated tools: status in 2013. *Nucleic acids research*, 41, D1063-D1069.
- VIZCAÍNO, J. A., CSORDAS, A., DEL-TORO, N., DIANES, J. A., GRISS, J., LAVIDAS, I., MAYER, G., PEREZ-RIVEROL, Y., REISINGER, F. & TERNENT, T. 2016. 2016 update of the PRIDE database and its related tools. *Nucleic acids research*, 44, D447-D456.
- WALTHER, T. C. & MANN, M. 2010. Mass spectrometry-based proteomics in cell biology. *The Journal of cell biology*, 190, 491-500.
- WANG, S., HUANG, X., SUN, D., XIN, X., PAN, Q., PENG, S., LIANG, Z., LUO, C., YANG, Y. & JIANG, H. 2012. Extensive crosstalk between O-GlcNAcylation and phosphorylation regulates Akt signaling. *PLoS one*, 7, e37427.
- WANG, W., EDDY, R. & CONDEELIS, J. 2007. The cofilin pathway in breast cancer invasion and metastasis. *Nature Reviews Cancer*, 7, 429-440.
- WANG, Z., HAO, Y. & LOWE, A. W. 2008. The adenocarcinoma-associated antigen, AGR2, promotes tumor growth, cell migration, and cellular transformation. *Cancer research*, 68, 492-497.
- WARNER, E., MESSERSMITH, H., CAUSER, P., EISEN, A., SHUMAK, R. & PLEWES, D. 2008. Systematic review: using magnetic resonance imaging to screen women at high risk for breast cancer. *Ann Intern Med*, 148, 671-9.

- WASHBURN, M. P., WOLTERS, D. & YATES, J. R. 2001. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature biotechnology*, 19, 242-247.
- WASINGER, V. C., ZENG, M. & YAU, Y. 2013. Current status and advances in quantitative proteomic mass spectrometry. *International journal of proteomics*, 2013.
- WATERS CORPORATION. 2017a. *Products: Mass spectrometry* [Online]. Available: http://www.waters.com/waters/home.htm?locale=en_US [Accessed 02/02 2017].
- WATERS CORPORATION. 2017b. *TargetLynx: Courses* [Online]. Available: http://www.waters.com/waters/en_US/TargetLynx-/nav.htm?cid=513791&locale=en_US [Accessed 04/02 2017].
- WEERAPANA, E., WANG, C., SIMON, G. M., RICHTER, F., KHARE, S., DILLON, M. B., BACHOVCHIN, D. A., MOWEN, K., BAKER, D. & CRAVATT, B. F. 2010. Quantitative reactivity profiling predicts functional cysteines in proteomes. *Nature*, 468, 790-795.
- WEIGELT, B., HORLINGS, H., KREIKE, B., HAYES, M., HAUPTMANN, M., WESSELS, L., DE JONG, D., VAN DE VIJVER, M., VAN'T VEER, L. & PETERSE, J. 2008. Refinement of breast cancer classification by molecular characterization of histological special types. *The Journal of pathology*, 216, 141-150.
- WEIGELT, B., KREIKE, B. & REIS-FILHO, J. S. 2009. Metaplastic breast carcinomas are basal-like breast cancers: a genomic profiling analysis. *Breast cancer research and treatment*, 117, 273-280.
- WEITZEL, J. N., BLAZER, K. R., MACDONALD, D. J., CULVER, J. O. & OFFIT, K. 2011. Genetics, genomics, and cancer risk assessment. *CA: a cancer journal for clinicians*, 61, 327-359.
- WELLS, J. M. & MCLUCKEY, S. A. 2005. Collision-induced dissociation (CID) of peptides and proteins. *Methods in enzymology*, 402, 148-185.
- WENGER, C. D. & COON, J. J. 2013. A proteomics search algorithm specifically designed for high-resolution tandem mass spectra. *Journal of proteome research*, 12, 1377-1386.
- WHITE, C. D., LI, Z. & SACKS, D. B. 2011. Calmodulin binds HER2 and modulates HER2 signaling. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1813, 1074-1082.
- WHITE, F. M. 2011. The potential cost of high-throughput proteomics. *Sci. Signal.*, 4, pe8-pe8.
- WHITEAKER, J. R., ZHANG, H., ZHAO, L., WANG, P., KELLY-SPRATT, K. S., IVEY, R. G., PIENING, B. D., FENG, L.-C., KASARDA, E. & GURLEY, K. E. 2007. Integrated pipeline for mass spectrometry-based discovery and confirmation of biomarkers demonstrated in a mouse model of breast cancer. *Journal of proteome research*, 6, 3962-3975.
- WIESE, S., REIDEGELD, K. A., MEYER, H. E. & WARSCHIED, B. 2007. Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. *Proteomics*, 7, 340-350.
- WILDING, J. L. & BODMER, W. F. 2014. Cancer cell lines for drug discovery and development. *Cancer research*, 74, 2377-2384.
- WILLIAMSON, J. C., EDWARDS, A. V., VERANO-BRAGA, T., SCHWÄMMLE, V., KJELDEN, F., JENSEN, O. N. & LARSEN, M. R. 2016. High-performance hybrid Orbitrap mass spectrometers for quantitative proteome analysis: Observations and implications. *Proteomics*.
- WISNIEWSKI, J. R., ZOUGMAN, A., NAGARAJ, N. & MANN, M. 2009. Universal sample preparation method for proteome analysis. *Nature methods*, 6, 359.
- WOLF-YADLIN, A., HAUTANIEMI, S., LAUFFENBURGER, D. A. & WHITE, F. M. 2007. Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proceedings of the National Academy of Sciences*, 104, 5860-5865.
- WOLTERS, D. A., WASHBURN, M. P. & YATES, J. R. 2001. An automated multidimensional protein identification technology for shotgun proteomics. *Analytical chemistry*, 73, 5683-5690.
- YAFFE, M. J. 2008. Mammographic density. Measurement of mammographic density. *Breast Cancer Res*, 10, 209.

- YAMAGUCHI, R., TANAKA, M., TSE, G. M., YAMAGUCHI, M., TERASAKI, H., HIRAI, Y., NONAKA, Y., MORITA, M., YOKOYAMA, T. & KANOMATA, N. 2015. Management of breast papillary lesions diagnosed in ultrasound-guided vacuum-assisted and core needle biopsies. *Histopathology*.
- YANG, H.-S., MATTHEWS, C. P., CLAIR, T., WANG, Q., BAKER, A. R., LI, C.-C. H., TAN, T.-H. & COLBURN, N. H. 2006. Tumorigenesis suppressor Pcd4 down-regulates mitogen-activated protein kinase kinase kinase 1 expression to suppress colon carcinoma cell invasion. *Molecular and cellular biology*, 26, 1297-1306.
- YANG, H., YE, D., GUAN, K.-L. & XIONG, Y. 2012. IDH1 and IDH2 mutations in tumorigenesis: mechanistic insights and clinical perspectives. *Clinical Cancer Research*, 18, 5562-5571.
- YOSHINARI, K., OSE, Y. & KATO, Y. 2000. Ion trap mass spectrometer. Google Patents.
- ZALLES, C., KIMLER, B. F., KAMEL, S., MCKITTRICK, R. & FABIAN, C. J. 1995. Cytology patterns in random aspirates from women at high and low risk for breast cancer. *The Breast Journal*, 1, 343-349.
- ZEIDAN, B. A., TOWNSEND, P. A., GARBIS, S. D., COPSON, E. & CUTRESS, R. I. 2015. Clinical proteomics and breast cancer. *Surgeon*, 13, 271-8.
- ZHANG, A., SUN, H., WANG, P. & WANG, X. 2013. Salivary proteomics in biomedical research. *Clinica Chimica Acta*, 415, 261-265.
- ZHANG, B., WANG, J., WANG, X., ZHU, J., LIU, Q., SHI, Z., CHAMBERS, M. C., ZIMMERMAN, L. J., SHADDOX, K. F. & KIM, S. 2014a. Proteogenomic characterization of human colon and rectal cancer. *Nature*, 513, 382.
- ZHANG, G., ANNAN, R. S., CARR, S. A. & NEUBERT, T. A. 2014b. Overview of peptide and protein analysis by mass spectrometry. *Current Protocols in Molecular Biology*, 10.21. 1-10.21. 30.
- ZHANG, H., LIU, Q., ZIMMERMAN, L. J., HAM, A.-J. L., SLEBOS, R. J., RAHMAN, J., KIKUCHI, T., MASSION, P. P., CARBONE, D. P. & BILLHEIMER, D. 2011. Methods for peptide and protein quantitation by liquid chromatography-multiple reaction monitoring mass spectrometry. *Molecular & Cellular Proteomics*, 10, M110. 006593.
- ZHANG, L., SHAO, Z. M., BEATTY, P., SARTIPPOUR, M., WANG, H. J., ELASHOFF, R., CHANG, H. & BROOKS, M. N. 2003. The use of oxytocin in nipple fluid aspiration. *The breast journal*, 9, 266-268.
- ZHANG, Y., NEWCOMB, P. A., EGAN, K. M., TITUS-ERNSTOFF, L., CHANOCK, S., WELCH, R., BRINTON, L. A., LISSOWSKA, J., BARDIN-MIKOLAJCZAK, A. & PEPLONSKA, B. 2006. Genetic polymorphisms in base-excision repair pathway genes and risk of breast cancer. *Cancer Epidemiology Biomarkers & Prevention*, 15, 353-358.
- ZHANG, Y., SIEUWERTS, A. M., MCGREEVY, M., CASEY, G., CUFER, T., PARADISO, A., HARBECK, N., SPAN, P. N., HICKS, D. G., CROWE, J., TUBBS, R. R., BUDD, G. T., LYONS, J., SWEEP, F. C., SCHMITT, M., SCHITTULLI, F., GOLOUH, R., TALANTOV, D., WANG, Y. & FOEKENS, J. A. 2009. The 76-gene signature defines high-risk patients that benefit from adjuvant tamoxifen therapy. *Breast Cancer Res Treat*, 116, 303-9.
- ZHANG, Y., WOLF-YADLIN, A., ROSS, P. L., PAPPIN, D. J., RUSH, J., LAUFFENBURGER, D. A. & WHITE, F. M. 2005. Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Molecular & Cellular Proteomics*, 4, 1240-1250.
- ZHAO, H., LANGERØD, A., JI, Y., NOWELS, K. W., NESLAND, J. M., TIBSHIRANI, R., BUKHOLM, I. K., KÅRESEN, R., BOTSTEIN, D. & BØRRESEN-DALE, A.-L. 2004. Different gene expression patterns in invasive lobular and ductal carcinomas of the breast. *Molecular biology of the cell*, 15, 2523-2536.
- ZHAO, Y.-S., PANG, D., WANG, F., XUE, Y.-W., GAO, D.-N., LI, H., LI, K., WANG, B.-Y., WANG, D. & LI, H.-Y. 2009a. Nipple aspirate fluid collection, related factors and relationship between carcinoembryonic antigen in nipple aspirate fluid and breast diseases in women in Harbin, PRC. *Cancer Epidemiology Biomarkers & Prevention*, 18, 732-738.

- ZHAO, Y., LEE, W.-N. P. & XIAO, G. G. 2009b. Quantitative proteomics and biomarker discovery in human cancer. *Expert review of proteomics*, 6, 115-118.
- ZHENG, Y., FANG, Y., LI, S. & ZHENG, B. 2013. Detection of plasma cofilin protein for diagnosis of lung cancer. *Nan fang yi ke da xue xue bao= Journal of Southern Medical University*, 33, 1551-1553.
- ZHONG, J., KRAWCZYK, S. A., CHAERKADY, R., HUANG, H., GOEL, R., BADER, J. S., WONG, G. W., CORKEY, B. E. & PANDEY, A. 2010. Temporal profiling of the secretome during adipogenesis in humans. *Journal of proteome research*, 9, 5228-5238.
- ZHOU, L., BEUERMAN, R. W., CHAN, C. M., ZHAO, S. Z., LI, X. R., YANG, H., TONG, L., LIU, S., STERN, M. E. & TAN, D. 2009. Identification of tear fluid biomarkers in dry eye syndrome using iTRAQ quantitative proteomics. *Journal of proteome research*, 8, 4889-4905.
- ZHUROV, K. O., FORNELLI, L., WODRICH, M. D., LASKAY, Ü. A. & TSYBIN, Y. O. 2013. Principles of electron capture and transfer dissociation mass spectrometry applied to peptide and protein structure analysis. *Chemical Society Reviews*, 42, 5014-5030.
- ZIESKE, L. R. 2006. A perspective on the use of iTRAQ™ reagent technology for protein complex and profiling studies. *Journal of experimental botany*, 57, 1501-1508.
- ZUBAREV, R. A. & MAKAROV, A. 2013. Orbitrap mass spectrometry. ACS Publications.