

The Influence of BRCA1's Ubiquitin Ligase Activity on Cell Motility



Sameer Sengupta

Keble College

Sir William Dunn School of Pathology

University of Oxford

Hilary Term 2014

A thesis submitted for the degree of Doctor of Philosophy

Abstract

Breast cancer type 1 susceptibility protein (BRCA1) has been established as an important tumour suppressor protein and its loss of function is associated with hereditary breast and ovarian cancer. An emerging body of work suggests that BRCA1 is involved in sporadic cases of breast and ovarian cancers and may also have a role in other cancers, indicating a more global role in tumourigenesis. BRCA1-mutated cancers can be early-onset and are characterised by being highly aggressive with a propensity to metastasize, thus from a clinical perspective there is a requirement to understand the molecular mechanisms in order to be able to tailor treatments and develop therapeutics.

BRCA1 has numerous cellular functions, many ascribed to its role in maintenance of genome integrity, transcription and checkpoint control. More recently, a number of extra-nuclear roles have been established. An interesting novel function is the role of the E3 ubiquitin ligase activity on cell motility. Abrogation of the ubiquitin ligase activity of BRCA1 results in cells exhibiting a hypermotile, invasive phenotype which may help to account for the metastatic nature of BRCA1-mutated tumours. Our aim was to further elucidate BRCA1's role in cell motility, starting with the identification of relevant candidate ubiquitin ligase substrates. To date, there has yet to be a systematic approach to identify BRCA1's ubiquitin ligase substrates. Thus we undertook an unbiased proteomic approach to identify extra-nuclear candidates by comparing the profiles of ubiquitinated proteins in breast cancer epithelial cells expressing either functional BRCA1 or ubiquitin ligase-dead BRCA1. We identified 55 candidates which were differentially enriched between the two cell lines and through pathway analysis we determined a significant proportion were cytoskeletal and translation related proteins. Using an ubiquitin-remnant profiling approach, we also identified the site(s) of ubiquitination for many of the candidates.

To assess the role of these candidates in cell motility initially we adopted an *in silico* approach. We used existing time-lapse movies from the online database (www.mitocheck.org) which systematically siRNA knocked down every single gene in the human genome. We developed a series of algorithms which track cell motility from these movies and used these to analyse 192,000 movies containing 3.5 billion cell steps. We have produced a complete database containing motility information after siRNA knockdown of every gene in the human genome, which has been annotated with gene ontologies, KEGG families, Gene Descriptions, SwissProt, Ensembl IDs and siRNA information. In addition to providing motility data of our candidates, we also carried out gene set enrichment analysis on the whole dataset to uncover structural or functional families that may be involved in up-regulating motility when knocked down by siRNA. This is the first report of a genome-wide motility database.

Based on overlaps between the results from these two large-scale unbiased proteomic and *in silico* datasets, we selected 4 candidates, namely, ezrin, moesin, fermitin-2 and delta-catenin. Through monolayer wound healing, cell spreading and single cell motility assays, we determined that ezrin was a particularly relevant and informative candidate. The hypermotile phenotype observed in cells expressing ubiquitin ligase dead BRCA1 was rescued through siRNA knockdown of ezrin and thus we suggest that BRCA1 may regulate cell motility through effects on ezrin.

This thesis has investigated candidate BRCA1's role in cell motility, identified candidate substrates for the E3 ubiquitin ligase activity, established a genome-wide motility database and proposed a possible pathway through which BRCA1 may mediate cell motility and by extension metastasis.

Dedicated to my late aunt, Sharmila Das.

An incredible woman with an inspirational thirst for knowledge.

Acknowledgements

Many people have contributed to the successful completion of this thesis and it is my great pleasure to be able to thank them here. Firstly, I would like to express my deep gratitude to my supervisor, David Vaux, for his continuous support and guidance throughout the project. His unfaltering enthusiasm and energy for science is so infectious and has served as an inspiration. David's generosity with his time, ideas and advice was invaluable to the project and my development as a scientist.

I would also like to extend my thanks to all the members of the Vaux Lab, for the helpful discussions, advice and great company. Thanks to Letitia Jean, Ashraf Malhas, and Chongsoo Lee for helping me navigate through the complicated experiments and providing indispensable advice with the thesis. Thanks to Aileen Maloney, Marek Drozd, Helder Carmen, Chris Goulbourne and Anja Nussler for making the lab such an enjoyable place to work. It was also a pleasure to supervise Alex West. His hard work and good humour helped to provide results for the final chapter, for which I am very grateful.

Further thanks must go to my departmental and college advisors Bass Hassan and Simon Hunt for their academic support throughout the four years.

I am indebted to my collaborators, Boguslaw Obara and Simon McGowan who helped with the bioinformatics. A huge thanks to Benjamin Thomas, Svenja Hester, David Trudgian and Gabriela Ridlova for their help with planning and executing proteomics experiments. I would also like to extend my appreciation to the many members of the Dunn School who helped with my project over the years. In particular, I am grateful to Ed Hutchinson for his help with the ubiquitination experiments and for reading my thesis, Alan Wainman for accommodating and solving my microscopy difficulties, and Fernando Martinez Estrada for his enthusiasm and computational skills with my mass spectrometry data.

I thank the Medical Research Council for funding my DPhil and the Keble Association for their financial support.

I could not have completed my thesis without a little help from my friends. I would like to thank Daniel Guinness, whose camaraderie and (un-)coordinated dance moves kept me sane during the write-up period. I am grateful for my supportive housemates over the years, Chris Gamble, AJ Connor, George Abraham and Tom Dunton, who have been a pleasure to live with. I am especially indebted to Dan, AJ and Tom for their help at the final adrenaline-fuelled, sleep-deprived stages and for getting me to the Examination Schools for the 'photo-finish'. Thanks to Tania Nguyen and Jayati Jain for their cutting wit, friendship and proof-reading. To my dear friends Foteini Dimirouli and Anbara Khalidi, who have always been just a coffee away from calming me down, putting things in perspective and making me laugh. Thanks to all my other close friends from Keble MCR, St Aidan's College, Core Four and Norwich School who are too many to mention but have all been so supportive of my academic endeavours.

My heartfelt appreciation to Maria Lacayo, whom I cannot thank enough. Her support, encouragement and belief in me kept me going through the tough times and made the good times so memorable. I am extremely grateful for our time together in Oxford and beyond.

My deepest thanks go to my family – to all my aunts, uncles, cousins and grandparents whose support across the years has been immeasurable. A special thanks to Nikhil and Arnov, the most amazing brothers anyone could ask for, and to whom the confines of this page cannot do justice. Finally, I would like to thank my parents, Sudeshna and Samit, who have always given me the love, freedom and encouragement to follow to my passions and the strength to see them through. To them, I owe everything and also dedicate this thesis.

Table of Contents

List of Abbreviations.....	9
List of Figures	11
List of Tables.....	12
Chapter 1: Introduction.....	13
1.1 Introduction	13
1.2 BRCA1 as a Tumour Suppressor	14
1.2.1 BRCA1 association with breast and ovarian cancer.....	14
1.2.2 BRCA1 association with basal-like breast cancer.....	15
1.2.3 BRCA1's links with other cancers	16
1.3 The Ubiquitination Pathway.....	18
1.4 Structure of BRCA1	21
1.4.1 The RING domain	22
1.4.2 The BRCT repeats.....	26
1.5 Localisation and Function of BRCA1	27
1.5.1 Shuttling of BRCA1	28
1.5.2 BRCA1 and the nucleus	29
1.5.3 BRCA1 and the mitochondria	31
1.5.4 BRCA1 and the centrosome	32
1.5.5 BRCA1 links from the nucleus to the cytosol.....	33
1.5.6 BRCA1, the cell membrane, and its role in cell motility	35
1.6 Aims	36
Chapter 2: Materials and Methods.....	40
2.1 Table of Antibodies.....	40
2.2 Mammalian Cell Culture.....	40
2.2.1 General Cell Culture.....	40
2.2.2 Culture of individual cell lines	40
2.2.3 Passaging of cells	41
2.2.4 Drug treatment of cells in culture	41
2.2.5 Transfection of mammalian cells in culture	41
2.3 Protein Extraction, Purification and Analysis	42
2.3.1 One dimensional SDS polyacrylamide gel electrophoresis.....	42
2.3.2 Determination of protein concentration by Bradford Ultra Assay	43
2.3.3 Determination of protein concentration by Direct Detect	44
2.3.4 Transfer of proteins to nitrocellulose membranes	44

2.3.5 Western blot analysis of proteins on nitrocellulose membrane	44
2.3.6 Post-nuclear cell lysis.....	45
2.4 Proteomic Techniques.....	45
2.4.1 Collecting and lysing cells for di-glycine peptide enrichment protocol.....	45
2.4.2 Extraction of peptides through Methanol-Chloroform precipitation.....	45
2.4.3 In-gel Tryptic digestion	46
2.4.4 In-solution tryptic digestion	47
2.4.5 SEP-PAK purification	47
2.4.6 Pull-down of ubiquitinated species using UbiQapture.....	48
2.4.7 Preparation of Di-Glycine antibody columns and immunoprecipitation of di-glycine peptides	48
2.4.8 Mass Spectrometry.....	49
2.4.9 Mass Spectrometry Data Analysis	50
2.4.10 Label-free protein quantitation by Normalised Spectral Index (SINQ)	51
2.4.11 Analysis of Mass Spectrometry data by Peaks de novo	51
2.4.12 Pathway Analysis	52
2.5 In-Silico Analysis.....	52
2.5.1 Data Handling of the MitoCheck time lapse-movies	52
2.5.2 Development of the algorithms and validation	52
2.5.3 Manual identification and tracking of cells.....	53
2.5.4 Database Annotation	53
2.5.5 Gene Set Enrichment Analysis.....	54
2.5.6 Correction of movies with internal shakes	54
2.6 Migration Assays	54
2.6.1 Single Cell Motility Assay	54
2.6.2 Wound Healing Assay.....	55
2.6.3 Spreading Assay.....	55
2.6.4 Immunofluoresence	56
2.6.5 Cell tracking in MetaMorph.....	56
2.6.6 Quantitating Immunofluoresence	57
2.6.7 Wound Healing Perimeter Analysis.....	57
Chapter 3 – A proteomic approach to identify candidate BRCA1 E3 ligase substrates.....	59
3.1 Introduction.....	59
3.1.1 BRCA1 as an E3 ubiquitin ligase.....	59
3.1.2 Enrichment of ubiquitinated species	60
3.1.3 Identification of ubiquitinated species	61

3.1.4 Aims	65
3.2 Results	69
3.2.1 Isolation of ubiquitinated species for UbiQapture Pulldown	69
3.2.2. Quantitative Proteomics on ubiquitin-conjugated proteins	74
3.2.3 Pathway Analysis of UbiQapture Pull-down results	81
3.2.4 Ubiquitin Remnant Pulldown	84
3.2.5 Generating a candidate list of potential BRCA1 E3 ubiquitin ligase substrates.	91
3.3 Discussion	92
3.3.1 Methods Development and data processing	92
3.3.2. The Biological Output from Proteomic Experiments	99
3.4 Conclusions	108
Chapter 4 – Linking candidate substrates to a motility phenotype using in-silico approach	110
4.1 Introduction	110
4.1.1 BRCA1 candidate substrates – next step	110
4.1.2 The MitoCheck Database	111
4.1.3 Aims of the Chapter	113
4.2. Results	115
4.2.1 Developing an algorithm to measure cell tracking	115
4.2.2 Establishing the variables for cell tracking algorithms	118
4.2.3 Validation of cell tracking algorithm	121
4.2.4 Genome-wide siRNA knockdown effect on motility	125
4.2.5 Establishing the Motility Database - Linking Gene Ontologies and KEGG Pathways to the Motility Database	131
4.2.6 Gene Set Enrichment Analysis of KEGG pathways	133
4.2.7 The Carbonic Anhydrase family	135
4.2.8 BRCA1 Candidates linked to the MitoCheck Database	137
4.3 Discussion	141
4.3.1 Developing an algorithm and tracking cells	142
4.3.2 Genome-wide analysis	144
4.3.3 Mining the database for interesting hits	146
4.3.5 Linking Candidates to Motility	148
4.4 Conclusions and Future Directions	151
Chapter 5: Establishing BRCA1's role in Cell Motility	153
5.1 Introduction	153
5.1.1 BRCA1 role in invasion and metastasis in breast and ovarian cancer	154
5.1.2 Selection of candidates based on proteomics and MitoCheck results	154

5.1.3 <i>In vitro</i> methods of analysing motility	155
5.1.4 Aims of the Chapter	156
5.2 Results	158
5.2.1 Single Cell Motility assay of HCC1937 cell lines.....	158
5.2.2 Wound healing assay of HCC1937 cell lines	160
5.2.3 siRNA knockdown of top candidate BRCA1 substrates	164
5.2.4 Single cell motility on multiple candidates	165
5.2.5 The effect of ezrin knockdown on cell spreading, cell motility and wound healing.....	168
5.3 Discussion	181
5.3.1 BRCA1 Ubiquitin ligase domain associated with motility and spreading.	181
5.4 Conclusions and Future Directions	188
Chapter 6: Discussion	189
6.1 Future work and clinical implications	192
Bibliography	195

List of Abbreviations

2D – two dimensional
3D – three dimensional
6xHis – 6 x histidine
aa – Amino acid
APC – adenomatous polyposis coli
ATP – adenosine triphosphate
BARD1 – BRCA1 associated RING domain protein 1
BRCA1 – breast cancer 1, early onset protein
BRCA2 – breast cancer 2, early onset protein
BRCT domain – BRCA1 C Terminus domain
C61G, C64G – cysteine to glycine mutation on aa 61, on aa 64
CA – carbonic anhydrase
CID – collision induced dissociation
CO₂ – carbon dioxide
CPFP – Central Proteomics Facility Pipeline
CTCF – Corrected total Cell Fluorescence
C-terminus – Carboxyl terminus
DAPI - 4',6-diamidino-2-phenylindole
DNA – deoxyribonucleic acid
DSS - disuccinimidyl suberate
DUBs – deubiquitinating enzymes
E1 – E1 ubiquitin activating enzyme
E2 – E2 ubiquitin conjugating enzyme
E3 – E3 ubiquitin ligase
ECM – extracellular matrix
ELISA – Enzyme Linked Immunosorbent Assay
EMT – epithelial to mesenchymal transition
ER – Estrogen Receptor
ERAD – Endoplasmic Reticulum Associated Protein Degradation
ERM – ezrin, radixin, moesin
ESI – Electrospray Ionisation
FDR – False Discovery Rate
FERM – 4.1 protein, ezrin, radixin, moesin
GFP – green fluorescent protein
GSEA – Gene Set Enrichment Analysis
HCC1937 – breast cancer epithelial cell line
HCC1937 + I26A BRCA1 – HCC1937 cells electroporated with full-length BRCA1 with point mutation I26A-pIRES2-EGFP
HCC1937 + WT BRCA1 – HCC1937 cells electroporated with full-length BRCA1-pIRES2-EGFP
HCD – high energy collisional dissociation

HECT - Homology to E6AP Carboxy Terminus
 HER2 – Human Epidermal Growth Factor Receptor 2
 HPLC – high performance liquid chromatography
 I26A – Isoleucine to Alanine mutation on aa 26
 ID-4 – Inhibitor of DNA binding protein 4
 IDs – identifications
 IF – immunofluorescence
 iTRAQ - Isobaric tags for relative and absolute quantitation
 KEGG – Kyoto Encyclopedia of Genes and Genomes
 Lys – Lysine
 MALDI - Matrix Assisted Laser Desorption/Ionization
 MAPK – Mitogen Activated Pathway Kinase
 MCAR – missing completely at random
 MG-132 – proteasome inhibitor
 MMPs – Matrix Metalloproteases
 mRNA – Messenger Ribonucleic Acid
 MS – mass spectrometry
 mtDNA – mitochondrial deoxyribonucleic acid
 NA – numerical aperture
 NEM – N-ethylmaleimide
 NES – nuclear export sequence
 NLS – nuclear localization sequence
 N-terminus – Amino terminus
 OeRC – Oxford e-Research Centre
 PBS – phosphate buffered saline
 PR – Progesterone Receptor
 pSer-X-X-Phe – phosphorylated serine, any aa, any aa, phenylalanine
 PTM - post translational modifications
 PVI – perimeter variability index
 qRT PCR – quantitative real time polymerase chain reaction
 RING domain – Really Interesting New Gene domain
 ROS – reactive oxygen species
 SDS - Sodium dodecyl sulfate
 SDS-PAGE – sodium dodecyl sulfate polyacrylamide gel electrophoresis
 SILAC – stable isotope labeling by/with amino acids in cell culture
 SINQ – Normalised Spectral Index Quantitation
 siRNA – small interfering RNA
 SUMO – Small Ubiquitin-like Modifier
 TNBC – Triple Negative Breast Cancers
 TUBEs – Tandem Repeated Ubiquitin Binding Entities
 Tyr – tyrosine
 UbcH5 - Ubiquitin-conjugating enzyme E2 D1
 UPS – ubiquitin proteasome system

List of Figures

- 1.1 The ubiquitination pathway
- 1.2 BRCA1 domains, binding sites and mutations of BRCA1
- 1.3 Contributing factors to the starting point of the thesis

- 3.1 Schematic of proteomic approaches for determining candidate BRCA1 E3 ligase substrates
- 3.2 Expression of BRCA1 in HCC1937 cell lines
- 3.3 Optimisation of lysis protocol
- 3.4 Western blot of pulldown of ubiquitinated species using UbiQapture
- 3.5 Results from UbiQapture pulldown
- 3.6 Plot of log₂ fold change hits from amalgamated UbiQapture data sorted on reproducibility
- 3.7 Tables of proteins most consistently enriched in WT and I26A from UbiQapture results
- 3.8 Gene ontology and pathway analysis results
- 3.9 KEGG Pathway analysis results
- 3.10 KEGG Pathway map for adherens junction
- 3.11 Method development of GX41 pulldown
- 3.12 Di-glycine pulldown results

- 4.1 Schematic of analysing time-lapse movies
- 4.2 F-measure plots to determine optimum kernel size and threshold
- 4.3 3D visualisation of MitoCheck videos
- 4.4 Validation and frame shifting
- 4.5 Genome-wide analysis of cell motility
- 4.6 Genome-wide analysis of cell motility with significant hits
- 4.7 Genome-wide motility database
- 4.8 The effect of acetazolamide and zinc sulphate on HeLa cell motility
- 4.9 Candidates overlaid on genome-wide motility graph

- 5.1 Single cell motility assays on HCC1937 cell lines
- 5.2 Wound Healing experiments on HCC1937 cell lines
- 5.3 Leading edge perimeter analysis of wound healing experiments
- 5.4 Knockdown of candidates in HCC1937 cell lines and single cell motility assay
- 5.5 Schematic of single cell motility, spreading and wound healing pipeline
- 5.6 siRNA knockdown of ezrin after 24 hours
- 5.7 The effect of ezrin knockdown on cell spreading
- 5.8 The effect of ezrin knockdown on single cell motility
- 5.9 The effect of ezrin knockdown on wound healing and immunofluorescence
- 5.10 Leading edge perimeter analysis of ezrin siRNA wound healing experiments
- 5.11 Possible model for BRCA1 regulation of motility through ezrin

List of Tables

2.1 Table of Antibodies

3.1.a BRCA1 substrate candidate list based on proteins enriched in HCC1937 + WT BRCA1 cells from UbiQapture pulldown and ubiquitin remnant profiling

3.1.b BRCA1 substrate candidate list based on proteins enriched in HCC1937 + I26A cells from UbiQapture pulldown and ubiquitin remnant profiling

4.1 Gene Set Enrichment Analysis (GSEA) of KEGG pathways

4.2.a BRCA1 substrate candidate table with MitoCheck and motility data

4.2.b BRCA1 substrate candidate table with MitoCheck and motility data

Chapter 1: Introduction

1.1 Introduction

Breast cancer type 1 susceptibility protein (BRCA1) has been established as one of the most important tumour suppressor proteins associated with breast and ovarian cancer. It is estimated that 10% of all breast cancers are hereditary and up to half of these cases are due to germline mutations within the BRCA1 or BRCA2 genes (Antoniou et al. 2003; Easton et al. 1993). Women with BRCA1 mutations are predisposed to developing early-onset, highly-metastasizing breast and ovarian cancer (Foulkes et al. 2003; Rodríguez-Pinilla et al. 2006).

The protein has been associated with a variety of cellular functions, including roles in the DNA damage repair pathways, cell cycle checkpoint, transcriptional regulation and centrosome amplification. Whilst many of these pathways are important to prevent tumourigenesis, the exact mechanisms of tumour suppression still remain unclear. BRCA1 has traditionally been studied as a nuclear protein, but there is an emerging literature that indicates BRCA1's reach may go far beyond the nucleus. This introduction seeks to review BRCA1's structure and function within the cell, specifically focusing on the more newly-associated extra-nuclear functions.

A vast literature exists, and a Scopus literature search for the term 'BRCA1' generates over 10,000 articles. However, many questions remain unanswered; how does BRCA1 exert its tumour suppression activity, why is there tissue specificity, can BRCA1 associated pathways be exploited for therapeutic interventions and why are BRCA1 cancers so highly metastasising? Substantial efforts have gone into answering these questions, with still elusive answers. We anticipate that some of these questions may be usefully addressed by examining BRCA1's function outside of the nucleus.

1.2 BRCA1 as a Tumour Suppressor

1.2.1 BRCA1 association with breast and ovarian cancer

BRCA1 was first identified as a gene associated with breast and ovarian cancer susceptibility in 1994 (Miki et al. 1994) and since then it has been shown that women carrying mutations within the BRCA1 gene have an 80% and 40% lifetime risk of developing breast and ovarian cancer respectively. BRCA1 germline mutations increase the risk of developing tumours in these tissues by 5-10 fold relative to the general population (Easton et al. 1995). As a result, enormous efforts have gone into investigating the biology of BRCA1 with the hope that this may improve our understanding of the complex mechanisms underlying hereditary breast and ovarian cancers so that these aggressive tumours can be treated more effectively.

BRCA1 also appears to have an important role in non-familial breast and ovarian cancers. Whilst somatic mutations within the BRCA1 gene in sporadic cases of breast and ovarian cancer are relatively rare (Futreal et al. 1994; Merajver et al. 1995), many studies have demonstrated a dysregulation of BRCA1 through epigenetic mechanisms. Hypermethylation of the BRCA1 promoter is observed in sporadic cases and is associated with poor survival rates (Esteller et al. 2000; Wei et al. 2005; Hsu et al. 2013). Upregulation of the transcriptional regulator ID4, which downregulates BRCA1 expression is also observed in sporadic cases of breast cancer (Turner et al. 2007; Beger et al. 2001). In addition BRCA1 protein and mRNA expression levels are reduced in invasive breast cancers relative to benign tumours, further demonstrating the proteins importance in tumour suppression and more specifically metastasis (Seery et al. 1999; Thompson et al. 1995; Wilson et al. 1999).

1.2.2 BRCA1 association with basal-like breast cancer

Understanding the nature and behaviour of a breast cancer is vital to administering the optimum treatment. There are multiple forms of classifications of breast cancers to aid this process, based on histopathology, gene expression, grade, stage and receptor status. Each tumour falls into a sub-category within each classification and is informative in assessing the state and prognosis of the cancer. In order to understand BRCA1's relevance in these classifications, it is important to appreciate the molecular subtypes, which are primarily based on gene expression patterns and receptor status.

The receptors that are of specific interest are the oestrogen receptor (ER), progesterone receptor (PR) and the human epidermal growth factor 2 (HER2) as their expression is often altered in breast cancers. Moreover, these receptors are targets for endocrine therapeutics such as Tamoxifen and Herceptin. Triple-negative breast cancers (TNBCs) are carcinomas that do not express any of the three aforementioned receptors which renders them difficult to treat (Lakhani et al. 2000). Further categorisation of breast carcinomas was proposed by Sørlie et al. based on gene expression patterns, which included receptor status. Five subtypes were proposed: normal-like breast cancer, luminal-A, luminal-B, ERBB2 and basal-epithelial like breast cancer (Sørlie et al., 2003; Sørlie et al., 2001). From these subtypes, the basal-epithelial like breast cancers exhibited the lowest overall patient survival. These tumours have subsequently been shown to be aggressive and invasive in their phenotype and the TNBCs were initially considered to fall into this category (Livasy et al. 2006).

The characterisation of the different subtypes has recently been refined based on the comprehensive study undertaken by the Cancer Genome Atlas Network. Molecular profiles of tumours from 825 patients were analysed for genomic mutations, DNA copy number, DNA methylation levels, mRNA expression levels and reverse phase protein

arrays. This study showed that 75% of TNBCs are basal-like (expressing basal cell keratins 5/6/17) with the remaining 25% having a different mRNA expression profile (The Cancer Geonme Atlas Network 2012).

Interestingly, BRCA1 mutated breast carcinomas tend to exhibit a phenotype which most closely resemble the basal-like breast tumours, with studies showing that 61-85% of BRCA1-mutated tumours display features that place them into the basal-like category (Melchor et al. 2008; Waddell et al. 2010; Jönsson et al. 2012). The basal-like breast tumours are the most aggressive subtype of carcinoma and given that the majority of BRCA1 tumours fall into this category further demonstrates that there is an important link between aggressive metastatic breast cancers and BRCA1 protein levels.

1.2.3 BRCA1's links with other cancers

BRCA1's primary association with cancer is through its links with the hormone-sensitive breast and ovarian tissues. However, more recently BRCA1 has also been connected to tumourigenesis within other tissues, both hormone sensitive and non-hormone sensitive.

Prostate cancer has been explored to determine whether there are any links with BRCA1, given that the prostate is another hormone-sensitive organ. Early studies linked both BRCA1 and BRCA2 to prostate cancer due to the chromosomal proximity of a candidate susceptibility gene and germline mutations in BRCA1 conferred an increased risk of prostate cancer (Thompson & Easton 2002; Struewing et al. 1997). A recent sequencing study of 886 patients further reinforced the involvement of BRCA1 with prostate cancer as deleterious BRCA1 germline mutations increased the relative risk of development of prostate cancer by 3.75 fold (Leongamornlert et al. 2012).

Germline mutations within BRCA1 have been implicated in early-onset colorectal cancer (Suchy et al. 2010) and furthermore loss of the BRCA1-interacting partner protein, BARD1 (important in the enzymatic activity of BRCA1, as discussed in section 1.4.1) is associated with poor clinical outcome (Sporn et al. 2011). Another study shows the relative risk of developing upper gastro-intestinal cancers, specifically oesophageal and stomach, increases in families with BRCA1 mutations (Moran et al. 2012) and a reduced expression of BRCA1 is correlated with reduced survival rates (Zhang et al. 2013). Dysregulation of BRCA1 has also been reported in pancreatic cancer (Lucas et al. 2013; Iqbal et al. 2012).

BRCA1 has also been studied for a possible association with lung cancer, but dysregulation of BRCA1 has different effects depending on the tissue and stage of tumourigenesis. Low mRNA expression levels of BRCA1 and promoter methylation were frequent in lung adenocarcinomas (Lee et al. 2007), consistent with the notion that BRCA1 is important for tumour suppression. However, high levels of BRCA1 mRNA expression in non-small cell lung cancer is associated with chemoresistance, specifically against platinum-based chemotherapeutics, and poor survival at later stages (Reguart et al. 2008; Taron et al. 2004). Platinum-based reagents such as cisplatin crosslink DNA thus stimulating apoptosis, but the presence of BRCA1 and its role in DNA damage repair may counteract this mechanism of drug action, promoting the survival of the cancer cells. This would be a particular selective advantage to cancer cells establishing in a colony at a secondary site. This makes BRCA1 expression levels of clinical significance and important when tailoring treatment.

Each of these studies showing the dysregulation of BRCA1 in a variety of different tissues further support the notion that BRCA1 has a more global role in tumourigenesis and is not restricted to breast and ovarian cancers. Germline mutations of BRCA1 increase

the risk of development of a variety of different cancers and are associated with poor survival rates, indicating it may be important in metastasis.

1.3 The Ubiquitination Pathway

One of the major functions of BRCA1 is its E3 ubiquitin ligase activity involved in the ubiquitination pathway. The study of this post-translational modification is still developing and is not as well characterized compared to phosphorylation. This is in part due to the complexity and variety of ubiquitin pathway regulation, enzyme-substrate interaction and downstream biological consequences. Prior to considering BRCA1's role as an ubiquitin ligase, we will briefly discuss the ubiquitination pathway, which has been extensively delineated by multiple reviews (Mani & Gelmann 2005; Komander 2009; Deshaies & Joazeiro 2009)

Ubiquitination is a complex post-translational modification, which involves the ligation of a small 76 amino acid protein, ubiquitin, onto substrate proteins. Similar to phosphorylation, ubiquitination is important in modulating the behaviour and determining the fate of the substrate. This post translational modification is a fascinating process with several different types of ubiquitination based on the site of ubiquitination on the target protein, the number of ubiquitin proteins involved and the nature of the linkage. The versatile nature of the process leads to a variety of downstream cellular consequences depending on these variables.

Further complexity arises from the three principal subtypes of proteins involved in the enzymatic cascade that results in ubiquitination of a target protein. The E1 ubiquitin activating enzyme activates the ubiquitin moiety in an ATP-dependent process. The E1 then transfers the ubiquitin moiety onto the cysteine of an E2 conjugating enzyme. The third protein in the pathway is the E3 ubiquitin ligase which brings together the ubiquitin-bound E2 conjugating enzyme and substrate and catalyses the C-terminal transfer of

ubiquitin to a lysine on the substrate, resulting in mono-ubiquitination of the protein.

Within this final category there are two principal classes of E3 ubiquitin ligases, HECT (Homology to E6AP Carboxy Terminus) domain E3 ligases and RING (Really Interesting New Gene) finger E3 ligases, the latter being the category to which BRCA1 belongs.

The number of proteins for each group vary, with there being approximately 10 E1 enzymes, 40 E2 enzymes and over 600 E3 ligases encoded by the human genome with overlapping specificity (Grabbe et al. 2011). Within this, target specificity can be determined by E3 ligases and the nature of the linkage can be influenced by the E2 protein involved in the complex (Christensen et al. 2007).

Proteins can be ubiquitinated in a variety of different ways (See Figure 1.1). There are two major forms of ubiquitination. Mono-ubiquitination is the conjugation of a single ubiquitin to the target protein and there can be multiple sites of mono-ubiquitination within the same protein. Secondly there is poly-ubiquitination, which is a chain of ubiquitin moieties on the substrate protein. The C-terminal glycine of the ubiquitin binds to the epsilon amino group of a lysine on the target protein. When forming poly-ubiquitin chains, the ubiquitin can bind to any of the lysine residues within ubiquitin, adding another layer of versatility to this complex post-translational modification. The nature of the linkage can determine the downstream biological consequences of the post-translationally modified protein's function or fate.

The characteristics that determine the nature of the poly-ubiquitin chains are yet to be fully understood or clarified. However, recent work suggests that the E2 and the amino acids adjacent to the lysine are partially responsible for the nature of the chain (Sadowski & Sarcevic 2010).

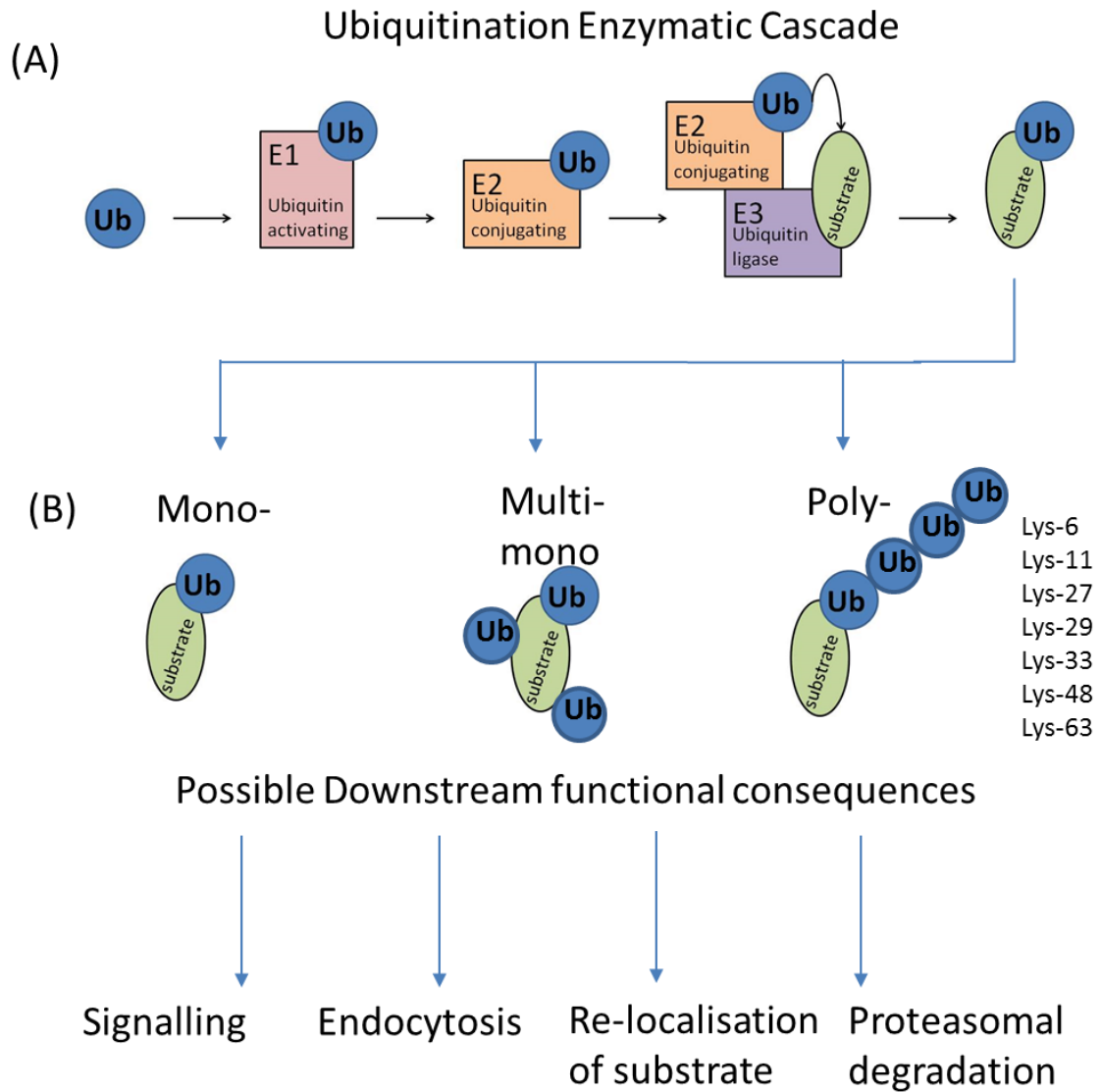


Figure 1.1 – The Ubiquitination Pathway – (A) Ubiquitin binds from free ubiquitin to the substrate through 3 enzyme cascade. The E1 binds ubiquitin and transfers it to the E2. The E3 can then either directly bind the ubiquitin from the E2 and transfer to the substrate (HECT) or act as a platform upon which the E2 and substrate are brought into proximity to transfer the ubiquitin moiety onto the substrate. (B) There are multiple forms of ubiquitination, mono-ubiquitination, multi-mono-ubiquitination and poly-ubiquitination, the latter can form chains through linkages with different Lysine residues. There are a variety of downstream consequences depending on substrate, enzyme, partner proteins, ubiquitin linkage and cellular localisation.

There are several forms of poly-ubiquitination with a variety of downstream functional consequences. Originally the poly-ubiquitination was thought to occur through ubiquitin proteins binding to each other through Lysine 48 residues, with 4 ubiquitin moieties required to target the substrate protein for proteasomal degradation. However now it has been established that there are 7 lysine residues in ubiquitin, all of which can be ubiquitinated. Poly-ubiquitin chains that form through lysine 27 and lysine 29 are associated with lysosomal degradation; lysine 11 has been shown to be involved with cell cycle control and endoplasmic reticulum associated protein degradation (ERAD). Ubiquitination is also associated with signalling and stabilisation of substrate proteins with lysine 63 poly-ubiquitin chains linked to signalling transduction pathways of NF- κ B (Chen & Sun 2009). As discussed in Section 1.4.1 BRCA1 mediates several different forms of ubiquitination.

1.4 Structure of BRCA1

To understand BRCA1's function it is vital to understand the structure of the protein (See Figure 1.1). BRCA1 is a large protein, with full length BRCA1 consisting of 1863 amino acids with a molecular weight of 220kDa. It has two domains which have been shown to be essential for many of its functions at the N and C terminus. The rest of the protein has not been studied in as much detail but still consists of some important domains including a DNA binding domain (aa 452-1079) (Paull et al. 2001), a nuclear export sequence (aa 81-99) (Rodríguez & Henderson 2000) and nuclear localisation sequence (aa 503-508) (Y. Chen et al. 1996). A coiled coil domain (aa 1369-1418) which binds PALB2 involved in DNA damage repair (Orelli et al. 2001).

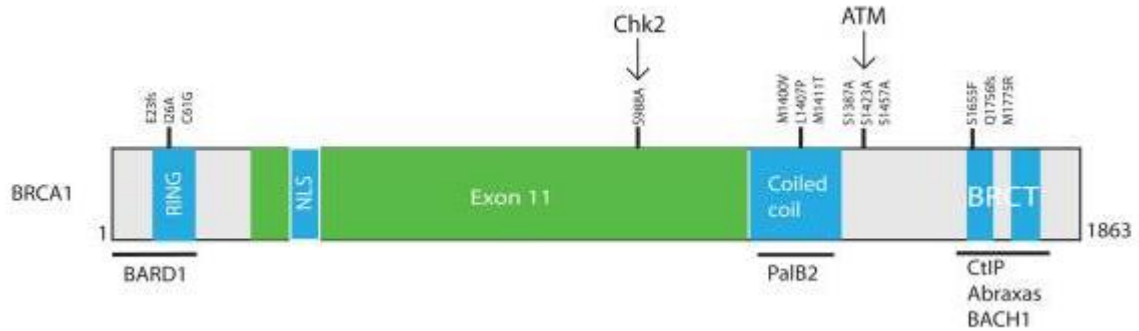


Figure 1.2. BRCA1 domains, binding sites and mutations of BRCA1: Full length BRCA1 consists of several important domains including the N-terminal RING domain, the C-terminal BRCT repeats, the nuclear localisation sequence and coiled-coil domain. Mutations in RING domain are of particular interest. C61G is a pathogenic mutation found in breast cancers which abrogates the binding to BARD1 and the E2 Ubch5. I26A is a synthetic mutant, used experimentally in this thesis and disrupts the binding to Ubch5, but still interacts with BARD1. Figure from Li & Greenberg, 2012.

BRCA1 undergoes multiple post-translational modifications. It has various phosphorylation sites that are cell cycle dependent (Cortez et al. 1999; Ruffner et al. 1999; Coene et al. 2005; Xu et al. 2001). Both Auto-ubiquitination and SUMOylation of BRCA1 increases the E3 ligase activity of the protein, thus disruption of these post-translational modifications may result in abrogation of BRCA1 enzyme activity (Morris et al. 2009; Mallery et al. 2002).

1.4.1 The RING domain

The first of BRCA1's highly conserved domains is the N-terminal zinc-binding RING (Really Interesting New Gene) domain, which has a role in the final step of the post-translational modification – ubiquitination. BRCA1 is an E3 ubiquitin ligase which coordinates the addition of ubiquitin to substrate proteins and this is the only known enzymatic activity of the protein. Whilst the domain only makes up approximately 5% of the protein, the proportion of cancer associated mutations found within the RING domain is very high, close to 20%, indicating the importance of the domain in its tumour suppression activity (Brzovic et al. 2003a).

The first 101 amino acids constitute the RING domain and this coordinates zinc cations via zinc binding residues which are flanked by alpha-helices. BRCA1 has a partner protein known as BARD1, which also contains a RING domain and BRCT repeats. The heterodimerization of these two proteins is of great importance regarding function of BRCA1, as it stabilises a four helix bundle is formed which stabilises the BRCA1-BARD1 complex *in vitro* and significantly increase the ubiquitin ligase activity (Hashizume et al. 2001; Xia et al. 2003). Up to 75% of BRCA1 is found in its heterodimeric state and BARD1 appears to have an important role in tumour suppression given that mutations are observed in ovarian and breast cancers (Baer & Ludwig 2002). Similarly BRCA1 auto-ubiquitination results in increased ubiquitin ligase activity.

Mutations in the BRCA1 RING domain that are associated with the development of cancer confer abrogation of the ubiquitin ligase activity. Two of the most common pathogenically relevant mutations are point mutations C61G and C64G, which result in the inability of the RING domain to coordinate the zinc ions, thus disrupting the BRCA1-BARD1 heterodimer. In addition, the complex fails to recruit the E2 enzyme Ubch5, all of which renders the BRCA1 ubiquitin ligase inactive (Hashizume et al. 2001; Ruffner et al. 2001).

Another mutant which has been synthetically developed and characterized to assess BRCA1's E3 ubiquitin ligase activity is the I26A point mutation. This mutation renders BRCA1 unable to bind to the E2 protein Ubch5 and thus unable to perform as an effective ubiquitin ligase (Brzovic et al. 2003a). This is of particular interest in relation to this thesis, as we assess BRCA1 ubiquitination using this model. Many studies have subsequently used the BRCA1 I26A variant to assess BRCA1's ubiquitination and identify substrates (Ma et al. 2010; Yu et al. 2006; Eakin et al. 2007).

BRCA1-BARD1 heterodimer has multiple ways to ubiquitinate substrates. The canonical ubiquitination pathway involves building a poly-ubiquitin chain through Lys-48 linkages which targets the modified protein for proteasomal degradation. However, in addition BRCA1 has been shown to mono-ubiquitinate a number of its established substrates including histone proteins, oestrogen receptor alpha and γ -tubulin. In addition, it can poly-ubiquitinate its substrates as well, but BRCA1 has shown to form poly-ubiquitin chains through Lys-6 and Lys-63 residues, as well as the more traditional Lys-48 (Yu et al. 2006; Sato et al. 2004; Wu-Baer et al. 2003). The fact that established BRCA1 E3 substrates are not exclusively targeted for proteasomal degradation is of paramount importance when developing approaches that seek to identify novel substrates. Many studies utilise proteasome inhibitors to elicit changes in the ubiquitination profiles of cells to uncover candidate substrates (Han et al. 2009; Vasilescu et al. 2005; Wagner et al. 2011). However in the instance of non-degradative ubiquitination, the use of proteasome inhibitors would mask any changes to potential substrates, thus one cannot rely on the use of proteasome inhibitors to increase the yield of ubiquitinated proteins for subsequent proteomic analysis. When studying the ubiquitination signal for these non UPS targeted putative substrates alternative strategies for enrichment of mono-ubiquitinated target protein will be required.

Since it represents the only known enzymatic activity of BRCA1, the ubiquitin ligase activity has come under some scrutiny as to its relevance in tumour suppression of breast and ovarian cancer. Ludwig and colleagues indicate that the ubiquitin ligase activity is not essential for DNA damage response by homology-directed repair at double stranded breaks and more recently showed that I26A mutations in murine models show no difference in tumour development compared to mice with functional BRCA1 (Shakya et al. 2011; Reid et al. 2008). Conversely, mutations within the BRCT domains promoted

tumour development, suggesting that it is the BRCT domain that is of significance in tumour suppression. However, studies have shown that the BRCA1-BARD1 heterodimer is an effective, specific ubiquitin ligase with a highly conserved RING domain. Moreover, another recent study investigated a murine model with the C61G mutation, concluding that the RING domain is important for tumour suppression, as BRCA1-C61G did not prevent tumour formation and resembled *BRCA1*-null tumours. In addition Zhu et al. established that the BRCA1 ubiquitination activity was required for heterochromatin silencing through ubiquitination of histone 2A, which could not be carried out by the I26A mutant (Zhu et al. 2011).

The discrepancies observed could be due to several factors. Firstly, the nature of the RING domain mutations are subtly different which could impinge on the biology of the protein. Shakya et al. use the I26A mutant, which still heterodimerizes to BARD1, but is unable to engage with the E2 ubiquitin conjugating enzyme. Whereas Drost et al. use the C61G mutant which is unable to bind BARD1 in addition to not binding to the E2 enzymes. Similarly, differences in results between Zhu et al. and Shakya et al. may be due to differences in the murine models used (Shabbeer et al. 2012).

The studies carried out in the last five years begin to show that the ubiquitin ligase activity of BRCA1 may not be an important tumour suppressive function with respect to the DNA damage repair response. Reduced activity of the DNA damage repair response is likely to be a key pre-requisite for increased mutation rates that are important for initial cell transformation. If so, loss of the ubiquitin ligase activity of BRCA1 may not affect this. Nonetheless, BRCA is an effective ubiquitin ligase when heterodimerized with BARD1 and has a highly conserved RING domain, which is susceptible to pathogenically relevant mutations. All of this suggests that the ubiquitin ligase activity may be more relevant at other stages in the tumorigenesis cascade. It would be interesting to explore

whether the ubiquitin ligase activity is important in metastasis, which could account for the aggressive nature of BRCA1-deficient tumours and may not be directly attributed to genomic instability.

The BRCA1-BARD1 heterodimer has a number of known substrates and they appear to be targeted with high specificity *in vitro*, since BRCA1/BARD1 does not promiscuously ubiquitinate the generic acceptor protein hen egg lysozyme. Some of these substrates will be discussed in more detail throughout Section 1.5, but some of the known ubiquitination targets have been discussed in this review (Wu et al. 2008).

1.4.2 The BRCT repeats

At the C-terminus, BRCA1 consists of a pair of structurally identical motifs known as the BRCT (Breast cancer C-terminus) repeats which are typically 90 amino acids in length. These regions are highly conserved and many pathologically relevant mutations occur within these repeats, indicating their importance in tumours suppression activity (Williams et al. 2003). Whilst there is no enzymatic activity associated with the BRCT repeats, the domain interacts with a range of proteins through a phospho-peptide binding module which has selectivity for a pSer-X-X-Phe motif and the structural integrity of the BRCT tandem repeats is crucial for efficient binding with phospho-peptides (Rodriguez et al. 2003). This motif is present in a number of proteins involved in the DNA damage pathway and has been shown to be important a number of process such as nuclear targeting of BRCA1 (Rodriguez et al. 2004), transcriptional regulation and chromatin structure organization (Yamane et al. 2000; Monteiro 2000).

There have been over 120 missense mutations within the BRCT tandem repeats which have been observed in hereditary cancers, but the definitive roles of these putative residues in BRCA1 function have yet to be resolved. Although there have been many studies examining the biological consequence of these mutations (Rodriguez et al. 2004),

there is still more work to be done in identifying the biochemical structural significance of these mutations. Many of the mutations result in a truncation of the BRCA1 protein, thus preventing binding to partner proteins, including for example BARD1, with implications for reduced ubiquitin ligase activity and subsequent enhanced tumour progression as demonstrated by mice models (Shakya et al. 2011).

The BRCT repeats are required for BRCA1 to bind the phosphorylated partner proteins in order to carry out its tumour suppressive functions in the nucleus through its roles in the DNA damage repair response pathway. This requires the shuttling of the protein into and out of the nucleus, using the nuclear import and export signals embedded in the protein alternately. Similarly, given BRCA1's extra-nuclear functions also requires chaperoning the protein out to the relevant subcellular localisation.

1.5 Localisation and Function of BRCA1

BRCA1 can be found in many regions of the cell and not just restricted to the nucleus as originally believed. Here we explore some of the novel localisations and functions associated with BRCA1, with a focus on the extra-nuclear targets.

For BRCA1 to effectively carry out its functions there needs to be a well-established controlled system in order to shuttle BRCA1 to its different positions. Dysfunction within these pathways can lead to the cell being compromised in regards to apoptosis, cell cycle regulation and DNA damage repair, processes that are all important in tumourigenesis and more importantly, all functions of BRCA1. The protein is extremely large, 220kDa and has been associated with a number of different complexes thus the protein is likely to require chaperone, partner and cargo proteins to aid its movement around the cell.

1.5.1 Shuttling of BRCA1

At present, two distinct pathways have been established for the nuclear import of BRCA1. Firstly the protein has two nuclear localisation sequences (NLS) in exon 11 and through binding with importin alpha it can enter through the canonical importin alpha/beta pathway at the expense of GTP (C. F. Chen et al. 1996; Thakur et al. 1997). Alternatively it can piggyback into the nucleus through forming the heterodimer with BARD1 (Fabbro et al. 2002). This heterodimerization to BARD1 masks the nuclear export signal, thus retaining pools of BRCA1 heterodimers in the nucleus (Fabbro et al. 2004).

However, BRCA1 is required at a number of cytoplasmic regions, therefore the regulation of nuclear export is also instrumental in determining correct localisation of the protein. BRCA1 has two Rev-like nuclear export sequences, which bind to nuclear export receptor CRM1 in order to translocate into the cytoplasm (Brodie & Henderson 2012).

The regulation of this process is complex and what stimulates BRCA1 to enter or leave the nucleus is not fully understood. An important consideration is the stability of the different BRCA1 complexes. BRCA1 has been shown to interact with a wide range of proteins and it often acts as a platform upon which other proteins integrate to form a functional complex. The relative stability and nature of the partner proteins in a BRCA1 complex may participate in determining its localisation at steady state. Furthermore, it may be through the regulation of partner proteins that localisation of BRCA1 is determined, for example BARD1 can regulate localisation of the heterodimer through the masking of the nuclear export signal (Fabbro et al. 2002). Similarly, MERIT40, a newly described protein, is involved in the RAP80/CCDC98-containing complex and it is required to retain the BRCA1 complex at DNA breaks by maintaining the stability of the RAP80/CCDC98 complex (Shao et al. 2009). Phosphorylation of BRCA1 by the

RAP80/CCDC98 complex is thought to be an important mechanistic aspect to the retention of BRCA1 at the site of double stranded breaks.

In order to get a better understanding of how BRCA1 is shuttled around the cell, we can draw information by looking at other tumour suppressor proteins, such as adenomatous polyposis coli (APC) (Brocardo & Henderson 2008), a protein found to localise to similar sites as BRCA1, such as the nucleus, mitochondria, centrosomes and cell membrane. APC exits the nucleus via the CRM1 pathway, which has shown to be the case for BRCA1 as well. Similarly, p53, FOXO and PTEN, other tumour suppressors, are regulated by mono-ubiquitination (Salmena & Pandolfi 2007), although with differing mechanisms. p53 is shuttled into the cytoplasm after mono-ubiquitination, whereas FOXO and PTEN are retained in the nucleus upon mono-ubiquitination. BRCA1 is known to be mono-ubiquitinated and auto-ubiquitinated, which significantly increases the ubiquitin ligase activity. Thus, the ubiquitination status of BRCA1 may influence its localisation through altering interactions with partner proteins.

1.5.2 BRCA1 and the nucleus

BRCA1 has multiple critical functions within the nucleus and these behaviours have been extensively studied. Many recent reviews comprehensively explore these major nuclear functions and their association with cancer (Huen et al. 2010; Silver & Livingston 2012; Starita & Parvin 2003; Roy et al. 2012). Its roles in maintaining genomic stability through the DNA damage repair responses, via homologous recombination and non-homologous end joining is a crucial factor to its tumour suppressive functions. Through these pathways, several E3 substrates have been identified.

In response to DNA damage repair several major BRCA1 complexes have been identified as being of importance, all with different partner proteins. BRCA1-A complex

interacts with MERIT40, Abraxas, Rap80 and deubiquitinases to engage with K-63 poly-ubiquitin at double strand breaks after ionising radiation (Li & Greenberg 2012). BRCA1-B involves TopBP1 and BACH1 to regulate the S-phase checkpoint, whereas BRCA1-C complex recruits CtIP, an E3 non-proteasome targeting substrate of BRCA1 and associates with chromatin after DNA damage, as well as play a role in G2/M checkpoint control (Yu et al. 2006).

In addition, the silencing of heterochromatin occurs through the mono-ubiquitination of histone 2A (H2A) and this has the downstream effect of transcriptionally repressing α -satellite DNA. This pathway may be involved regulating a number of the observed nuclear functions of BRCA1, as overexpression of α -satellite DNA is associated with BRCA1-deficient phenotypes such as reduced homologous recombination, centrosome duplication and checkpoint deficiency (Zhu et al. 2011).

RNA Polymerase II (RNAPII) is involved with transcription and stalls at lesions in the template strand of DNA. Upon stalling, BRCA1 ubiquitinates the hyper-phosphorylated subunit of RNAPII, Rbp1 as an initial step in the DNA damage repair response pathway and this has been associated with the targeting of RNAPII to the proteasome (Starita et al. 2005). PCNA and SAFB2 are two other proteins that have implicated in BRCA1-mediated ubiquitination, the latter of which is associated with non-proteasome targeting (Tian et al. 2013; Song et al. 2011).

Thus multiple BRCA1 E3 substrates have been identified with a range of different functions, based on previous known interactions. Many of these are not Lys-48 or proteasome linked, which indicates the versatility of BRCA1 as an E3 ubiquitin ligase.

1.5.3 BRCA1 and the mitochondria

BRCA1 has several well-characterised functions within the nucleus, such as its roles in the DNA damage repair pathways, in homologous recombination and non-homologous end joining. Mitochondria have been implicated in carcinogenesis through several different pathways and their function is necessary for cancer cell viability (Wallace 2012). Mutations in both nuclear and mitochondrial DNA can lead to the build-up of mitochondrial metabolites, activating stress response pathways. Furthermore, this can increase the production of reactive oxygen species (ROS), culminating in the alteration of transcription factors resulting in cell proliferation. The proximity of the mtDNA, localised in the mitochondrial matrix, to the ROS makes it susceptible to DNA damage and mutation rates within the mtDNA are often high in tumour cells (Lee & Wei 2009). As a result, mitochondria require DNA damage repair proteins in order to maintain the genomic integrity of the mtDNA.

Several studies have identified pools of BRCA1 in the mitochondria associated with the mtDNA, indicating that BRCA1 may carry out some of its established nuclear DNA damage repair functions in the mitochondria as well (Coene et al., 2005; Maniccia et al., 2009). This is further reinforced by the identification of Mre11 in the mitochondria, a member of the protein complex Mre11/NSB1/RAD50 involved in non-homologous end joining (Dmitrieva et al. 2011). The presence of the nuclear DNA repair machinery including BRCA1 suggests that there is an analogous system present in the mitochondria in order to protect mitochondrial genome integrity, although the nature of the DNA protection and repair has yet to be fully described. Further work must be done to understand the functional regions of BRCA1 relevant in the mitochondria in order to try elucidate the mechanism of the mitochondrial DNA damage repair.

However, these genes are coded by nuclear DNA and absent from the mtDNA, so further work must go into understanding how these proteins are imported into the mitochondria. It is likely to be a complex pathway which may require chaperones which unfold and refold proteins such as BRCA1, considering its size. Further insights into the functions of BRCA1 in mitochondria will come from comparison of the sites and rate of mtDNA mutation in cell lines expressing either functional or non-functional BRCA1.

1.5.4 BRCA1 and the centrosome

BRCA1 has also been shown to localise to the centrosome, the cytoplasmic organelle involved in microtubule organization, cell cycle and cell division. Centrosomes undergo maturation prior to mitosis, which involves the recruitment of γ -tubulin, a centrosomal protein important in the nucleation of microtubules. Dysregulation of centrosome formation and duplication is a characteristic hallmark of cancerous tissues (Nigg & Raff 2009). To investigate the function, localisation and interaction between BRCA1 and the centrosome several important studies have been undertaken over the past 5 years.

BRCA1 is required for maintaining centrosome number in breast epithelial cells, as the loss of BRCA1 leads to centrosomal amplification; the molecular basis of this is likely to be BRCA1's direct mono-ubiquitination of γ -tubulin (Starita et al. 2004). The ubiquitination of γ -tubulin is required for the maintenance of centrosome number and prevention of over-amplification. During mitosis, the inhibitory effect of BRCA1 needs to be reversed in order to allow centrosome duplication to occur. This process is thought to be regulated by AURKA which phosphorylates BRCA1 and inhibits the ubiquitin ligase activity thereby allowing centrosome amplification (Starita et al. 2004). BRCA1 engages

with a number of centrosomal proteins including cyclin B and cdc25B and targets them for degradation (Shabbeer et al. 2012).

Another interesting observation involving BRCA1 at the centrosome is the fact that a range of other proteins that are associated with homologous recombination such as XRCC2 and RAD51 which interact with BRCA1 in the nuclear context of homologous recombination, also associate with centrosomes. This is suggestive of a tight functional linkage between centrosomes and HR proteins. Moreover this could be a potential mechanism of ensuring important DNA damage repair related proteins, especially those of low copy number, are equally distributed into daughter cells during mitosis through interacting with the centrosomes.

BRCA1 interacts with numerous protein complexes involved in several pathways associated with cell cycle and mitotic progression. But the numerous papers that have been published over the last few years have predominantly demonstrated co-localisation and interaction of specific protein complexes with BRCA1. It is yet to be understood which of these processes are the most relevant to the mechanisms that ultimately lead to tumourigenesis.

1.5.5 BRCA1 links from the nucleus to the cytosol

An important recent discovery is the finding that BRCA1 ubiquitinates the hormone receptors, progesterone receptor and oestrogen receptor ER α . The progesterone receptor PR-A is ubiquitinated by BRCA1 and targeted for proteasomal degradation in the absence of progesterone. Moreover, under progesterone treatment, BRCA1-BARD1 heterodimer epigenetically silences promoter regions by mono-ubiquitination of H2A through interaction of PR-A (Calvo & Beato 2011).

BRCA1 mono-ubiquitinates a cytosolic receptor that binds the hormone oestrogen. Once ligand and receptor are bound the complex localises to the nucleus and binds hormone responsive DNA elements. The oestrogen pathway plays a major role in breast cancer, with over 70% cases of breast cancer over-expressing ER α . Moreover, the most notable drug to treat breast cancer, tamoxifen, is an inhibitor of the ER α . By means of synthesising differential sized BRCA1 and BARD1 constructs, Eakin et al. determined not only does BRCA1/BARD1 complex ubiquitinate ER α at the ligand binding domain, but also determined that the BRCA1 residues 1-241 and BARD1 residues extending beyond 170 were necessary for efficient ubiquitination (Eakin et al. 2007). It was proven that BRCA1/BARD1 was required for the mono-ubiquitination of ER α through the lack of ubiquitination when the cells expressed mutant BRCA1 C61G, C64G and the designed I26A construct which are all ligase dead. The proposed function of the mono-ubiquitination of ER α is regulation of ER α transcriptional activation by BRCA1. This is further reinforced by studies by Ma et al. who show that BRCA1 over-expression results in reduction of ER α activity. Moreover, they propose a model pathway, suggesting that BRCA1 ubiquitination is in direct competition with acetylation of ER α , which has the converse effect to ubiquitination, of increasing the sensitivity of ER α to oestrogen. Ubiquitination has an important role in competing with acetylation sites which subsequently alters the activity of the substrate (Ma et al. 2010). However this is not entirely conclusive, as all the interactions demonstrated are *in vitro*, it does not take into account all the other regulatory mechanisms of the oestrogen pathway, so there are likely to be other partner proteins which affect the behaviour of this interaction.

Nevertheless, these studies propose novel cytosolic functions of BRCA1-dependent ubiquitination, which may help us to understand the link between BRCA1 mutations and its propensity for tumour development in hormone sensitive tissues.

1.5.6 BRCA1, the cell membrane, and its role in cell motility

An important novel identification of BRCA1's functions is its role in cell motility. The aggressiveness of BRCA1 tumours have indicated that BRCA1 may be involved in the metastatic cascade and a recent study by Coene et al. may provide insights into how that may be the case. This article established that BRCA1 localise at membrane ruffles and focal adhesion sites. Moreover, they showed disruption of the ubiquitin ligase domain of BRCA1 results in increased cellular motility, impaired cell spreading and a hypermotile phenotype. Both the amino and carboxy terminals were required to restore the cells to a normal motile level. Further investigation identified that BRCA1 colocalises with ezrin, radixin and moesin protein complex (ERM), which is localised at the plasma membrane at focal adhesion sites, leading edges and ruffles and it interacts with the underlying cytoskeleton. Moreover, ezrin is a key component in tumour metastasis, making this family of proteins interesting candidates in relation to BRCA1-mediated motility (Coene et al., 2011).

Coene et al. performed wound healing assays with breast cancer cell lines transfected with functional wild type BRCA1 and BRCA1 with a point mutation rendering it ubiquitin ligase dead. Not only did the cell line with ligase dead BRCA1 wound repair at a significantly increased rate, the leading edges were not as uniform and structured. Additionally, cells 6-7 cells back from the leading edge were seen to migrate individually across the leading edge and into the wound space. These cells demonstrated a loss of cell-cell adhesion and a hypermotile phenotype which has been taken to be indicative of a metastatic characteristic.

This study, coupled with other observations of BRCA1's involvement in cell motility and cell polarity, suggest that mutations with BRCA1 may lead to abnormal

motile behaviour which is the first step of the metastatic cascade, which could be of enormous importance when considering the biology and treatment of BRCA1-mutated cancers (Maxwell et al. 2011; Yasmeen et al. 2008).

1.5.6.i Transcriptional control of BRCA1 and its effect on cell motility

Recent evidence suggests that BRCA1 may in addition have an effect on motility through transcriptional control. BRCA1 may repress a transcriptional repressor, slug, which is associated with the epithelial to mesenchyme transition (EMT), given that knockdown of BRCA1 increases slug expression (Proia et al. 2011). EMT is an important process during embryonic development and many of the same cellular changes, namely loss of cell-cell contact and gain of invasive behaviour have also been associated with the initiation of metastasis. Increased slug protein levels is associated with basal-like breast cancers, which is another phenotype associated with BRCA1-mutated tumours.

Wu et al. have recently shown that it is likely to be a feedback loop, with slug repressing BRCA1. The stability of slug through canonical Wnt signalling results in H3K4 demethylase being recruited to BRCA1's promoter and epigenetically silencing BRCA1 (Wu et al. 2012). These results further illustrate the importance of BRCA1 to cell motility and invasion.

1.6 Aims

BRCA1 is an important multi-functional caretaker protein with important roles to play in tumour suppression in breast and ovarian cancers. Emerging evidence suggests that BRCA1 may have a more global role in tumour suppression, with epigenetic abrogation of BRCA1 expression in sporadic cases of breast and ovarian cancer and mutations frequently observed in a range of other cancers. The BRCA1 mutated tumours

are characterised by being extremely aggressive and early onset, with a high propensity to metastasise. Whilst a tremendous amount of work has been carried out to establish BRCA1's role in the DNA damage repair response important for tumour suppression, there is a growing body of evidence to indicate that BRCA1 is important in metastasis through different means.

Coene et al. established that BRCA1 can regulate cell motility and a compelling argument is that BRCA1 exerts its metastatic-preventing functions through regulating cell motility. In order to try to further characterise BRCA1's role in motility and metastasis, we look to its primary biological functions. BRCA1 is an E3 ubiquitin ligase and whilst several substrates have been identified, the majority of these have been based on building upon known interactions. We seek to identify candidate BRCA1 substrates based upon a large-scale unbiased proteomic approach to identify ubiquitinated proteins. This will involve comparison of proteomic profiles between cells expressing functional BRCA1 and ubiquitin ligase dead BRCA1. This approach may uncover novel partners and pathways that BRCA1 may be associated with and shed light on ubiquitin signal mediated functions of BRCA1.

We then seek to identify whether any of these BRCA1 ubiquitination targets have a role in regulating cell motility to further characterise the viability of the candidates having a role in BRCA1 mediated metastasis-suppression. We have decided to undertake a genome-wide unbiased *in silico* approach to measure cell motility and then overlay the BRCA1 candidate substrates to identify the relevance of any of our candidates in spontaneous single cell motility.

Combining these two studies will allow us to distil a list of candidate proteins that regulate motility and exhibit BRCA1 dependent changes in ubiquitination. These candidates will then be prioritised based on their ubiquitination profile, their motility

profile and their relevance in the existing literature. We will then investigate the effects of altering the expression levels of these candidate target proteins on the spreading and motility profiles of breast cancer cells in order to assess their roles in BRCA1 ubiquitin ligase-mediated regulation of cancer cell motility.

Our general hypothesis is that one of the ways BRCA1 acts as a tumour suppressor is by ubiquitinating protein(s) that are involved in regulating cell motility and invasion. Mutations in the BRCA1 gene result in the disruption of this pathway, resulting in the possibility of tumours that are more likely to metastasize. Establishing these proteins and pathways may help elucidate further the nature of hereditary breast and ovarian cancers and illuminate novel targets for therapeutics. In Chapter 3 the proteomic approach will identify candidates involved in this motility-regulating pathway that are differentially ubiquitinated between cells expressing functional BRCA1 and ubiquitin ligase dead BRCA1. In Chapter 4, comparing these candidates to the curated database of an *in silico* genome-wide study of motility will allow us to further narrow in on our potential candidates. In Chapter 5, we aim to test these candidate *in vitro* to determine their relevance in the regulation of cell motility, invasion and metastasis.

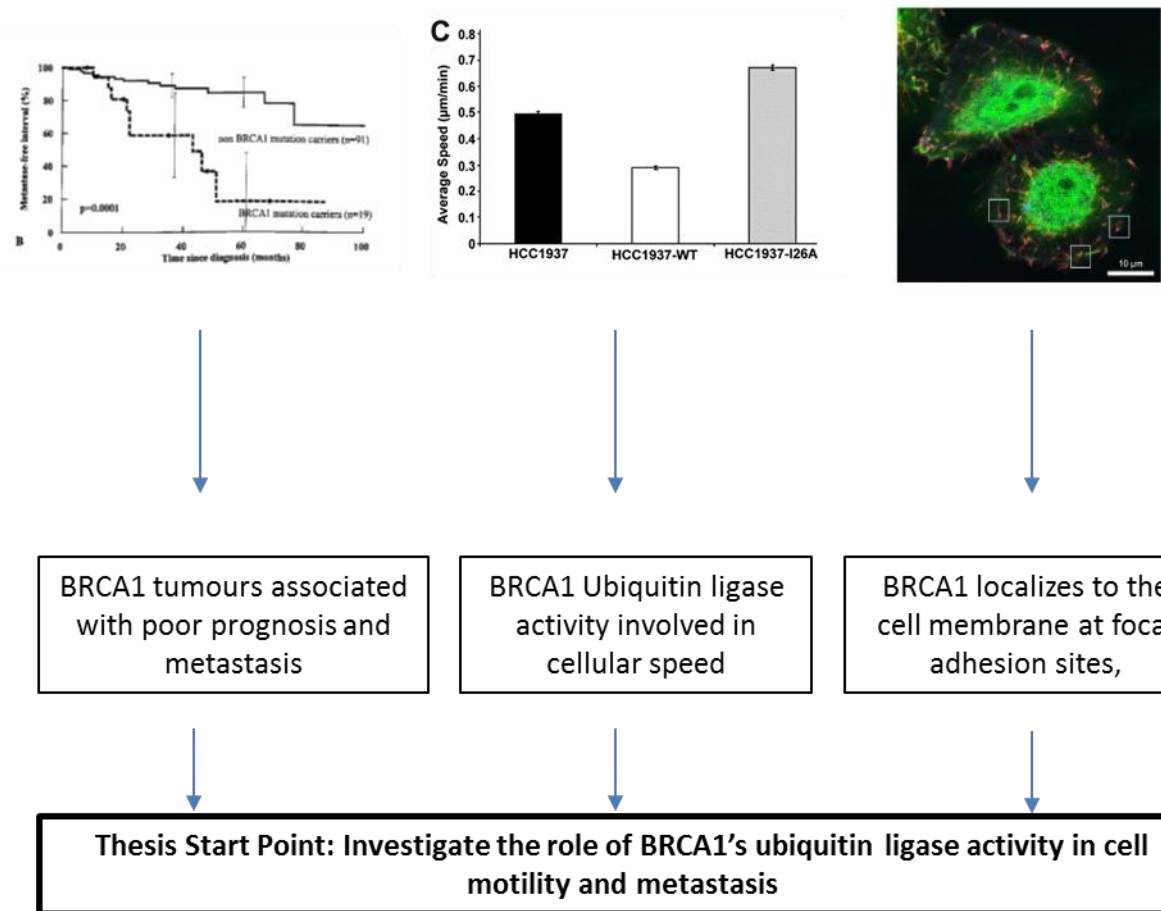


Figure 1.2: Three contributing factors to the start point of the thesis. (1) Kaplan-Meier plot of patients with BRCA1 tumours v non-BRCA1 tumours and their metastasis free interval. From Stoppa-Lyonne et al. 2000. (2) Mutated BRCA1 Ubiquitin ligase results in hypermotility. (3) BRCA1 C-terminal domain construct localises to the cell membrane with F-actin and ERM proteins. From Coene et al. 2011.

Chapter 2: Materials and Methods

2.1 Table of Antibodies

Antibody	Dilution	Host Species	Source
BRCA1 Ab-4	1:200	Mouse	Oncogene Science
Ubiquitin (P4D1)	1:500	Mouse	Santa Cruz Biotech
Lamin-B	1:1000	Mouse	Abcam
β -Actin	1:3000	Mouse	Sigma-Aldrich
Ezrin	1:500	Mouse	Abcam
Moesin	1:500	Goat	Abcam
δ -Catenin	1:200	Mouse	Abcam
Fermitin-2	1:500	Rabbit	Abcam
HRP-linked Anti-Mouse	1:5000	Goat	Sigma-Aldrich
HRP-linked Anti-Goat	1:20000	Mouse	Jackson ImmunoResearch
HRP-linked Anti-Rabbit	1:5000	Goat	Sigma-Aldrich
Cy3-conjugated Anti-Mouse	1:200	Goat	Jackson ImmunoResearch

2.2 Mammalian Cell Culture

2.2.1 General Cell Culture

Cell culture was performed in a category two laminar flow hood using sterile technique. Antibiotics were not used frequently in cell culture applications. Cultured cells were incubated at 37°C in the presence of 5% CO₂.

2.2.2 Culture of individual cell lines

HCC1937 cell lines were cultured in RPMI-1640 media, HeLa cells were cultured in Dulbecco's Modified Eagle Medium (DMEM), MCF-7 cells were cultured in Eagle's Minimum Essential Medium. All culture media were supplemented with 10% fetal calf serum and non-essential amino acids. Lab stock of the different HCC1937 cell lines were utilised. Generation of BRCA1 WT or BRCA1 I26A reconstituted HCC1937 cells was carried out by introducing BRCA1-pIRES2-GFP and I26A BRCA1-pIRES2-EGFP to

HCC1937 cells by electroporation as described by Coene et al. (Elisabeth D Coene et al. 2011).

2.2.3 Passaging of cells

Cells were typically grown to 80-90% confluency and passaged by trypsinisation with TrypLE (Gibco 12604-013). Trypsinisation was stopped by the addition of growth medium in equal volume. Cells were pelleted by centrifugation at 500 g at 4°C, supernatant removed, resuspended in growth media and seeded at the appropriate density. Cells that were destined for frozen storage were resuspended in 1ml of freezing medium (FCS with 10% DMSO). Cells were frozen overnight to -80°C in a polystyrene holder to achieve an appropriate cooling rate and then in liquid nitrogen.

2.2.4 Drug treatment of cells in culture

All drugs were dissolved in PBS or DMSO depending on solubility described by the supplier. Reconstituted HCC1937 cells were treated with 600µg/ml G418 as a selection drug. Media for these cells were changed every 2-3 days, with fresh G418 added. Serial dilutions of Acetazolamide were established in 10 µl DMSO (in order to minimise cytotoxicity effect of solvent). Serial dilutions of Zinc acetate were dissolved in 10 µl DH₂O. HeLa cells were treated with Acetazolamide and Zinc acetate for 48 hours, with fresh media and drug replacement after 24 hours.

2.2.5 Transfection of mammalian cells in culture

Transient transfections of siRNA against proteins of interest were carried out using Lipofectamine RNAiMAX (Invitrogen). HCC1937 cells and MCF-7 cells grown for

single cell assays and cells were plated at 50% confluency and grown in antibiotic free, 10% serum supplemented RPMI media and EMEM media respectively. siRNA treatment was carried out 24 hours after seeding cells. Volumes, concentrations and protocol of siRNA and Lipofectamine RNAiMAX were adjusted and carried out as suggested by product data sheet. Typically 20nM of siRNA in RNase free H₂O was diluted in 50µl FCS free media and combined with 1µl RNAiMAX lipofectamine in serum free media, combined and incubated for 20minutes at room temperature. 100 µl suspension was added to 500 µl fresh media in each 24 well plate well. Media was changed 6 hours after initial transfection. Cells were assayed either 24 or 48 hours after transfection. Expression of the targeted protein was determined by Western Blot analysis using monoclonal antibodies against the protein of interest.

Ezrin siRNA targeted ezrin mRNA at 5'-CAGGACTGATTGAATTACGGA-3'.
Moesin siRNA targeted moesin mRNA at 5'-AAAGGTTTCTCCACCTGGCTG-3'.
Fermitin-2 siRNA targeted fermitin-2 mRNA at 5'-CTGGTGGAGAACTCGATGTA-3'.
Delta-catenin siRNA targeted delta-catenin mRNA at 5'-
TGGGAGCTTATTGCTAATTTA-3'. Both ezrin and moesin siRNAs had been validated in HeLa S3 cell line with 80% and 77% knockdown respectively. All siRNAs were generated by Qiagen.

2.3 Protein Extraction, Purification and Analysis

2.3.1 One dimensional SDS polyacrylamide gel electrophoresis

SDS polyacrylamide gel electrophoresis was performed based on the protocol first described by Laemmli (Laemmli 1970) using a Bio-Rad Miniprotean 3 apparatus (Bio-Rad). Gels were cast with a stacking gel (3.75% acrylamide/ bisacrylamide 37.5:1, 0.145 M Tris pH 6.8, 0.01% SDS (w/v), 0.3% TEMED (v/v)) overlaying the resolving gel (10% acrylamide/ bisacrylamide 37.5:1, 0.45 M Tris pH 8.8, 0.1% glycerol (v/v), 0.1% SDS

(w/v), 0.15% TEMED (v/v)). The resolving gel was left to polymerise for at least 30 minutes at room temperature. The buffer chambers were filled with SDS running buffer (25mM Tris, 250 mM glycine, 0.1% SDS). Samples were run at a constant 80 volts through the stacking gel and a constant 150 volts through the resolving gel.

Samples were prepared in Laemmli sample buffer (50mM Tris-Cl pH 6.8, 100mM DTT, 2% SDS, 0.1% bromophenol blue, and 10% glycerol) and were heated to 95 for ten minutes. Cells were then centrifuged at 15000 g for 1 minute at room temperature in a bench top microfuge. Samples were either frozen down or loaded onto the gel with gel loading tips, to maximise accuracy and minimise cross-contamination.

Cultured cells were washed three times with ice cold PBS. When cells were cultured in 6 well, 12 well or 24 well plates, they were lysed directly in with the addition of 5X Laemmli sample buffer. In the case of cells cultured in flasks, cells were scraped and collected into falcoln tubes. Cells were then lysed in 5X Laemmli sample buffer or lysis buffer (PBS, 1% Triton-X 100, Protease inhibitor cocktail (Roche)).

2.3.2 Determination of protein concentration by Bradford Ultra Assay

A dilution series was made using Bovine Serum Albumin (BSA) (Sigma) from 1.5mg.ml⁻¹ to 25µg.ml⁻¹. BSA was dissolved in the same buffer that samples were in. Samples were mixed with Bradford Ultra reagent (Expedeon) at a 1:15 sample:reagent ratio for 0.1mg.ml⁻¹ to 1.5mg.ml⁻¹ range and a 1:1 ratio for the 1µg.ml⁻¹ to 25µg.ml⁻¹ range. Samples, standards and blanks were prepared in triplicates in 96 well plates and absorbance was measured at 595nm using FLUOstar Optima fluorescence plate reader.

2.3.3 Determination of protein concentration by Direct Detect

The Direct Detect FTIR Spectrometer (Merck Millipore) was used to calculate protein concentration. 2 µl of sample solution from a stock solution of 500 µl for was applied to the wells on the Direct Detect cards and inserted into the machine. Results were obtained from the automated display of the IR spectrum and normalising against blank values.

2.3.4 Transfer of proteins to nitrocellulose membranes

Proteins were transferred from gel to nitrocellulose membrane using the Mini Trans-Blot Electrophoretic Transfer Cell (Bio-Rad, Munich, Germany). The nitrocellulose membrane was overlaid onto the gel and placed between two layers of Whatman 3 mm paper and foam pads on either side. The transfer was performed at 4°C at 90 Volts for 60 minutes in transfer buffer (25 mM Tris, 145 mM glycine and 20% methanol). Membranes were stained with Ponceau S solution (Sigma) to detect protein bands, destained with transfer buffer and washed extensively with PBS.

2.3.5 Western blot analysis of proteins on nitrocellulose membrane

Membranes were blocked with 5% Bovine Serum Albumin (BSA) in PBS with 0.05% Triton (PBST) for 1 hour. Primary antibodies were diluted in the blocking buffer at recommended concentrations, and incubated for 1.5 hours at room temperature or overnight at 4°C. Membranes were washed 3x 5 minutes with PBS 0.1% Tween 20. Membranes were then incubated with secondary antibody for 1 hour, washed 6x 5 minutes with PBS 0.1% Tween 20. Protein on the membrane was detected by enhanced chemiluminescence (ECL) by exposure of membrane to Immobilon Western ECL HRP substrate (Millipore) for 5 minutes prior to developing.

2.3.6 Post-nuclear cell lysis

Cell lysis was performed using a mild lysis buffer that consisted of PBS, 1% Triton-X 100 (Sigma Aldrich) and protease inhibitor cocktail (Roche). Cells were subsequently physically homogenised using a Balch homogenizer at 4°C typically using 10 repetitions. Lysates were rotated for 15 minutes on a rotating wheel and then centrifuged for 10 minutes at 14000rpm. The supernatant was the lysate used for further analysis.

2.4 Proteomic Techniques

2.4.1 Collecting and lysing cells for di-glycine peptide enrichment protocol

Cells were cultured in T75 and T 175 flasks. For the first experiment, 2 T75 flasks were grown to confluency, for the second 6 T75 flasks and for the third repetition 6 flasks of T175 flasks were grown to 90% confluency. Cells were washed three times with RPMI and then scraped using a cell scraper in sterile Phosphate-buffered saline (PBS) on ice. Cells were then pelleted in falcon tubes through centrifugation at 500 rpm for 5 minutes.

Cells were resuspended in isolation buffer; Dulbecco's PBS, 1% Triton-X 100. Lysates were prepared as described in 2.3.6. Lysis buffer also consisted of N-ethylmaleimide (10mM). After lysis and aliquot was taken for protein concentration assay by Bradford Ultra Assay or Direct Detect. Proteasome inhibitor MG-132 (10µM) was added to lysate immediately after retrieval of the aliquot.

2.4.2 Extraction of peptides through Methanol-Chloroform precipitation

Methanol-chloroform extraction was used to isolate proteins from solution, particularly when buffers needed to be exchanged. Samples in lysis buffer were split into

200µl aliquots. 600 µl of methanol and 150 µl chloroform was added to each aliquot. Samples were vortexed, 450 µl Milli-Q H₂O was added to each aliquot and then vortexed again. Samples were centrifuged at room temperature for 1 minute at 14000rpm. The upper aqueous phase was removed without disrupting the precipitate and a further 450 µl of methanol was added to each aliquot. After vortex, samples were centrifuged for 2 further minutes and the supernatant was then removed. Samples were left to dry for 10 minutes, after which the samples were resuspended in the desired buffer. Resuspension was aided by vortexing, sonication and homogenisation of the pellet.

2.4.3 In-gel Tryptic digestion

20 µl of lysates were loaded onto a 3.75% SDS PAGE stacking gel and proteins entered the gel by electrophoresis. Subsequently another 20 µl of the same lysate was loaded into the relevant wells, in order to maximised the amount of protein in the gel. Electrophoresis was carried out for 24 minutes at 80 volts until the samples had entered approximately 10mm into the 10% resolving gel, as determined by the leading edge. Electrophoresis was stopped and the bands of the resolving gel which contained the protein, were excised and cut into 1mm cubes. Gel fragments were washed using a wash buffer (50% methanol, 5% acetic acid in Milli-Q H₂O) at room temperature whilst shaking overnight. Gel pieces were rinsed for 2-3 hours followed by the removal of wash solution. 200 µl of acetonitrile is added and gel pieces were dehydrated for 5 minutes at room temperature, twice. 30 µl of 10mM DTT buffer was added and reduced for 30 minutes. 30 µl of 50mM chloroacetamide and alkylate for 30 minutes. 200 µl of acetonitrile was added and gels were dehydrated for 5 minutes, followed by the removal of acetonitrile. Trypsin was prepared on ice as a 20ng/ µl concentration in 50mM ammonium bicarbonate. 30 µl of trypsin added to the gel pieces and rehydrated for 10 minutes. Excess trypsin was

removed and digestion was undertaken overnight at 37°C. 50 µl of 50mM ammonium bicarbonate is added to gel pieces and mixed for 10 minutes. 50 µl of extraction buffer (50% acetonitrile, 5% formic acid in Milli-Q H₂O) was added for 10 minutes followed by a second extraction buffer of (85% acetonitrile, 5% formic acid in Milli-Q H₂O) and subsequently lyophilised. Proteins were resuspended in buffer containing 2% acetonitrile and 0.1% formic acid.

2.4.4 In-solution tryptic digestion

After methanol/chloroform extraction, proteins were resuspended in 100 µl 6M Urea buffer. 5 µl of 200mM DTT was added, vortexed and incubated for 60 minutes. 20 µl of 200mM chloroacetamide alkylated the sample for a further 60 minutes. Then another 20 µl of DTT was added and left to stand at room temperature. Urea concentration was diluted by adding 775 µl of Milli-Q H₂O. Trypsin was added at a 1:50 ratio in relation to total protein as determined by Bradford Ultra assay. Proteins were mixed and digestion was carried out overnight. Double digest was carried out, meaning half of the amount of trypsin was added after the first 2 hours. Reaction was subsequently stopped the next day by decreasing the pH with acetic acid. Peptides were purified using C18 SEP PAK purification.

2.4.5 SEP-PAK purification

SEP-PAK columns (Waters purification) were used. 3 ml of wetting buffer (0.1% Trifluoroacetic acid (TFA) in Acetonitrile) were run through the column twice. 2 ml of equilibration buffer (0.1% TFA + 100% Milli-Q H₂O) was run through the column three times. Then our sample was loaded onto the column and run through slowly twice. 3ml of equilibration buffer was used to wash the columns. Peptides were subsequently eluted of

first with 3 ml elution buffer 1 (0.1% TFA + 50% acetonitrile + 50% Milli-Q H₂O) followed by elution buffer 2 (0.1% TFA + 80% acetonitrile + 20% Milli-Q H₂O). Peptides were subsequently concentrated to a powder using a SpeedVac (Thermo Scientific).

2.4.6 Pull-down of ubiquitinated species using UbiQapture

Ubiquitinated proteins were pull enriched using the protocol from UbiQapture-Q Kit (Enzo Life Sciences). The protocol provided with the kit was followed (http://static.enzolifesciences.com/fileadmin/files/manual/BML-UW8995_insert.pdf). 25 µg of post-nuclear lysate (as determined by Bradford Ultra assay) was added to 40 µl of UbiQapture-Q matrix. For the 3rd repetition, the amounts were doubled to capture more protein. Post-nuclear lysates were incubated with affinity matrix, rotating overnight at 4°C. Samples were centrifuged at 5000g and the supernatant was removed, labelled as the unbound fraction and stored in -20°C. Matrix was washed with 200 µl PBS twice and conjugated proteins were eluted with PBS + 4% SDS (in order to verify protein concentration by Bradford Assay). Eluted fractions were heated at 95°C for 10 minutes and subsequently loaded onto SDS-PAGE for separation of proteins by electrophoresis. Samples for mass spectrometry analysis were loaded into the resolving gel for in-gel digestion as described in section 2.4.2.

2.4.7 Preparation of Di-Glycine antibody columns and immunoprecipitation of di-glycine peptides

Immunoprecipitation and di-glycine antibody columns were prepared based on the Pierce Crosslink Immunoprecipitation Kit protocol (Thermo Scientific - <http://www.piercenet.com/instructions/2162134.pdf>). 20µl of the Pierce Protein A/G plus Agarose slurry was added to Spin Columns. 10, 20 and 25 µg of the di-glycine antibody

(GX41) was coupled to the resin for experiment one, two and three respectively. Antibody was crosslinked to the resin using disuccinimidyl suberate (DSS).

For immunoprecipitation, the digest (peptide output from the trypsin digest) was first pre-cleared as recommended. The amount of digest added to the columns varied based upon the repetition of experiment. For the final two runs approximately 0.75mg and 2.5mg of peptide were added to the columns. The digest was rotated with the immobilised antibody overnight at 4°C. Peptides were eluted into 60 µl elution buffer. Eluted samples were acidified with formic acid and then treated by SEP-PAK purification.

2.4.8 Mass Spectrometry

2.4.8.i Mass Spectrometry Method – Orbitrap

Samples were analysed on an Ultimate 3000 RSLCnano HPLC (Dionex) system run in direct injection mode coupled to a LTQ XL Orbitrap mass spectrometer (Thermo Electron). Samples were resolved on a 15cm by 75 micron inner diameter picotip analytical column (New Objective) which was packed in-house with ProntoSIL 120-3 C18 Ace-EPS phase, 3 µ bead (Bischoff Chromatography). The system was operated at a flow-rate of 300nL min⁻¹. A 120 min gradient was used to separate the peptides. The mass spectrometer was operated in a “Top 5” data dependent acquisition mode. Precursor scans were performed in the orbitrap at a resolving power of 60,000, from which five most intense precursor ions were selected and fragmented in the linear ion trap using CID at a normalised collision energy of 35%. Charge state +1 ions were rejected from isolation for fragmentation. Dynamic exclusion was enabled for 40s. Data were converted from .RAW to .MGF using ProteoWizard

2.4.8.ii Mass Spectrometry Method – Q-Exactive

Samples were analysed on an Ultimate 3000 RSLCnano HPLC (Dionex) system run in direct injection mode coupled to a Q-Exactive Orbitrap mass spectrometer (Thermo Electron). Samples were resolved on a 25cm by 75 micron inner diameter picotip analytical column (New Objective, Woburn, MA, USA) which was packed in-house with ProntoSIL 120-3 C18 Ace-EPS phase, 3 μ bead (Bischoff Chromatography). The system was operated at a flow-rate of 300nL min⁻¹. A 120 min gradient was used to separate the peptides. The mass spectrometer was operated in a “Top 10” data dependent acquisition mode. Precursor scans were performed in the orbitrap at a resolving power of 70,000, from which the ten most intense precursor ions were selected by the quadrupole and fragmented by HCD at a normalised collision energy of 25%. The quadrupole isolation window was set at 3 m/z . Charge state +1 ions and undetermined charge state ions were rejected from selection for fragmentation. Dynamic exclusion was enabled for 40s. Data were converted from .RAW to .MGF using ProteoWizard

2.4.9 Mass Spectrometry Data Analysis

RAW data files were converted to mzXML files using ReAdW (v. 4.2.1) and submitted to the in-house Central Proteomics Facilities Pipeline (CPFP). This implements automatic identification of MS/MS spectra using several search engines, Mascot, X!Tandem and OMSSA, to maximise coverage of the peptide sequence. Reconciliation of the different search parameters, modification specifications and data formats of the search engines were performed automatically. All searches were performed against a concatenated target/decoy database, providing a false-discovery rate that can be compared with the false discovery rate from the results. Results are reported at a target/decoy 1% FDR. Searches were undertaken with a fixed modification of carbamidomethyl cysteine

and variable modifications of oxidised methionine di-glycine tagged lysine and LeuArgGlyGly tagged lysine. Precursor mass tolerance was set to ± 10 ppm and MS/MS fragment ion tolerance at ± 0.5 Da.

2.4.10 Label-free protein quantitation by Normalised Spectral Index (SINQ)

Files were handled through the CPFP and quantitative label free analysis was undertaken through this platform. The method to accurately predict protein abundance was Normalized Spectral Index (SI_N), based on peptide count, spectral count and fragment ion intensity (Griffin et al. 2010). The CPFP allows comparison of relative protein abundance calculated from the SI_N of two samples and is carried out in an automated fashion (Trudgian et al. 2011; Trudgian et al. 2010). Peptide Spectrum Match False Discovery Rates were set at 0.01% and the fragment ion matching tolerance was set at 0.5Da.

2.4.11 Analysis of Mass Spectrometry data by Peaks de novo

Mass spectrometry .mzXML files were uploaded into PEAKS de novo v. 5.3 (Bioinformatics Solutions Inc.). The following settings were established for our search: Parent ion error tolerance set at 10ppm, fragment monoisotopic ion error tolerance at 0.01Da. Fixed modifications were set for carbamidomethylation and variable modifications were set for ubiquitination (GlyGly). Enzyme was specified as trypsin and up to 3 cleavage sites were permitted. The taxa was set as *Homo sapiens* and The FDR was estimated with decoy-fusion and the resultant proteins were filtered by assigned score to give a <5% FDR for peptide matches. Proteins that were carried forward for further analysis and comparison had to have an average localised score greater than 30

(recommended as the threshold in PEAKS user manual) and contain a GlyGly modified lysine (Zhang et al. 2012)

2.4.12 Pathway Analysis

Gene ontology and KEGG pathway analysis was carried out through DAVID EASE Bioinformatics resource (Huang et al. 2009; Kanehisa & Goto 2000; Kanehisa et al. 2012). Analysis was undertaken using the functional annotation tool and the background gene set used was the entire human gene set. For biological process and molecular function gene ontology analysis, results were obtained from the PANTHER subset. A p-value cut-off of 0.05 was implemented as generated by the DAVID EASE software.

2.5 In-Silico Analysis

2.5.1 Data Handling of the MitoCheck time lapse-movies

The movies were made accessible by Dr Jean Karim Hériché from the Ellenberg group in EMBL, Heidelberg. The .avi movie files were downloaded, converted to .mat (matlab compatible) files and stored on the OeRC Supercomputer.

2.5.2 Development of the algorithms and validation

The algorithms to identify cells, track cells and merge tracks were written in Matlab by Dr Boguslaw Obara. The algorithms can be found in the Supplementary Material. In order to validate that the algorithm had effectively managed to track cell movement and speed, 10 movies were randomly selected using a random number generator in Matlab. 3 frames were randomly identified and centroids were manually identified, in the same method in which to identify centroids for threshold and kernel size

optimisation. For centroid detection it was ensured that F-measure values generated were above 0.95. For merging and tracing, 3D plots were generated and examined to ensure that the generated traces were tracked cells effectively without going off course or mis-identifying cells.

2.5.3 Manual identification and tracking of cells

In order to optimise the cell identification and tracking we needed to establish examples of gold standard cell tracking. This was achieved by manually tracking cells. Images were uploaded into GNU Image Manipulation Program (GIMP v.2.6) and all the cells in consecutive or randomly selected frames were manually identified, centroids marked in the centre of the cell on an overlay image with a size 9 marker. The overlays with the marked centroids for these frames were used as templates to establish the efficacy of the algorithm in its ability identify centroids correctly.

2.5.4 Database Annotation

The MitoCheck motility database consists of Ensembl IDs (v.66) which were directly acquired from Dr Jean Karim Hériché. These were linked directly to the relevant time lapse movies. In addition these IDs were used to generate the SwissProt Accession and IDs, gene descriptions and RefSeq IDs through NCBI gene IDs map (<ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/>). KEGG pathway IDs and Gene Ontology IDs were acquired by conversion using an online software Gene ID converter (Alibés et al. 2007). Dr Simon McGowan (CRBG, Oxford) helped with the database organisation and ID conversions.

2.5.5 Gene Set Enrichment Analysis

Gene Set Enrichment Analysis was carried out using the software available from the Broad Institute website (www.broadinstitute.org/gsea/index.jsp). A tab-delimited text file containing the gene list of the output from the MitoCheck motility analysis was ranked based on mean speed and was uploaded. The search was run against ‘all genesets’ and the number of permutations was set to 1000. Chip platforms were left blank (Mootha et al. 2003; Subramanian et al. 2005).

2.5.6 Correction of movies with internal shakes

In order to correct for global movements visible between consequential frames which we believe was caused by the constant level shaking on microscopy stage, a normalized cross-correlation approach has been used implemented by Dr Bogulsaw Obara based on the method described by Haralick et al. (Haralick & Shapiro 1993).

2.6 Migration Assays

2.6.1 Single Cell Motility Assay

For single cell motility experiments, cells were seeded at 20-50% confluency. 24 hours after seeding, cells were washed twice with media. The infrared fluorescent DNA dye DRAQ5 was added to media which had been equilibrated to 37°C and 5% CO₂ in a 1:2000 dilution. The DRAQ5 containing media was added to the cells and incubated at room temperature for 5 minutes. Wells were then topped up with equilibrated media and the plate/dish was sealed in parafilm. Images were acquired using the Zeiss LSM 510 Meta Confocal Microscope at x10 magnification. Images were taken every 2 minutes for 2

hours. Multiple locations were recorded per well and automatically captured using the automated XY stage.

2.6.2 Wound Healing Assay

Cells were seeded at 50% confluency and cultured at 37°C and 5% CO₂ until reaching 90-95% confluency. Cells were then washed twice with media and a scratch was made using a sterile p10 tip. Cells were washed twice in media to remove dead cells from the wound. The dish was then filled with equilibrated media (media at 37°C and 5% CO₂), sealed in parafilm. Images were acquired using the Zeiss LSM 510 Meta confocal microscope, equipped with the Multiple Time Series Macro, Release 4.2. Multiple positions (2-4) were recorded for each well using the macro. Time-lapse phase images were taken at x10 magnification every 10 minutes for 10 to 18 hours, depending on wound closure.

2.6.3 Spreading Assay

Cells were cultured to 80-90% confluency and then trypsinised for the minimal amount of time possible to detach the cells (approximately 5 minutes at 37°C). Cells were resuspended in fresh media and seeded at single cell low density (20-50%) and left to adhere for 20 minutes (at 37°C and 5% CO₂). Wells were filled with equilibrated media and recorded by time-lapse confocal microscopy using the Zeiss LSM 510 confocal microscope under cell culture conditions (37°C and 5% CO₂). Images were taken every 2 minutes for 3 hours to capture cell spreading. Multiple locations were recorded in each experiment.

2.6.4 Immunofluorescence

After wound healing and/or single cell motility assays, wells from 24-well plates were washed twice with PBS fixed with 4% paraformaldehyde in PBS (v/v) for 15 minutes, incubated with 25mM Glycine for 10 minutes and permeabilised with 0.1% Triton X-100 in PBS (v/v) for 10 minutes. Cells were blocked with 1% Fish Skin Gelatin (Sigma) for 30 minutes at room temperature, or left rocking overnight at 4°C. Primary antibody was added at supplier's recommended concentration in 1% Fish Skin Gelatin and cells were incubated for 1 hour at room temperature. Cells were washed 3 times with PBS for 5 minutes. Secondary antibody was added at recommended concentration in 1% Fish Skin Gelatin and incubated on a rocker for 30 minutes at room temperature. Cells were washed again 3 times with PBS, 10 µl of Mowiol with DAPI was added to the cells and coverslips were placed on top of cells in wells from 24-well plate. Cells were left at 4°C overnight before imaging using Zeiss LSM 510 Meta Confocal Microscope.

For imaging cells that have been probed using secondary antibodies conjugated to Cy3 and treated with DAPI, a Helium-Neon Laser (543nm) and an Argon laser (405nm) line were used to excite the respective fluorophores.

2.6.5 Cell tracking in MetaMorph

To track cell speed, .lsm files were uploaded onto MetaMorph (v. 7.5) and converted to 8-bit images. For analysis of single cell motility time lapse movies, the first 60 frames were used and phase images were discarded for analysis. The drop-in app 'Track Objects' was utilised and cells were manually selected to the best approximation in the middle of the cell. Search regions were kept constant between movies. The tracking procedure was observed to ensure there were false positives. Individual cells were labelled

and recorded in order to associate the relevant cells in the spreading and immunofluorescence assays.

2.6.6 Quantitating Immunofluorescence

Immunofluorescence was calculated in ImageJ. The immunofluorescent region of interest was selected using the drawing tools. The area, integrated density and mean grey value were measured. Background values for these metrics were also recorded based on an area with no fluorescence. This was repeated for every relevant cell in the field of view. Finally corrected total cell fluorescence (CTCF) was calculated based on the following equation:

$$\text{CTCF} = \text{Integrated Density} - (\text{Area of cell} \times \text{Mean fluorescence of background readings})$$

2.6.7 Wound Healing Perimeter Analysis

Wound healing perimeter analysis was carried out as described in Coene et al (E D Coene et al. 2011). The MetaMorph journal was developed by Mr Nick White (Cardiff University). The disorder of the leading edge was measure by the perimeter. The journal output generates two values; the mean width of the wound and the combined length of the perimeter (reduced perimeter) for every time point.

For each movie, the wound was orientated to be horizontal. A four-pass filter was applied to enhance the cell edges and a lomo filter was applied to reduce noise and join up the touching cells. Threshold intensities were determined for each frame based on being 3-5 standard deviations above the mean background. Subsequently the wound area was found by searching for a single object greater than 20,000 pixels and this object generated the mean wound width and reduced perimeter as stated in Coene et al. 2011. The

corrected perimeter was calculated for each frame by dividing the reduced perimeter by the length of wound. The absolute difference of the corrected perimeter between frame n and $n+1$ was calculated for every frame. The normalised perimeter variability index is the sum of all the absolute corrected perimeter divided by the number of frames analysed.

Chapter 3 – A proteomic approach to identify candidate BRCA1 E3 ligase substrates

3.1 Introduction

3.1.1 BRCA1 as an E3 ubiquitin ligase

BRCA1 is a protein important in tumour suppression, particularly in breast and ovarian cancer. The only enzymatic activity associated with BRCA1 is its function as an E3 ubiquitin ligase, due to its N-terminal RING domain as discussed in Section 1.4.1.

Several individual BRCA1 E3 substrates have been identified by building upon established interactions and performing in vitro ubiquitination assays, as discussed in Section 1.5. The majority of the identified substrates are canonically nuclear proteins associated with BRCA1's established functions such as transcriptional regulation and DNA damage repair. Substrates that have been identified include CtIP (Yu et al. 2006), RNA polymerase II (Kleiman et al. 2005), RPB8 (Wu et al. 2007) and the estrogen receptor α (Eakin et al. 2007).

Whilst many studies have focused on BRCA1's nuclear functions (Huen et al. 2010; Starita & Parvin 2003) and binding partners, there has yet to be a study to systematically determine the substrates of BRCA1 E3 ligase. This is due to a combination of the complexities of ubiquitination as a post-translational modification and limitations with the techniques available to enrich ubiquitinated species. In addition there is uncertainty over the extent of influence of E3 ubiquitin ligases with respect to how many targets a single E3 ubiquitin ligase can ubiquitinate.

The fundamental pathway of the E1, E2 and E3 proteins in the process of ubiquitination has been extensively researched (see Section 1.3). The number of proteins for each group vary, with there being approximately 10 E1 enzymes, 40 E2 enzymes and over 600 E3 ligases encoded by the human genome with overlapping specificity (Grabbe et

al. 2011). One of the major challenges now is to identify the specific substrates related to particular E3 ubiquitin ligases. Studies that attempt to address this have the fundamental strategy of comparing a system with a functional E3 ligase and a deficient E3 ligase. These systems' ubiquitination profiles are obtained and analysed through two principal steps: the enrichment and identification of ubiquitinated proteins.

3.1.2 Enrichment of ubiquitinated species

When embarking on this project, several techniques were available to successfully enrich ubiquitin-conjugated proteins from complex mixtures for subsequent identification. One such technique involves the expression of 6xHis tagged ubiquitin in mammalian or yeast cells. Proteins that have been covalently modified by this ubiquitin are purified by immobilised metal ion affinity chromatography as first reported by Peng et al. (Peng et al. 2003; Hitchcock et al. 2003). These landmark studies were the first to enrich significant numbers of ubiquitinated proteins, with over 200 ubiquitinated proteins identified.

The second technique involves immobilising antibodies that recognise ubiquitin onto agarose beads. Cell lysates are then incubated with the beads and ubiquitinated proteins are eluted off and identified by mass spectrometry (Vasilescu et al. 2005). This study identified 70 ubiquitinated proteins and had the advantage of not over-expressing modified ubiquitin,

Another similar technique utilises ubiquitin binding domains isolated from E2 and E3 proteins to bind ubiquitinated moieties. Initially, several ubiquitin binding domains were investigated and used to pulldown ubiquitinated species (Raasi et al. 2004). More recently, kits have become available, including the UbiQapture-Q kit, which can pulldown both poly- and mono-ubiquitinated species, and Tandem Repeated ubiquitin binding

entities (TUBEs), which developed to have a higher affinity for poly-ubiquitinated species than individual ubiquitin binding domains (Hjerpe et al. 2009).

These different methods have their advantages and disadvantages and optimum utilisation of the approach is dependent on the aim of the experiment. One important consideration is the nature of the ubiquitinated species that are being enriched. Given that ubiquitination is such a varied, complex modification, enrichment techniques may have different sensitivities depending on the nature of the ubiquitin-linkage, including whether the protein is mono- or poly-ubiquitinated. As discussed in Section 1.3, different types of ubiquitination result in a varied array of downstream consequences. The most well established process is poly-ubiquitination of Lys48 residues, which targets the substrate for proteasomal degradation. Mono-ubiquitination on the other hand is associated with non-proteolytic functions, such as cell signalling, protein stabilisation and trafficking (Chen & Sun 2009). Using TUBES in combination with proteasome inhibitors is a sensible approach if considering the proteolytic implications of an ubiquitin ligase. However, if considering non-proteolytic functions of ubiquitination, the use of antibodies that can identify both poly- and mono-ubiquitinated species becomes advantageous.

Partway through the studies described in this chapter, a novel technique based on isolation of tagged ubiquitination sites in peptides after digestion, rather than whole ubiquitinated proteins, was published (Xu et al. 2010). This was based on a monoclonal antibody, GX41, which recognised the di-glycine tryptic fragment on lysine ubiquitinated residues. This is discussed in more detail in section 3.1.3.ii.

3.1.3 Identification of ubiquitinated species

Several techniques can be used to identify proteins from a list of known or suspected candidates, including Western Blotting or ELISA after enrichment by one of the

methods described above. For de novo identification of unknown targets peptide sequencing by Edman Degradation was initially the only approach. However, the most routinely used current technique for large-scale protein identification is mass spectrometry.

3.1.3.i Mass Spectrometry

A crucial aspect to cell biology is the understanding of protein function, structure, presence and localisation. An important tool in helping us to elucidate some of these aspects is mass spectrometry. At the fundamental level, mass spectrometry can characterize and determine the sequence of proteins. There are 3 fundamental steps involved in determining peptide sequence through mass spectrometry: ionisation of peptides, separation of peptide ions based on their mass to charge ratio (m/z) and the detection of the ions.

There are multiple techniques and technologies available for each of these steps. For the ionisation step, the two major techniques currently used are ESI (Electrospray Ionisation) and MALDI (Matrix-assisted laser desorption/ionisation). MALDI involves mixing the sample with a matrix and the molecules are bombarded with a laser. ESI, which allows for greater sensitivity, involves the mixing of the sample in a polar, volatile solvent which is sprayed out as an aerosol of highly charged droplets. The next step is the separation of peptide ions and Nano-High Performance Liquid Chromatography (HPLC) is a relatively novel development. Nano HPLC coupled with electrospray ionisation has become one of the most widely used due to its increased sensitivity, separation efficiency and the added advantage of minimising dilution of the sample (Mitulovic & Mechtler 2006). In regards to mass analysers, the Orbitrap mass analysers are reported to have high resolution, high dynamic range and mass accuracy, although takes more time to analyse ions (Jedrychowski et al. 2011). More recently, a new mass spectrometer, the Q-

Exactive has improved sensitivity, dynamic range and resolving power enabling identification and quantitation of more peptides (Michalski et al. 2011). Throughout this project, these were the systems we used.

Intact proteins are often too large and complex to effectively analyse by mass spectrometry, so a typical initial step in a bottom-up proteomics study is the proteolytic cleavage of proteins into their constituent peptides. Smaller peptides are easier to ionise and fragment, improving the coverage of the protein. A common method is digesting proteins with trypsin either in-solution or in-gel. Trypsin typically cleaves after lysine and arginine residues (except when followed by a proline or blocked by PTMs), to produce smaller peptide fragments that can be more easily ionised (Olsen et al. 2004). This trypsinisation step has interesting consequences for ubiquitinated proteins which can be utilised to identify ubiquitinated hits. Each peptide produces a mass spectrum based on its fragmentation and is identified by comparing the fragmentation spectra with library of all theoretical spectra, generated by *in silico* digestion.

These tools are invaluable to identifying proteins and more recently has been incorporated and developed to help identify post-translational modifications. However, there are some inherent difficulties with investigating post—translational modifications which are discussed in more detail in the following section.

3.1.3.ii Ubiquitination and Mass Spectrometry

Proteomics of ubiquitinated proteins has been an area of great interest over the past decade and a useful product of trypsinising ubiquitinated proteins has helped identify specific sites of ubiquitination. During the proteomic sample preparation, proteins are trypsinised to generate peptides to increase sequence coverage. The process of trypsinisation on proteins that have an ubiquitin moiety attached leads to a signature di-

Glycine tag on the lysine to which the ubiquitin was bound. Trypsin cannot cleave at the site of modified lysine, such as that of an ubiquitin-conjugated protein. Tryptic digestion of ubiquitin conjugated protein results in a missed cleavage lysine with a di-glycine signature residue covalently bound to the lysine. This di-glycine tag is a remnant signature from the C-terminus of the ubiquitin moiety, which has a characteristic 114.1 Da shift of the lysine. Thus, peptides with a missed cleavage site and a 114.1 Da mass shift on the internal lysine are indicative of a post-translational modification. This can be utilised to determine the sites of ubiquitination. In addition, this mass shift can also be observed on peptides without a missed cleavage site on serine, threonine and cysteine residues, representing non-canonical ubiquitination sites (Shimizu et al. 2010; Cadwell & Coscoy 2005). These modes of ubiquitination may be of biological importance but their physiological significance has yet to be established.

Nonetheless, there are still many hurdles associated with the identification of ubiquitinated proteins. A common difficulty in the identification of any post-translation modified peptide is the limitation in the dynamic range of mass spectrometers. This makes obtaining good coverage of substoichiometric modified peptides difficult. In addition, most post-translational modifications are labile, making them difficult to capture, and risking loss of the PTM during fragmentation. There is the added difficulty with ubiquitination that many ubiquitinated proteins are directed for proteosomal degradation and thus have a high turnover. In addition, ubiquitin modification is highly labile and can be removed from proteins by deubiquitinating enzymes (DUBs). As a result, retrieval of ubiquitinated proteins is a difficult task, which usually returns low yields. Moreover, the proteins that are retrieved may not represent proteins that are highly ubiquitinated due to the high turnover. In order to try to overcome these issues, most studies take steps to enrich for ubiquitinated cells that are treated with DUBs inhibitors, such as N-

ethylmaleimide, and proteasome inhibitors, such as MG-132 (Udeshi et al. 2012; Lopitz-Otsoa et al. 2012; Xu et al. 2010).

Whilst undertaking the project, a novel antibody was generated by the Jaffrey lab (Xu et al. 2010). This antibody can identify the di-glycine signature on peptides left on the site of ubiquitination. We used this antibody to try to help further characterise specific sites of ubiquitination. We prepared affinity columns containing this antibody in order to capture peptides that may have been ubiquitinated. This was an additional tool to the UbiQapture experiments, as one can identify proteins of interest, but also enrich for the sites of ubiquitination. This may uncover further information regarding BRCA1's candidate substrates and the biologically relevant sites of modification.

3.1.4 Aims

BRCA1 ubiquitin ligase activity is a relatively unexplored area, yet one of great clinical relevance. A significant proportion of patients with hereditary BRCA1 mutations that develop cancer have mutations within the N-terminal RING domain. Furthermore, it has recently been shown that BRCA1 may be involved in regulating cell motility (Elisabeth D Coene et al. 2011), which may have some role to play in invasion and metastasis. To try to understand this pathway in greater detail, we have taken an unbiased proteomic approach to identify candidate BRCA1 E3 extra-nuclear substrates.

Whilst BRCA1 has many important functions within the nucleus, including transcriptional regulation, and DNA damage repair, we would like to identify specifically BRCA1's roles and ubiquitination targets in the cytoplasm. The nuclear functions have been studied extensively but emerging roles for BRCA1 at the cell membrane are comparatively unexplored, with promising interactions suggesting links to metastasis. Several cytoplasmic E3 ubiquitin ligases have important roles in tumour suppression

(Satija et al. 2013) and BRCA1 has already been observed in non-nuclear compartments of the cell (Section 1.5.3-6).

BRCA1's ubiquitin ligase function in the nucleus is a contentious issue as a number of nuclear substrates have been identified but the ligase activity has been claimed to be unnecessary for some of these functions. However, studying the nuclear ubiquitination behaviour of BRCA1 was beyond the scope of the work envisaged under this chapter. Many of the approaches we have developed and utilised can be adopted and tailored to further explore nuclear functions of BRCA1.

Our principal two aims are:

- 1) Identify candidate BRCA1 substrates through a comparative label free quantitative proteomic approach.
- 2) Identify specific sites of ubiquitination in potential candidate proteins.

To accomplish this we have implemented a label-free quantitative proteomic approach that was novel at the start of the project. This involved using a kit called UbiQapture, which pulled down ubiquitinated proteins through beads coated with ubiquitin binding domains. Applying this approach to lysates from human breast cancer cells either expressing or lacking functional BRCA1 enabled us to generate candidate BRCA1 target lists. Through examining the differences in the profiles of the protein lists reported and pathway analysis, we have built up a list of candidates. Sites were specifically identified using the di-Glycine antibody.

We hypothesise that there are a subset of proteins that are differentially ubiquitinated between the HCC1937 + WT BRCA1 and HCC1937 + I26A BRCA1 cells and we anticipate a set of proteins will be enriched in the HCC1937 + WT BRCA1 cells

due to direct ubiquitination by BRCA1. We hypothesise that within this group there will be proteins involved in cytoskeletal regulation and signalling, which may underlie and explain the observation that cells without functional BRCA1 have a different motility phenotype to cells with functional BRCA1. We further hypothesise that there will be an overlap of proteins enriched in HCC1937 + WT BRCA1 cells in the two proteomic approaches of pulling down ubiquitinated proteins and peptides with di-glycine signature peptides. These proteins will be our strongest candidates when considering BRCA1 ubiquitination targets. Mining the proteomic results for families enriched in either group of cells may provide further evidence for pathways regulated by BRCA1 by ubiquitination.

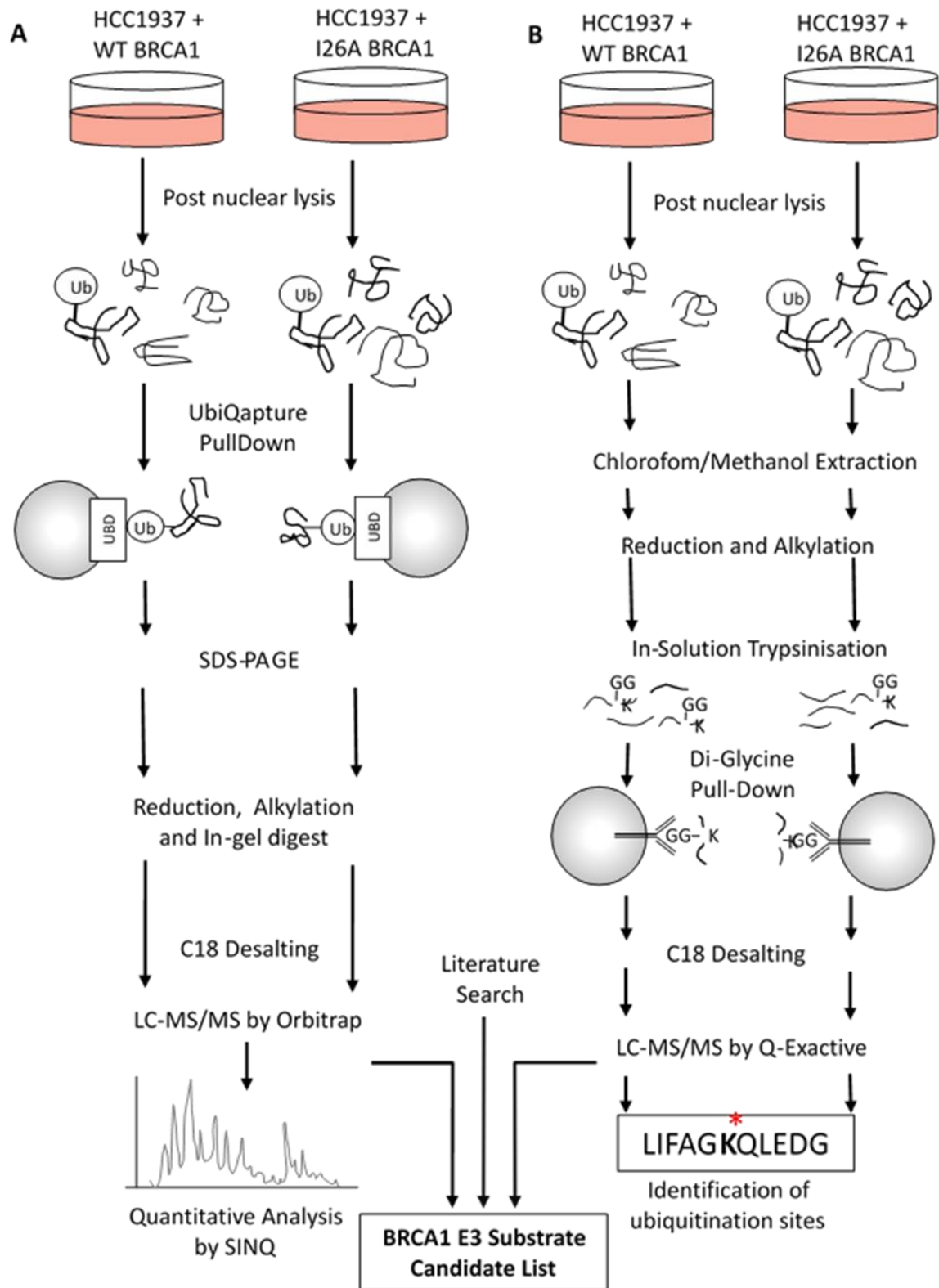


Figure 3.1: Schematic of proteomic approaches for determining candidate BRCA1 E3 ligase substrates: (A) UbiQapture approach to pulldown ubiquitinated species from post nuclear lysates. Quantitative analyses were undertaken by comparing SINQ values. (B) Ubiquitin remnant profiling approach, using an antibody that recognises the di-glycine tryptic signature on ubiquitinated peptides.

3.2 Results

3.2.1 Isolation of ubiquitinated species for UbiQapture Pulldown

3.2.1.i BRCA1 levels in HCC1937 cell lines

To try identify proteins ubiquitinated in BRCA1 dependent manner, we used the HCC1937 breast cancer epithelial cell line which has one BRCA1 null allele and the other allele has a 5382insC mutation (Tomlinson et al. 1998). Cells were transfected with wild type functional BRCA1 (WT) and full length BRCA1 with an isoleucine to alanine point mutation that renders it ubiquitin ligase dead (I26A)(Brzovic et al. 2003b). Schematic of BRCA1 in all 3 cell lines is displayed in Figure 3.1.A.

Throughout the thesis I will use three terms to describe the cell lines used; “HCC1937” or abbreviated to “HCC” represents the untransfected parent breast cancer epithelial cell line from ATCC. “HCC1937 + WT BRCA1” or the abbreviated form “WT” represents the HCC1937 cells transfected with a plasmid stably expressing wild-type functional BRCA1 under drug selection. “HCC1937 + I26A BRCA1” or the abbreviated form “I26A” represents HCC1937 cells transfected with plasmid stably expressing full length BRCA1 with a isoleucine to alanine point mutation at the 26th amino acid residue which has non-functional ubiquitin ligase activity.

Expression levels of BRCA1 in these cell lines was validated by Western Blot (Figure 3.1.B). We see strong bands close to the 220kDa region in the HCC1937 + WT BRCA1 and HCC1937 + I26A BRCA1 lanes, which are likely to correspond to BRCA1. There is a very faint band in the HCC1937 lane, which may correspond to low level of endogenous truncated BRCA1, which cannot bind to BRCA1’s interaction partner BARD1. The lack of this interaction results in a reduction in the rate of ubiquitination. The beta-actin loading control shows confirmed equal loading of protein to the wells.

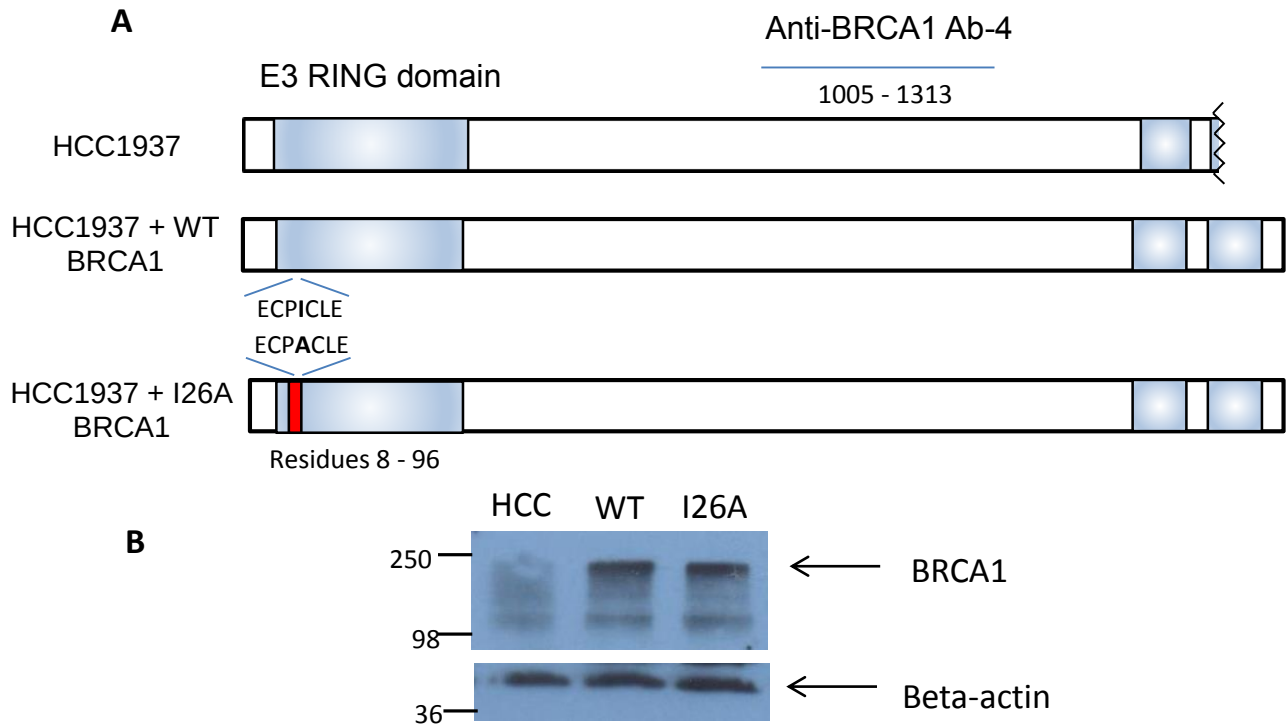


Figure 3.2: Expression of BRCA1 in HCC1937 cell lines: (A) Schematic of the BRCA1 protein. The WT BRCA1 protein, mutant protein I26A, and the isoform endogenous to HCC1937 cells are all shown. (B) Western Blot of whole cell lysates of HCC1937 cells, HCC1937 + WT BRCA1 cells and HCC1937 + I26A BRCA1 cells, probed with anti-BRCA1 Ab-4. The transfected cell lines have a strong signal at the 220kDa level, which is most likely BRCA1. The transfected cell line has minimal signal. The loading control was beta-actin which confirms total protein concentration was similar.

These levels of BRCA1 in the transfected cells are comparable to the signal observed in

MCF-7 cells, an epithelial breast cancer cell line which expresses endogenous BRCA1.

Transfected cell lines were grown under G418 selection, as the plasmid encodes for *neo*

gene. Transfection of plasmid was further validated by treating all three cell lines are

treated with $600\mu\text{g.ml}^{-1}$ G418 and recording cell death by light microscopy after 72 hours.

100% cell death was observed in the HCC1937 untransfected cell line and minimal cell

death in the transfected HCC1937 cell lines. These data both provide evidence for the

HCC1937 cell lines expressing their BRCA1 variants.

3.2.1.ii Optimisation of lysis procedure

In this study, we are specifically trying to identify the extra-nuclear BRCA1 substrates, therefore lysis conditions had to be optimised to prepare post-nuclear lysates. We wanted lysis conditions that maximised retrieval of cytosolic protein fractions and minimised nuclear proteins. We opted for a mild lysis buffer (Section 2.4.1) and used a Balch homogeniser that sheared the cell membrane but leave the nuclei intact. Figure 3.2 shows the optimisation of the number of passages through the shear chamber used to retrieve proteins, but minimise the amount of nuclear protein as measured by the presence of Lamin B. The Coomassie stained gel represents the total amount of protein retrieved from the prepared lysate. The amount of Lamin B detected increases as the number of repetitions that the lysate is pushed through the homogeniser increases. All 4 conditions retrieve substantial amount of protein as detected by Coomassie Blue staining. We decided to use 10 repetitions through the homogeniser, as the amount of lamin B detected was relatively small and the protein retrieval would be maximised by the use of the Balch homogeniser, particularly as observed in the Coomassie stained gel at lower molecular weights.

3.2.1.iii Pulldown of ubiquitinated species using UbiQapture

We used the UbiQapture Q kit to pull down ubiquitinated proteins (Section 2.4.6). From Figure 3.4, one can see the amount of ubiquitinated proteins in our input lysates from the HCC1937 + WT BRCA1 and I26A BRCA1 cell lines are comparable and similar in profile to the control ubiquitin-rich lysate provided with the kit. The signal for the input of I26A is stronger than that of the control and WT, but this can be accounted for by loading differences as shown by the beta-actin loading control. The eluted fractions have an increased signal particularly at a lower molecular weight, indicating the enrichment of

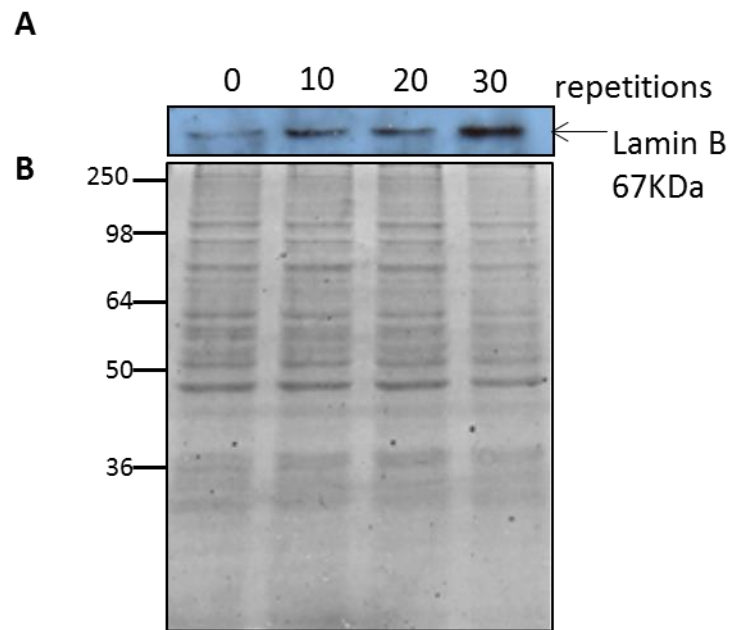


Figure 3.3: Optimisation of lysis protocol: To minimise amount of nuclear material in lysates. (A) Western Blot probing whole cell lysates that have been prepared with varying repetitions in Balsch homogeniser with Anti-Lamin B. (B) Lysates were separated by 10% SDS-PAGE and stained with Imperial Coomassie Reagent

ubiquitinated species has been successful. Moreover, the specificity of the UbiQapture beads was reinforced by the fact that the abundant cytoskeletal beta-actin was not concentrated in the eluted fractions. The unbound fraction has a much weaker signal than the input and eluted fractions, potentially indicating bead saturation and, more importantly, that the majority of ubiquitinated proteins from the post-nuclear lysate have been successfully enriched in the eluted fraction, given the weak signal in the unbound fraction. It should also be noted, this pull-down is likely to include non-ubiquitinated binding factors that are associated with the ubiquitinated proteins, given our gentle lysis procedure.

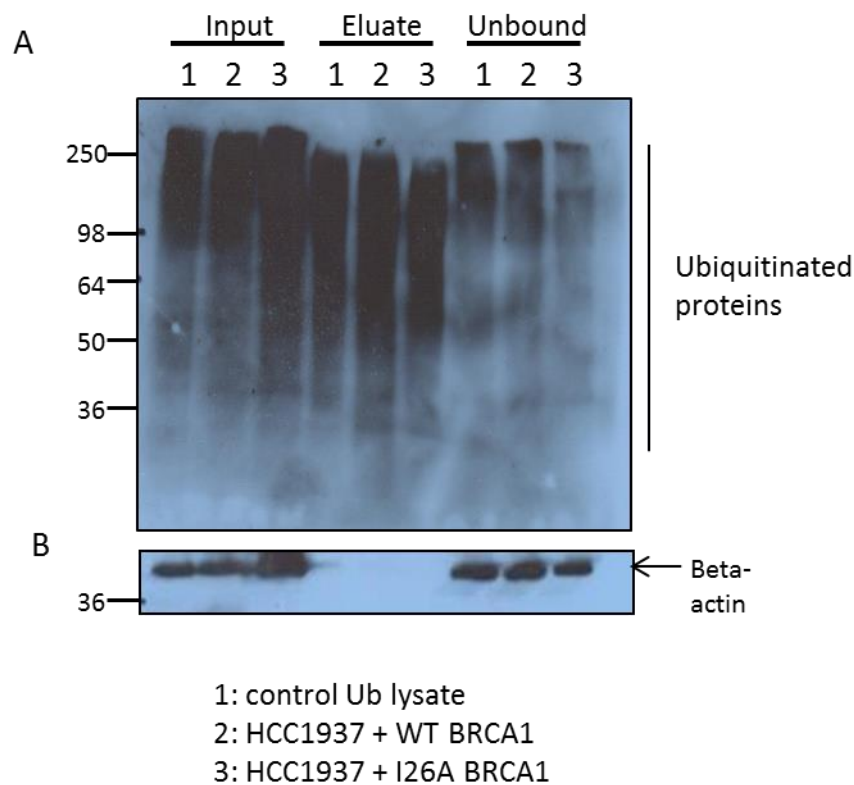


Figure 3.4. Western Blot of Pulldown of ubiquitinated species using UbiQapture: HCC1937, HCC1937 + WT BRCA1 and HCC1937 + I26A BRCA1 post nuclear lysates were incubated with UbiQapture beads which bind ubiquitinated species through ubiquitin binding domains. (A) Western Blot shows the equalised concentrations of the input, eluate and unbound fractions of ubiquitinated proteins. Membrane was probed with primary antibody Anti-ubiquitin P4D1 and secondary Anti-Mouse conjugated to HRP. (B) Loading control, membrane probed with Anti-Actin and secondary Anti-Mouse conjugated to HRP

The pull-down was repeated and eluted samples were run into the top of the resolving gel by SDS-PAGE as described in section 2.4.3, excised, followed by in-gel digestion and submitted for analysis by mass spectrometry. An aliquot of the eluate was taken and analysed by Bradford Ultra Assay to ensure protein concentrations were similar. In total, three biological replicates of this experiment were analysed. With subsequent experiments, minor alterations to the protocol were made. Specifically, starting material was increased two-fold. In addition, the amount of UbiQapture beads used was doubled.

Samples were eluted with PBS + 4% SDS as opposed to the suggested sample buffer to avoid interference with the Bradford Ultra Assay.

3.2.2. Quantitative Proteomics on ubiquitin-conjugated proteins

3.2.2.i Identification of proteins

The digested samples were run on the LTQ Orbitrap Mass Spectrometer as described in Section 2.4.8.i. The chromatograms (not shown) for each run demonstrated that the samples were abundant with protein, with improved spectra with each iteration of the experiment, most likely due to improved enrichment strategies.

The proteomic experiments were successful in identifying a large number of proteins from the UbiQapture pulldowns, ranging between samples and experiments from 155 to 526 proteins. As visualised in Figure 3.5.A (Venn Diagrams), there was a trend for an increase in the number of proteins reported for each iteration of the experiment. A significant proportion of the protein hits were reproducibly observed in each experiment for WT and I26A hits, with 127 proteins found in each WT run and 109 proteins identified in all three I26A runs. A 0.01 q-value threshold was applied to all the datasets that corresponded to a 1% False Discovery Rate. At least 2 unique peptides needed to be identified to correspond to a protein hit. Note that in this approach there is no enrichment for ubiquitin modified peptides as the selection operates on the intact protein prior to fragmentation. Label-free absolute quantitation was carried out on the data through the Central Proteomics Facility Pipeline as described in section 2.4.10. The method used was normalised spectral index quantification (SINQ). The histogram in Figure 3.5 shows the distribution of the fold change values based on the raw SI_N values for WT and I26A as calculated by the Pipeline (not including unique hits in WT or I26A). For all 3 experiments, the data exhibit a normal distribution with 70% of the fold change values

falling between $\log_2$ fold change WT:I26A of -1.5 to 1.5 and 90% of the hits falling between a $\log_2$ fold change value of -2 and 2. This is encouraging, as one would expect the majority of the protein hits to centre around a $\log_2$ fold change value of 0 – which equates to a 1:1 ratio. This would be consistent with the fact that the vast majority of ubiquitin ligases are present in both cell types and we are only perturbing the behaviour of one. Furthermore, it confirms the results from the protein quantitation Bradford assays (not shown), that the overall protein concentration was relatively similar (within 10% of each other). There is a slight skew in the distribution towards the enrichment of proteins in the I26A sample, with experiment 1 and 2 having its normal distribution centred on $\log_2$ fold change of -1. For experiment 3, the centre is closer to a $\log_2$ fold change value of 0. This is reinforced by Figure 3.3 C.

Whilst there are subtle differences between the repetitions, these figures demonstrate that all 3 experiments have similar overall profiles in their fold change distribution, justifying aggregation of data during further analysis. In all three experiments the histograms and plot in Figure 3.3 show similar overall trends but this view of the data does not indicate which protein are specifically enriched and if the individual hits are reproducibly enriched. This requires closer inspection of the data.

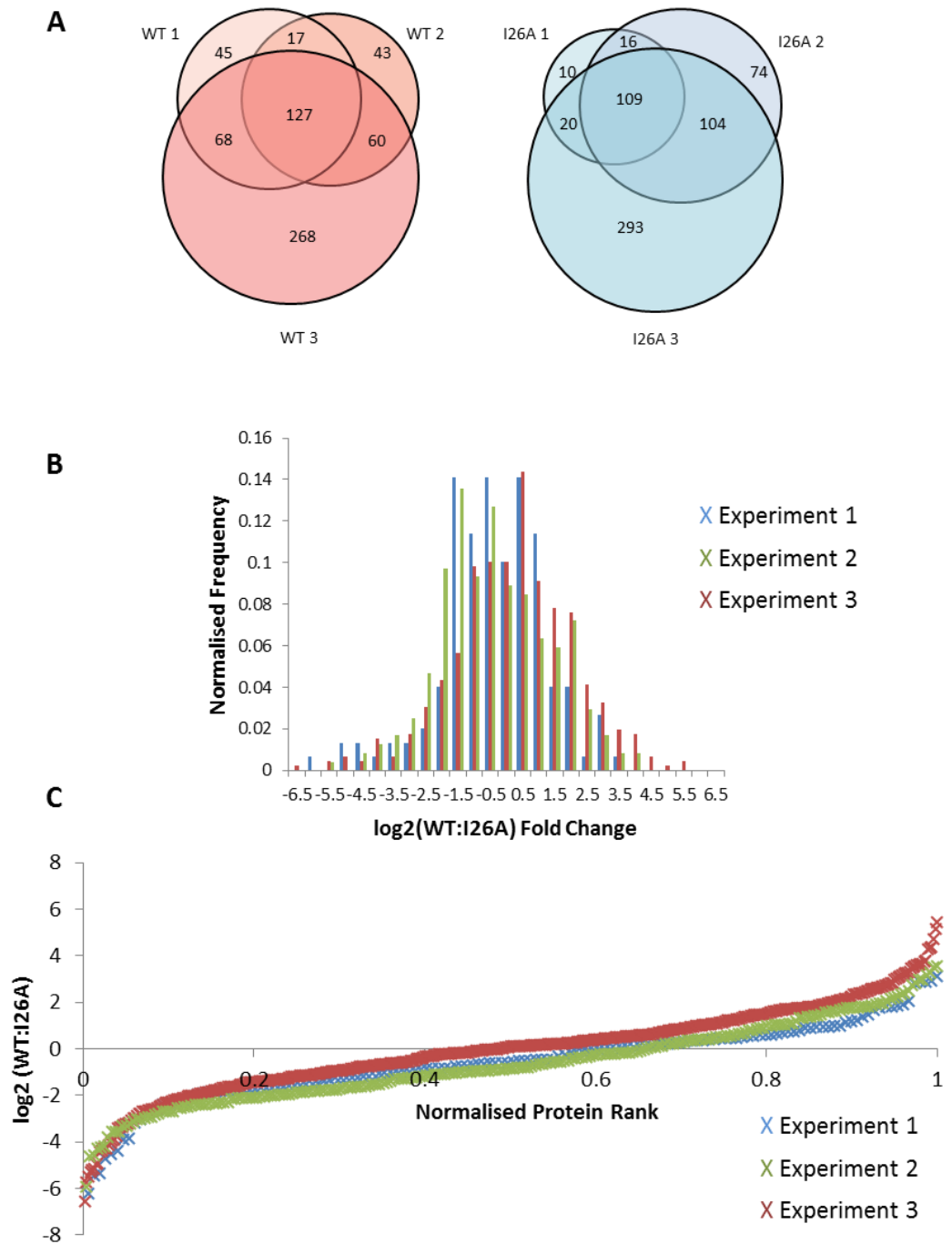


Figure 3.5. Results from the UbiQapture pulldown: (A) Venn Diagrams representing the common protein hits between the 3 repeats of the experiments for both WT and I26A samples. (B) Histogram showing the distribution of the $\log_2(\text{WT:I26A})$ fold change from all 3 experiments. Fold change values were calculated on the basis of the Normalised Spectral Index generated from the CFP Pipeline. The normalised frequency represents the number of proteins over the total number of proteins observed in said experiment. The majority of $\log_2$ fold change values fall between -1 and +1, representing a modest fold change showing the majority of proteins are likely to be at similar stoichiometric levels in cells expressing functional BRCA1 and I26A BRCA1.

3.2.2.ii Categorisation of quantitative mass spectrometry results

To assess the hits that are reproducibly enriched in either WT or I26A and visualise the data more clearly, we have sorted the data into categories. The inherent nature of proteomic data is that there is a substantial amount of variability and this is exaggerated further when trying to identify post translational modifications (Kim et al. 2011; Lopitz-Otsoa et al. 2012). Thus it is difficult to get consistent fold change values for every protein in all three repeats.

In these variable experiments, there is more confidence in a protein hit based upon the number of times it appears in the repeats and the consistency of the relative enrichment. We therefore established nine categories and each protein is placed in one, based on their consistent reproducibility. The first two categories are WT 3:0 and I26A 3:0. These represent proteins that have the highest confidence as they are observed in all three repeats and consistently enriched in either WT ($\log_2$ fold change >0) or I26A ($\log_2$ fold change <0). The next two categories are WT 2:0 and I26A 2:0, representing proteins that appear in two experimental repeats with a consistent enrichment and then WT 1:0 and I26A 1:0, where the protein is enriched in only one repeat in either WT or I26A. The next categories WT 2:1 and I26A 2:1 represent hits that have inconsistent enrichment profiles but are identified in all three repeats. For example, protein X is enriched in WT on two occasions and I26A in one repeat. This is labelled WT 2:1 and I26A 2:1. The final subset represent hits observed in two experiments, once enriched in WT and once in I26A, labelled 1:1. Every protein has subsequently been categorised and plotted based on their category and the mean $\log_2$ fold change in Figure 3.6.

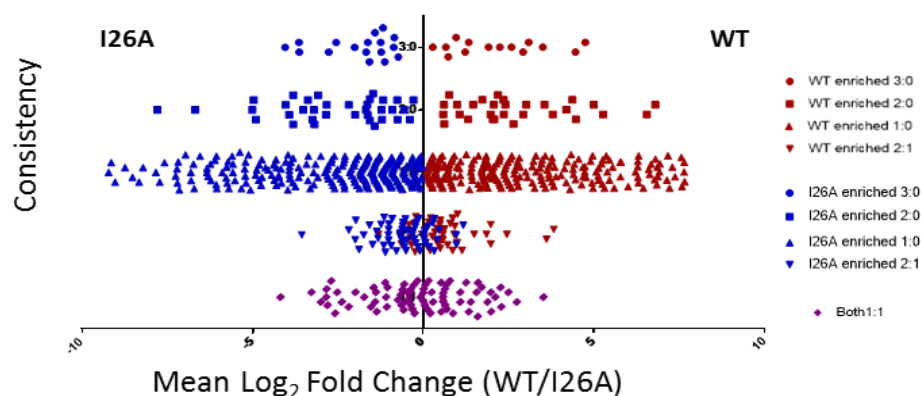


Figure 3.6 - Plot of log₂ Fold Change hits from amalgamated UbiQapture data sorted on their reproducibility: The categories distinguish the consistency of the enrichment of protein hits. The sum the two numbers in the category represent how many times the protein was identified in the 3 repeats. 3:0 represents proteins that have been observed in all 3 repeats, consistently enriched in WT or I26A. 2:0 represents proteins that have appeared in 2 of the experiments and consistently enriched in WT or I26A. 2:1 represents proteins that have been observed in all 3 experiments, but enriched twice in WT and once in I26A (or vice versa). 1:1 represents proteins that have been identified in 2 experiments

Within these lists there is a subset of proteins that are uniquely identified in a WT or I26A for each repeat. For these hits it is not possible to establish a numerical fold change value based on the normalized spectral index, due to a missing value. However, for data handling purposes, it is still useful to include and assign a numerical log₂ fold change to these hits. For each experiment, the minimum observed Normalised Spectral Index value for that repeat was used as a substitute for the zero value assigned to the unobserved proteins to give a conservative consistent fold change value. This is equivalent to assuming that the protein is present in the sample with the missing value, but at a level just below the significant detection threshold. The rationale for this conservative estimation method is explained further in Section 3.3.1.ii. This allows us to plot every single observed protein on Figure 3.6.

This plot reveals a number of noteworthy points about the data. Firstly, the number of consistently enriched hits in all three experiments is quite modest with up to 29 proteins enriched consistently in either WT or I26A in all three experiments. The number of proteins increase as one goes down the categories to 2:0 and 1:0 respectively, with a greater spread regarding $\log_2$ fold change values. The large number of hits in the 1:0 category represent proteins that have only been identified in one experiment, a large proportion of which correlate with the unique hits found experiment 3 as shown by the Venn diagrams in Figure 3.4.A. The $\log_2$ fold change values here range from -9 to 8. The larger fold change values are most often generated by hits which missing values had imputed in order to calculate a fold change. The majority of the hits in this category cluster between a $\log_2$ fold change of -2 and +2 and have an overall range of -4 to +4. This indicates a large proportion of hits which are probably at relatively similar levels within the cell but are harder to pick up. In the next category, 2:1, the spread of the $\log_2$ fold change values are reduced and apart from a few outliers are clustered around a $\log_2$ fold change of zero. Interestingly, there are instances of red hits, which are enriched twice in WT and once in I26A, which has an average $\log_2$ fold change value below 0 and vice versa. In these instances, the large $\log_2$ fold change value when the protein is enriched in I26A outweighs the relatively modest $\log_2$ fold change values when enriched in WT on two occasions. This is indicative of the variable nature of large-scale proteomic experiment. Similarly proteins in the 1:1 category cluster around zero.

3.2.2.iii Highly reproducible hits from UbiQapture Pull-Down

Using the categorisation plotted in Figure 3.6, we visually represent in a quantitatively accurate manner the size of the categories in Figure 3.7.A and establish a list of proteins that fall into the WT 3:0 (Figure 3.7.B) and I26A 3:0 (Figure 3.7.C) categories.

These are the most confident hits and a number of them appear to be of functional interest. Enriched in the cells expressing functional BRCA1 are a number of ribosomal proteins from both the small 40S subunit and large 60S subunit, with varying $\log_2$ fold change values from 0.26 to 4.45. One of the proteins with the highest average $\log_2$ fold change value is the cytoskeletal related ERM family protein moesin with an average $\log_2$ fold change value of 2.91.

Proteins that appear enriched in the I26A BRCA1 transfected cells consistently in all three experiments include a number of cytoskeletal related proteins such as vimentin, beta-tubulin and cytoplasmic actin. A number of mitochondria-related proteins are also enriched, such as the ATPase family AAA domain-containing protein, the mitochondrial phosphate carrier protein and the tricarboxylate transport protein with $\log_2$ fold change values of -2.76, -4.04 and -2.54 respectively.

Whilst inspection of comparative enrichment lists reveals functionally plausible candidates (see Section 1.5.5-6), the complete dataset contains far more information than can be assessed by manual inspection. Accordingly, we then turned to systemic analysis for enrichment of multiple targets in previously defined functional pathways.

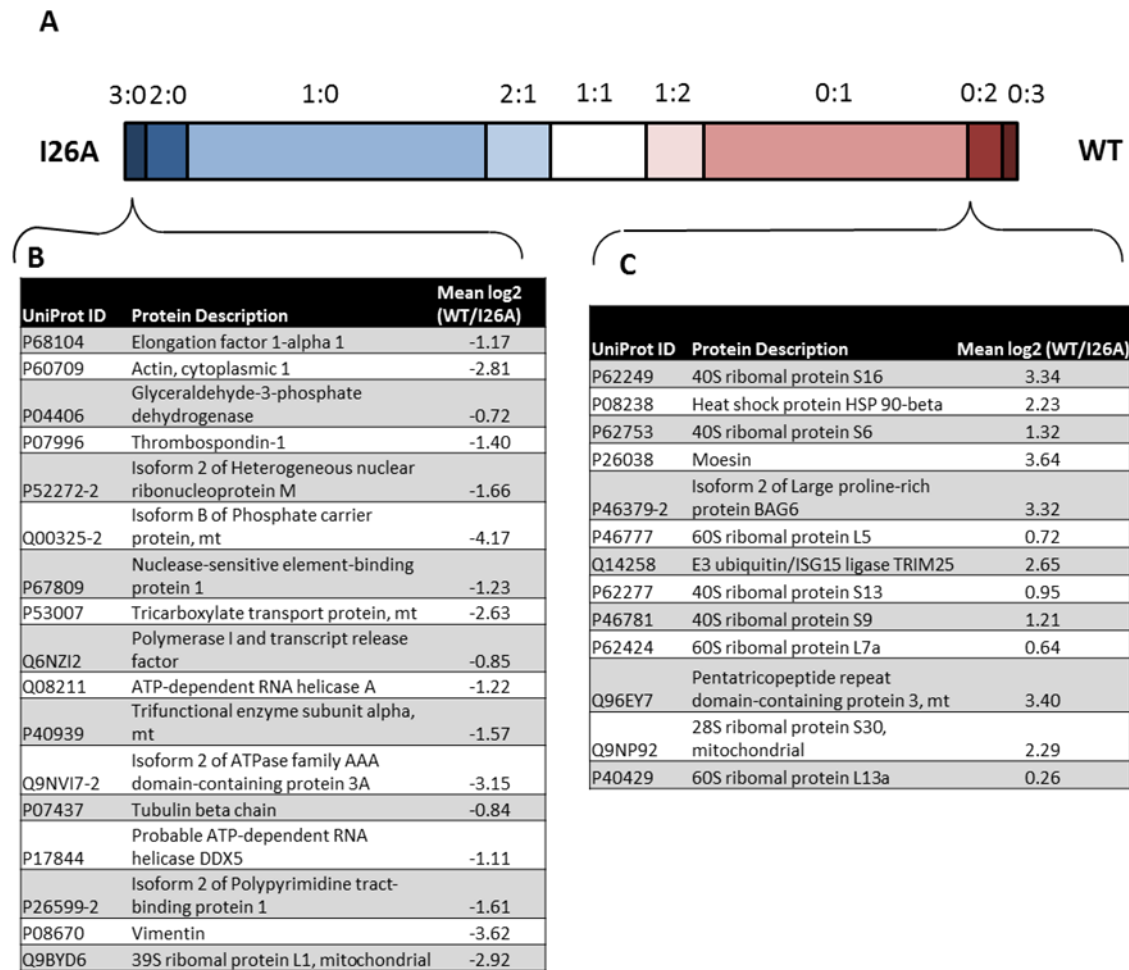


Figure 3.7. Tables of proteins most consistently enriched in WT and I26A from UbiQapture Experiment: (A) Infographic proportionally representing number of proteins associated in each consistency category as explained in Figure 3.6. Blue represents proteins that were enriched in I26A and red represents proteins that were enriched in WT. (B) Table of proteins identified by mass spectrometry from UbiQapture Pulldown which are in the top category for consistency in I26A with their associated mean log₂ fold change. (C) Table of proteins identified by mass spectrometry from UbiQapture Pulldown which are in the top category for consistency in WT with their associated mean log₂ fold change.

3.2.3 Pathway Analysis of UbiQapture Pull-down results

3.2.3.i Gene Ontology Results

The Gene Ontology and Pathway analysis reveals a number of interesting features in our proteomic dataset. Two lists were assembled, WT and I26A. The WT list consisted of all proteins that had a log₂ fold change value greater than 0 and the I26A list had hits

with a $\log_2$ fold change value less than 0. An alternative option that was tested involved generating two lists based on all the hits observed in any of the repeats for WT and I26A. The majority of proteins that we detect from our proteomic experiments appear in at least one of the WT and I26A repeats and, as the gene ontology enrichment tool DAVID EASE (Huang et al. 2009) does not account for fold change or enrichment, the output showed that the gene ontology and pathways profiles were very similar. We decided to mine the data further, generating the lists based upon their relative enrichment of the proteins. Categories were only accepted on the basis of having a p-value less than 0.05.

The Biological Processing Gene Ontology Analysis produces 17 categories which are relatively enriched. Figure 3.8.A shows a comparable distribution of proteins between WT and I26A in the majority of the biological process categories, with the exception of mRNA splicing and processing enriched in I26A relative to WT. Similarly protein biosynthesis and cell structure and motility are two categories that are enriched in WT.

From the Molecular Function Gene Ontology Analysis, we see 20 categories that are shown to be enriched in our cell lines with variable numbers of proteins in each category for WT and I26A. A large number of proteins in our sample are part of a ribosomal complex and associated with nucleic acid binding (with 143 and 148 proteins in this category in WT and I26A respectively – not shown in graph). Similarly, there are a number of proteins enriched in mRNA processing. We see some clear differences between WT and I26A, with mitochondrial and transfer carrier proteins enriched in the I26A cells but not in WT. Conversely, we see a significant enrichment of ribosomal proteins and cytoskeletal proteins within the cells expressing WT BRCA1.

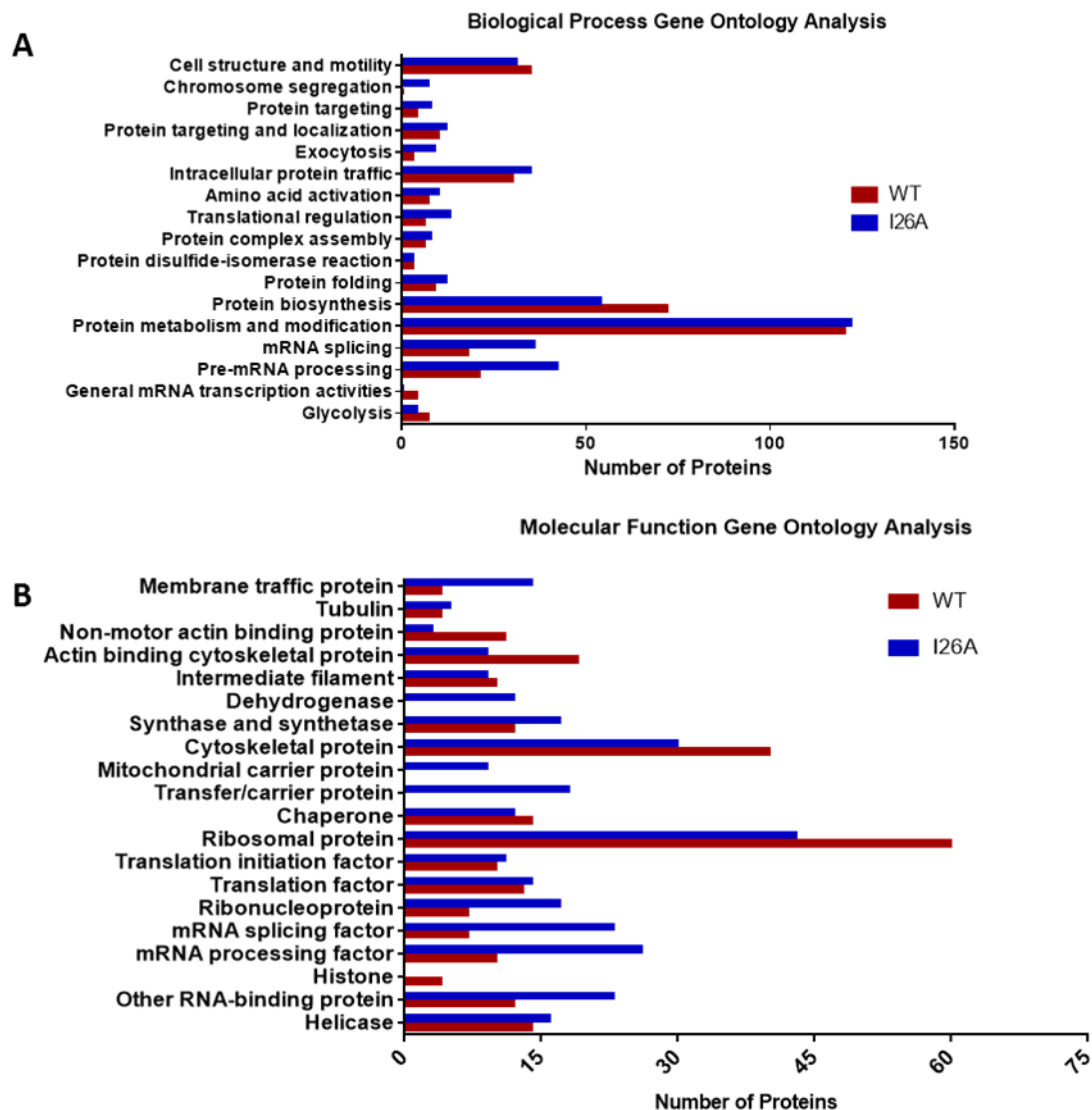


Figure 3.8. Gene Ontology and Pathway Analysis Results (A) Comparison of enriched proteins in WT and I26A Gene Ontology Biological Processing Panther categories based on DAVID EASE analysis. (B) Comparison of enriched proteins in WT and I26A Gene Ontology Molecular Function Panther categories based on DAVID EASE analysis. Hits were included if found to be enriched by DAVID EASE functional annotation and if the associated p-value < 0.05.

3.2.3.ii KEGG Pathway Analysis

The KEGG pathway analysis represented in Figure 3.9, as carried out as discussed in Section 2.4.12 is consistent with and reinforces the differences observed by Molecular Function GO as there is a WT enrichment of proteins involved with the ribosome, adherens junction and transendothelial migration. Conversely, protein export, antigen processing, proteasome associated proteins and lysine degradation are KEGG categories enriched in the I26A cells. Interestingly, we see ribosomal proteins KEGG category associated with both WT and I26A, however, there are nearly double the number of ribosomal proteins observed in WT compared to I26A and three times as many hits under the spliceosome category in I26A compared to WT. Figure 3.10 shows the specific KEGG pathway for Adherens Junction, with the highlighted hits enriched in WT. This is a representative example of the differences in enrichment and this pathway is of particular interest given its links to cell motility.

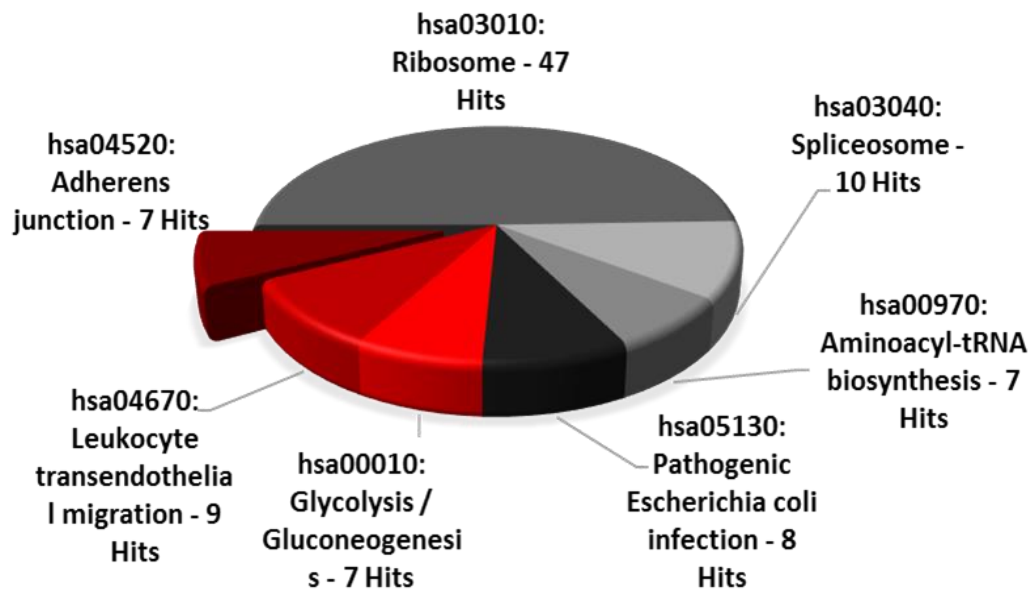
3.2.4 Ubiquitin Remnant Pulldown

In order to assess the candidates revealed in the study described above for their involvement in BRCA1-regulated control of cell migration we would like to know the sites of ubiquitination that depend on BRCA1. In order to do this, we turned to an experimental approach that identifies the site of ubiquitination as well as the identity of the protein targets.

3.2.4.i Method Development of GX41 Protocol

The pull-down of di-glycine modified peptides protocol was optimised and altered with each iteration of the experiment, thus the 3 repeats were not technical repeats. The first iteration involved an in-gel digestion approach followed by the di-glycine pulldown. This was subsequently altered to accommodate an in-solution digestion protocol, to

A KEGG PATHWAY WT ENRICHMENT



B KEGG PATHWAY I26A ENRICHMENT

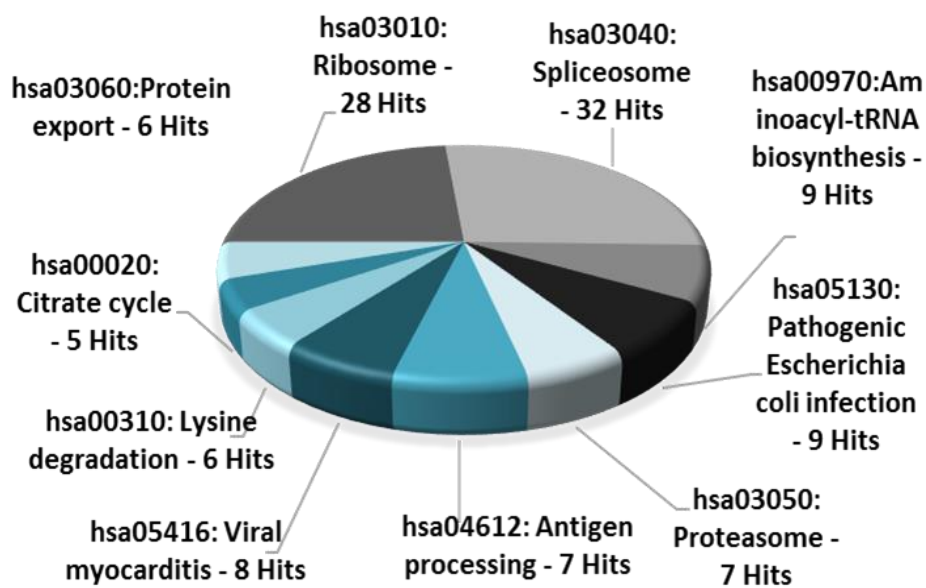


Figure 3.9. KEGG Pathway Analysis Results (A) KEGG Pathway analysis of proteins enriched in WT (B) KEGG Pathway analysis of proteins enriched in I26A.

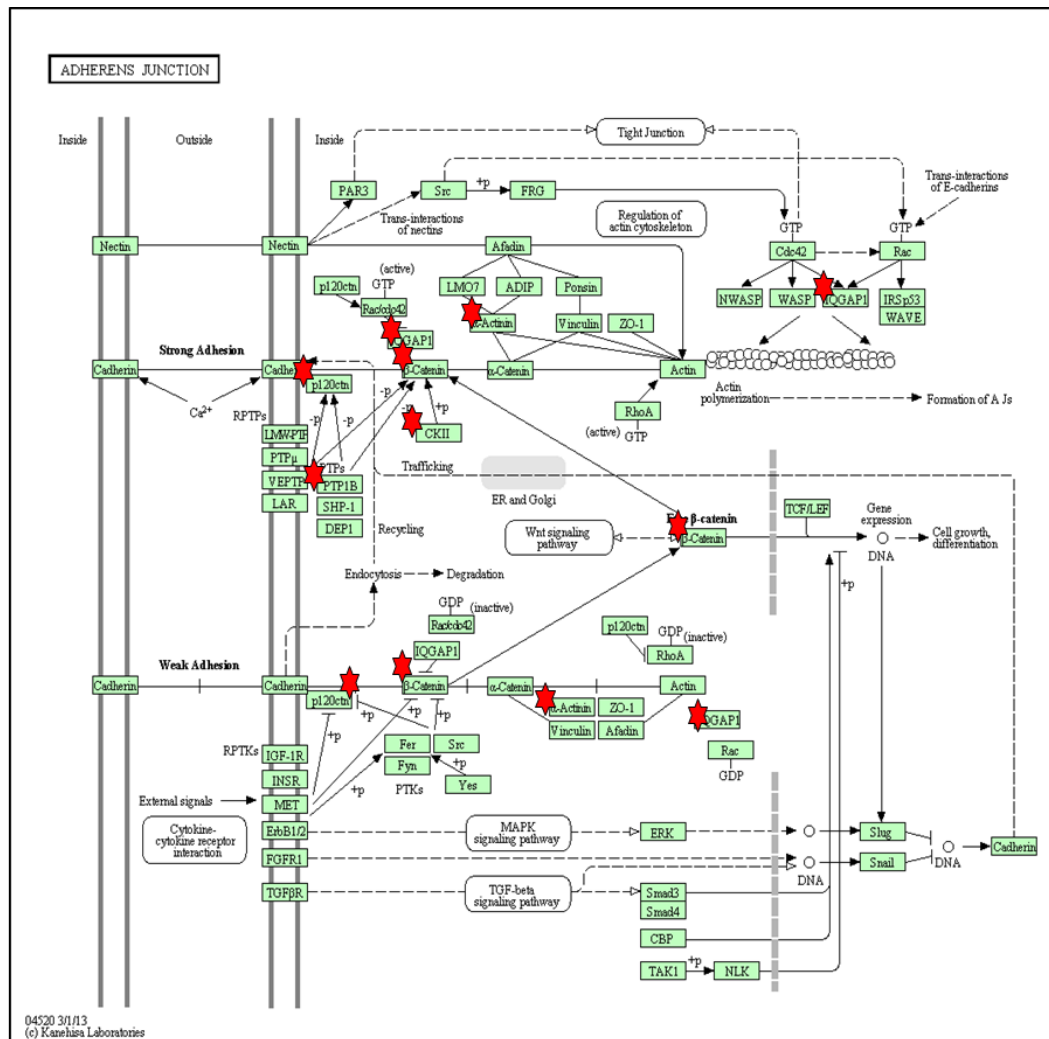


Figure 3.10. KEGG Pathway Map for Adherens Junction: A representative KEGG pathway map from our enriched pathways in the WT cell line. This map is of the Adherens Junction, with all the WT enriched proteins highlighted with a red star. Hits were included if found to be enriched by DAVID EASE functional annotation and if the associated p-value < 0.05.

increase the amount of post nuclear lysate that would undergo trypsinisation in order to maximise the retrieval of di-glycine modified peptides. The total amount of lysate we were able to process through the in-gel digestion was 80µl (16% of the cell lysate) due to the loading constraints of the gels.

To maximise the retrieval of di-glycine peptides, we took several steps. Firstly, we increased the starting material to 5mg post-nuclear lysate; secondly, we performed an in-solution digest to ensure digestion of all of the lysate, and not merely a proportion of it,

due to the constraints of the gel. We also increased the amount of GX41 antibody utilised in our columns to try increase the amount of peptide. Whilst we did not grow cells in the presence of proteasome inhibitors or DUB inhibitors, we also added MG-132, a proteasome inhibitor and N-ethylmaleimide (NEM), a DUB inhibitor to our lysis buffer, to reduce degradation of ubiquitinated proteins and removal of ubiquitin from substrate proteins during lysis.

To integrate the in-solution digest we had to perform the digest in 8M urea, which required precipitating the post-nuclear lysate so that the digestion occurred in a detergent-free buffer. All precipitation and re-solubilisation techniques incur some loss of protein, but the chloroform-methanol extraction was deemed to be the best protocol for minimising protein loss (Fic et al. 2010). Thus after the post-nuclear lysis, a chloroform-methanol extraction approach was implemented as described in Chapter 2.4.2

The final GX41 protocol had multiple complicated steps; thus several intermediate validation steps were taken to ensure good digestion and retention of peptide. To ensure complete digestion, a 20ul aliquot was run on the Q-Exactive for one hour and analysed by RawMeat to measure the distribution of charged species (for repeats 2 and 3), as described in Chapter 2.4.8.ii and 2.4.9. Figure 3.11.A is the output from RawMeat for the second GX41 WT run and demonstrates a good digestion, as the majority of the ions are 2+ and 3+ ions, whereas an incomplete digestion would have a greater proportion of highly charged ions.

Similarly, there are multiple extraction and lyophilisation steps, which invariably reduce protein yield. To calculate loss of protein we quantitated protein and peptide concentrations of input and eluate by Direct Detect. This IR spectroscopy method can also evaluate the complexity and cleanliness of the sample through detection of carbohydrate

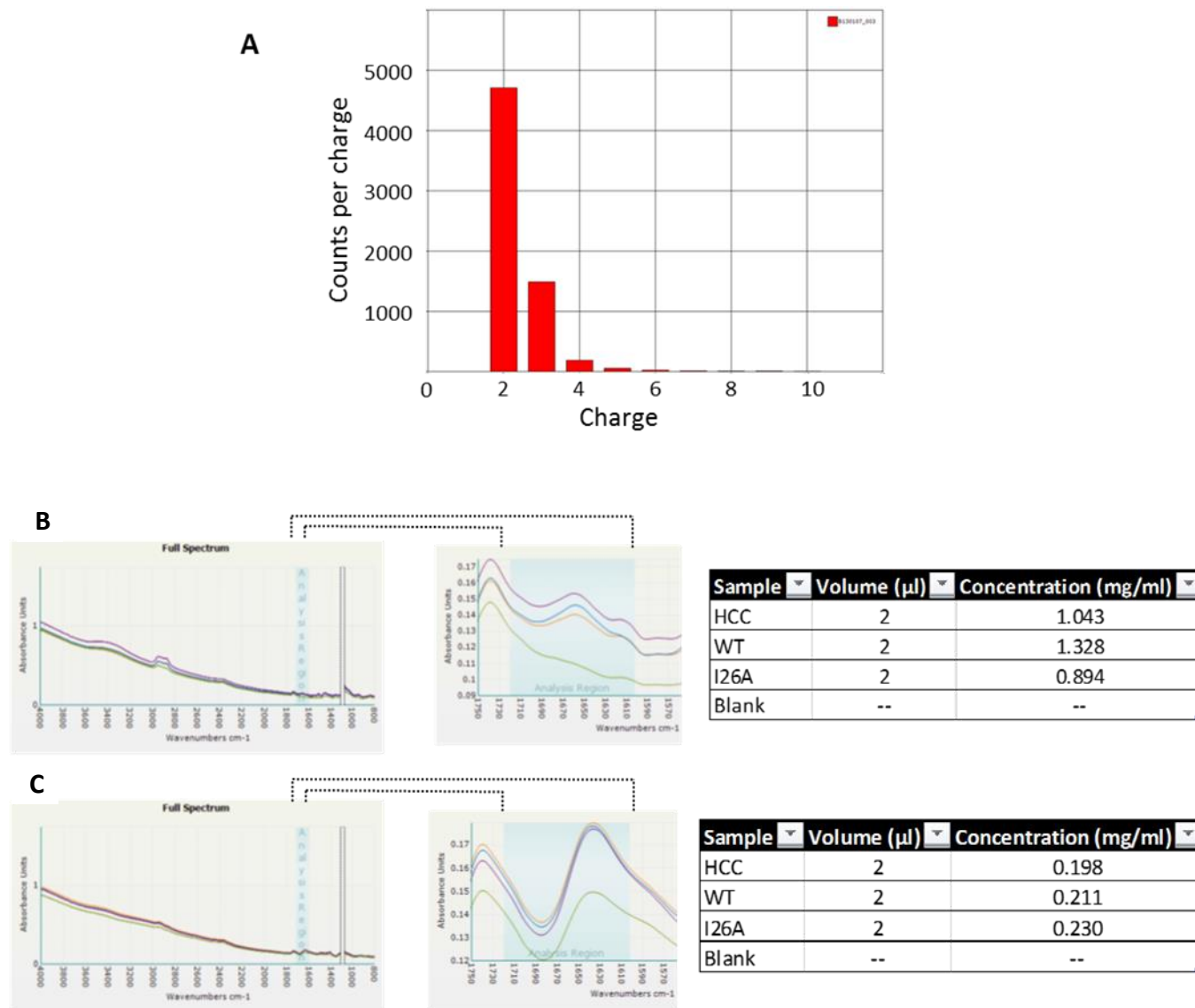


Figure 3.11. Method Development of GX41 Pull down (A) To verify good digestion which is vital for the generation of di-glycine residues on ubiquitination sites, we ran an aliquot of the sample on the Q-Exactive and analysed by RawMeat, which identifies the relative number of charged species in the sample. For good digestion the majority of peptides should have a charge of 2+ or 3+ as demonstrated here which was the WT sample of experiment 2. (B) Protein Quantitation carried out using Millipore Direct Detect to establish relative amount of protein/peptide in input and eluate after GX41 Pulldown. First graph shows the full spectrum of the whole cell lysate with the second graph showing the protein peak analysis region. The table shows the concentration for whole cell lysate. (C) Graph showing the spectrum for the GX41 eluted fraction, with full spectrum, protein peak analysis region and tabulated results. The IR spectra also allow to determine how crude the sample is. Here we see that the eluted species have a much tighter profile and protein concentration, which is also cleaner as demonstrated by the fewer peaks.

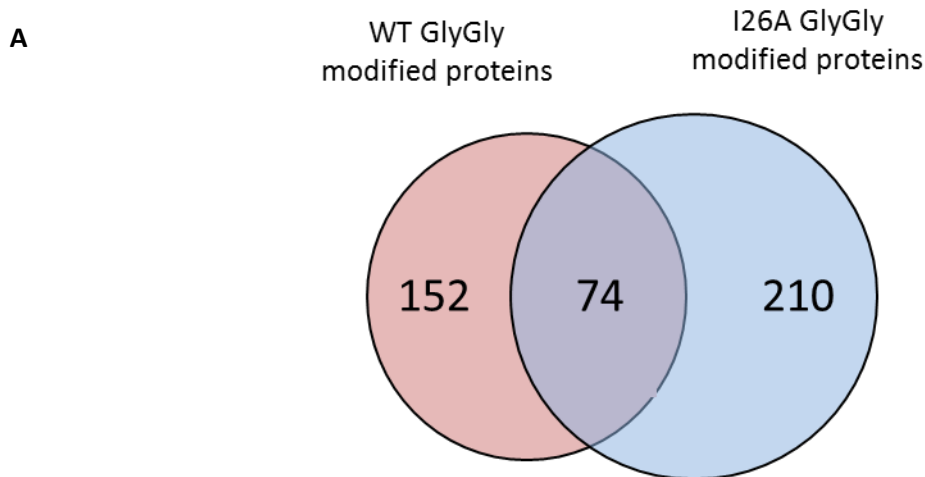
and lipid. Figure 3.11 B shows that the eluate has a concentration of approximately 0.2mg.ml^{-1} compared to the input which has approximately 1mg.ml^{-1} . Moreover, we see our eluted sample is very clean relative to the input, with minimal carbohydrate or lipid contamination, verifying the extraction and pulldown are specifically enriching peptides.

Dot blots were also carried out to assess the relative amount of di-glycine peptides in input relative to eluate and input (not shown). However, the results were inconclusive, due to antibody leakage from the columns. This ultimately masked the true signal of di-glycine peptides, as the secondary antibody detected the leaked GX41 antibody. However, the output from the mass spectrometry results verified the enrichment of di-glycine through comparison of the number of hits identified in these experiments relative to the UbiQapture experiments.

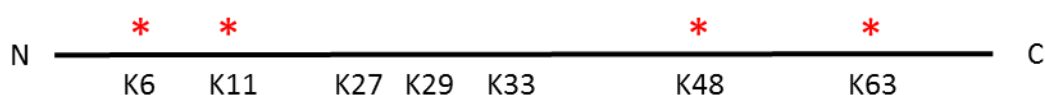
3.2.4.ii Identification of di-Glycine modified peptides

The GX41 Pulldown resulted in the enrichment of many di-glycine peptides. With each iteration of the experiment, the number of di-glycine peptides identified significantly increased due to changes made to the protocol significantly improved the retrieval of gly-gly peptides. Peptides were identified through both Mascot and Pipeline. The .RAW data was also converted to .mzXML files and analysed by de novo sequencing by PEAKS de novo, to capture any peptides that may have been missed by Mascot and Pipeline.

Figure 3.11 shows the amalgamation of all three experiments and the relative number of di-glycine peptides identified in WT and I26A by Mascot and Peaks. We see a total of 226 and 284 di-Glycine modified proteins identified in WT and I26A respectively, with 74 proteins present in both. The total number of peptides is even greater, with some proteins having multiple GlyGly modified peptides. Mascot hits were



B Ubiquitination sites Detected on Ubiquitin



C

UniProt ID	Protein Description	WT Peptide	I26A Peptide
P07355	Annexin A2	K*LMVALAKGRR	TK*GVDEVTIVNLTNR
P07900	Heat shock protein 90kDa α (cytosolic)	VILHLK*EDQTEYLEERR	KK*LSELLR
P11142	Heat shock 70kDa protein 8	ATVEDEK*	LQGK*INDEDK
		NQVAMNPTNTVFDK*RLIGR	ILDK*CNEIINWLDK
P26641	Eukaryotic translation elongation factor 1 γ	KEK*GSREEK	K*FPAGK
		RILGLLDAYLK*	TNRTPEFLRK*FPAGK
P35579	Myosin, heavy chain 9	QQAADK*YLYVDKNFINNPLA	TVGQLYK*EQLAK
		QADWAAK	QQQLTAMK*VLQRNCAAYLK*LR
			VKVNKDDIQK*MNPPK
P46821	Microtubule-associated protein 1B	KDSVAAGK*PK	QQAADKYLYVDK*NFINNPLAQADWAAK
P53618	Coatomer protein complex, β 1	EVIK*TNNVSEHEDTDKYR	K*PAALKPK
Q13200	Proteasome 26S subunit, non-ATPase, 2	NGFVNAAFQGDKLLTDDGNK*	K*TLQLALDLVSSR
Q14008	Cytoskeleton associated protein 5	K*PGWKETNFQVMQMKLHIV	DAGDKDK*EQELSEEDK
Q14204	Dynein, cytoplasmic 1, heavy chain 1	ALIAQK*	K*EGLDEVAGIINDAK
Q15084	Protein disulfide isomerase family A6	KLACK*VDGHK	QFEK*QVELYR
Q15149	Plectin	SLLQALNEVK*GR	PTIK*IFGSNK
		GK*VKLAAVDATVNQVLASR	GVARLSAEAEK*
		IIITVVEEQEQK*GR	ELMWLNEK*EEEEVGFDWSDR
Q9Y490	Talin 1	LGAASLGAEDPETQVVLINAV	LAQVAKAVSHSLNNVCNCLPGQKDVDVAL
		K*DVAK	K*SIGESS

Figure 3.12. Di-Glycine Pull Down Results.(A) Total number of Proteins with GlyGly modified peptides established through Mass Spectrometry, between the WT and I26A cell lines. (B) Representation of ubiquitin and the lysine residues on which GlyGly signature peptides were detected from the pulldown. (C) Proteins that are detected in both cell lines, but with different sites of ubiquitination. K* represents that di-glycine modified peptide.

generated with an overall false discovery rate of 5%. For the Peaks output, hits were accepted with an average localised score of 55% and a 5% FDR.

The total number of Gly-Gly hits between WT and I26A are comparable, with there being 30 more hits in WT for experiment 2 and approximately 100 more hits in I26A for experiment 3. This analysis suggests that we have successfully managed to enrich di-glycine species. Moreover, Peaks appears to be a useful tool in determining post-translational modification sites, and may increase the number of hits identified in a proteomic run.

We have also identified, in all runs, multiple sites of ubiquitination on ubiquitin, with 4 of 7 lysines appearing in our analysis, as illustrated by Figure 3.11. We did not pick up Lys27, 29 or 33, however, these are reported to be less common than the more conventional Lys 48 and Lys 63 (Peng et al. 2003).

Figure 3.11 also contains a table of an interesting subset of the proteins that were identified in both WT and I26A, but with different peptides identified in the two samples. These on first glance would be discounted from the analysis, but may be of functional relevance, with differential ubiquitination sites given the profile of the E3 ligases in the cell.

3.2.5 Generating a candidate list of potential BRCA1 E3 ubiquitin ligase substrates.

To generate a final candidate list of potential BRCA1 E3 substrates, several thresholds and criteria were established. Firstly, the proteins that appeared within the UbiQapture hits within the 3:0 and 2:0 categories were considered. Secondly, there had to be a log₂ enrichment score greater than +/-2. Proteins were also considered if peptides were uniquely identified through the ubiquitin remnant profiling experiment. Finally, the lower categories were manually searched for potentially interesting hits based on the

literature and their relative enrichment. The candidate list was subsequently trimmed to remove any proteins that appeared to be unusual or anomalous (such as keratin, which is most likely to be a contaminant).

3.3 Discussion

Hereditary breast and ovarian cancers are highly metastatic and a significant proportion of these cancers are in part due to BRCA1 mutations. This Chapter seeks to identify potential BRCA1 E3 substrates through two proteomic strategies in order to reveal cytoplasmic functions of BRCA1 that may influence tumourigenesis, specifically with the idea of regulation of cell motility in mind.

Several large-scale experiments were carried out which were newly developed and brought about their own challenges. As a result the discussion will be split into two principal sections: ‘method development and processing of data’ and the ‘the biological output’ from the proteomic experiments.

3.3.1 Methods Development and data processing

In this Chapter we undertook two proteomic experiments. The first experiment involved the pull-down of ubiquitinated proteins from post-nuclear lysates using the UbiQapture kit, which binds mono- and poly-ubiquitinated species via an ubiquitin binding domain. Eluates were analysed by mass spectrometry and samples were compared by label-free quantitation by Spectral Index quantitation to identify proteins which were significantly upregulated in cells expressing functional BRCA1 versus cells expressing ubiquitin ligase dead BRCA1.

UniProt ID	Gene Description	Category	Mean Log2 Fold Change	Peptide
P26038	moesin	3:0	2.91	LLPQRVLEQHKLNK
Q96EY7	Pentatricopeptide repeat domain 3	3:0	3.49	
P08238	heat shock protein 90kDa alpha (cytosolic), class B member 1	3:0	2.23	
Q14258	tripartite motif containing 25	3:0	2.57	
P62249	ribosomal protein S16	3:0	4.45	
O43707	actinin, alpha 4	2:0	1.45	
Q04637	eukaryotic translation initiation factor 4 gamma, 1	2:0	1.72	
P35268	ribosomal protein L22	2:0	6.54	KLVVK(+114.04)GGKK
P52294	karyopherin alpha 1 (importin alpha 5)	2:0	1.96	
P62899	ribosomal protein L31	2:0	4.98	
O60716	catenin (cadherin-associated protein), delta 1	2:0	2.18	K(+114.04)PLEE
P15311	ezrin	2:1	0.40	EAQDDLKTK(+114.04)TK
P46940	IQ motif containing GTPase activating protein 1	2:1	0.78	SLNIK(+114.04)TDPVDIYS
P26373	ribosomal protein L13	2:1	0.53	YHTKVR
O00410	importin 5	1:0	2.58	
Q9UM54	myosin VI	1:0	3.47	
P60981	destrin (actin depolymerizing factor)	1:0	3.56	
P22102	phosphoribosylglycinamide formyltransferase, phosphoribosylglycinamide synthetase, phosphoribosylaminoimidazole synthetase	1:0	4.20	NLIESMQINGSVLKNGSLTNHFSFEKK(+114.04)
P53618	coatamer protein complex, subunit beta 1	1:0	5.76	EAADPLASKLNK
Q01518	CAP, adenylate cyclase-associated protein 1 (yeast)	1:0	0.17	K(+114.04)PGPYVKEMNDAAMFYTNRLKE YK(+114.04)
P27824	calnexin	1:0	4.67	QTSGMEYK(+114.04)KTDAPQ
Q29RF7	PDS5, regulator of cohesion maintenance, homolog A (S. cerevisiae)	1:0	0.25	LFGSK(+114.04)
O75083	WD repeat domain 1	1:0	3.67	WDSGSSVGELTGHK(+114.04)
O00629	karyopherin alpha 4 (importin alpha 3)	1:0	6.62	
O95782	adaptor-related protein complex 2, alpha 1 subunit	1:0	0.42	KK(+114.04)GPGAGSALDDGR
P50402	emerin	1:0	6.29	
P50570	dynamin 2	1:0	2.51	
P05362	intercellular adhesion molecule 1	1:0	4.32	
P23528	cofilin 1 (non-muscle)	1:0	1.76	SKMIYASSK(+114.04)DAIKK
P40227	chaperonin containing TCP1, subunit 6A (zeta 1)	1:0	1.76	Multiple peptides
Q9NTI5	PDS5, regulator of cohesion maintenance, homolog B (S. cerevisiae)	1:0	1.80	SALLPVLHHKSK(+114.04)
Q00610	clathrin, heavy chain (Hc)	1:0	0.85	Multiple peptides
P35606	coatamer protein complex, subunit beta 2 (beta prime)	1:0	1.83	

Table 3.1.a: BRCA1 substrate candidate list based on proteins enriched in HCC1937 + WT BRCA1 cells from UbiQapture Pulldown and Ubiquitin Remnant Profiling: Contains UniProt ID, Category, mean log2 fold change and the peptide if established from the ubiquitin remnant profiling. Mainly consists of hits reproducibly enriched in WT.

UniProt ID	Gene Description	Category	Mean Log2 Fold Change	Peptide
P07437	tubulin, beta class I	0:3	-0.84	
P04406	glyceraldehyde-3-phosphate dehydrogenase	0:3	-0.72	VVDLMAHMASK(+114.04)
P60709	Actin, cytoplasmic 1 OS	0:3	-3.63	
P06748	nucleophosmin (nucleolar phosphoprotein B23, numatrin)	0:2	-4.96	SIRDTPAK(+114.04)NAQS
Q562R1	Beta-actin-like protein 2 OS	0:2	-7.78	
Q9Y678	coatamer protein complex, subunit gamma	0:2	-1.41	ATFYLVLEQK(+114.04)
P08240	signal recognition particle receptor (docking protein)	0:1	-2.77	GTLGGMFGMLK(+114.04)GLVGSS
Q9NQC3	reticulon 4	0:1	-7.24	LPSDTEK(+114.04)EDS
Q99460	proteasome (prosome, macropain) 26S subunit, non-ATPase, 1	0:1	-1.36	KEEKEK
P30050	ribosomal protein L12	0:1	-2.00	IGPLGLSPK(+114.04)
P10809	heat shock 60kDa protein 1 (chaperonin)	0:1	-1.11	DVK(+114.04)FGADARALMLQGVDLLADAVAV TMGPG
Q7Z406	myosin, heavy chain 14, non-muscle	0:1	-4.80	LTAEIWK
O60568	procollagen-lysine, 2-oxoglutarate 5-dioxygenase 3	0:1	-2.80	K(+114.04)FVQSGSL
Q10567	adaptor-related protein complex 1, beta 1 subunit	0:1	-0.56	VFFVK(+114.04)
Q96RT1	erbB2 interacting protein	0:1	-1.43	IYDILSDNGPQQPSTTVKITSVDGK(+114.04)
P27816	microtubule-associated protein 4	0:1	-2.99	Multiple peptides
Q9P2E9	ribosome binding protein 1 homolog 180kDa	0:1	-0.30	K(+114.04)QLVAREQEIT
P49368	chaperonin containing TCP1, subunit 3 (gamma)	0:1	-0.86	P.RIVLLDSSLEYK(+114.04)G
P21980	transglutaminase 2 (C polypeptide, protein-glutamine-gamma-glutamyltransferase)	1:2	-1.73	ILGEPK(+114.04)QR
Q9UHB9	signal recognition particle 68kDa	1:2	-0.60	TLNFK(+114.04)MGNH
P53621	coatamer protein complex, subunit alpha	1:2	-0.59	LNDLIQRLQLCYQLTTVGK(+114.04)FEE

Table 3.1.b: BRCA1 substrate candidate List based on proteins enriched in HCC1937 + I26A BRCA1 from UbiQapture Pulldown and Ubiquitin Remnant Profiling: Contains UniProt ID, Category, mean log2 fold change and the peptide if established from the ubiquitin remnant profiling. Mainly consists of hits reproducibly enriched in I26A.

The second experiment utilised the same post-nuclear lysates. Proteins were isolated and digested in-solution. The resultant peptides were run through columns that contained the GX41 antibody that identified the di-glycine post-trypsin signatures on the ubiquitination sites. The eluates were analysed by mass spectrometry to identify sites of ubiquitination. Multiple considerations were made when deciding upon proteomic approaches that will be discussed further here.

3.3.1.i Quantitative proteomics methods for pulldown of ubiquitinated proteins using UbiQapture

Multiple strategies and techniques are available for undertaking quantitative proteomics, broadly split into 3 categories; labelling proteins in cell culture SILAC, labelling proteins directly by iTRAQ or Label-Free methods. Whilst the labelling approaches are more established label free methods have become widely used over the past 5 years. More recently, studies have shown that label-free methods are more reproducible and able to quantitate a greater number of proteins in comparison to both SILAC and iTRAQ (Trinh et al. 2013; Liu et al. 2013). In addition label-free methods are more cost-effective.

3.3.1.ii Missing Data for Label Free Quantitation

One difficulty associated with Label Free Quantitation is the presence of zero values, which is not problematic for label-based studies. The UbiQapture experiment generates a list of proteins based upon peptide identification for each sample. Within these lists there are a subset of peptides and proteins that are uniquely identified in a particular sample. For these hits it is not possible to establish a numerical fold change value based on the normalized spectral index, resulting in missing values. These zero values may be

reported due to the peptide and parent protein truly being absent in one sample. However, there is also the possibility that the peptides are present but at an abundance which falls below the detection threshold. Alternatively, the peptide may just be incorrectly identified or missing completely at random (MCAR) due to unknown factors such as poor ionisation. Missing data are a common problem in the field of high-throughput genomics and proteomics and whilst several models have been proposed,(Karpievitch et al. 2012) there is no consensus on how to most effectively handle these data.

A difficulty is knowing how to interpret and handle these missing values. Due to the complex nature of our sample and the known limitations of the dynamic range of the Orbitrap relative to the Q-Exact mass spectrometer, a significant proportion of peptides uniquely identified in one sample may still be present in the other sample but just below the threshold of detection. For data handling purposes, we have taken the conservative approach of assuming that these peptides and proteins are present and below the detection threshold. For each experiment, the minimum observed Normalised Spectral Index value was used as a substitute for the zero value assigned to the unobserved proteins. The rationale is that this is a conservative measure of the detection threshold, below which proteins may be present with a normalised spectral index value that does not get reported.

This allows us to generate conservative estimates for the normalised spectral index fold change for every protein in the WT and I26A samples. The two obvious caveats to this approach is that there may well be instances when the protein is truly absent or at a much lower abundance and we have assigned a fold change value that overestimates the presence of said protein. The other possibility is that the peptides from a protein may fall into the MCAR category, underestimating the true biological variation. However, MCAR values have been reported to make up a relatively small proportion of missing data values (Karpievitch et al. 2009). Thus the majority of the newly imputed values will report an

underestimated fold change value, allowing us to process the data in a more systematic way whilst reducing the probability of generating false positive hits.

3.3.1.iii Reproducibility of the UbiQapture data

The proteomics data from the UbiQapture experiment generate a large number of proteins, a significant proportion of which are identified in all triplicate experiments for each sample. From the WT triplicate analyses, 76% of the proteins appear in at least two repeats, and 66% of proteins appear in at least two repeats of the I26A runs.

Several thresholds were applied to the data processing steps to ensure reporting of optimum protein hits. Reported proteins had to have 2 unique peptides identified and come from a group with a q-value corresponding to a 1% False-Discovery Rate. Given the complex nature of samples, achieving consistently reproducible list of proteins between experiments is a hard task in proteomics. Even when using a standardised sample, studies have shown reproducibility of 60% for the protein profiles and down to 20% reproducibility of peptides (Tabb et al. 2010). Other studies have shown 40% of peptides are unique when the same sample has been run 5 times.

A number of proteomic studies, whilst highly cited and published in high-impact journals, still only present data from a single mass spectrometry run or with no biological repeats (Jeon et al. 2007)(Lai et al. 2011). Subsequently our data are relatively robust, given that each iteration of the experiment has minor alterations to the protocol in order to try to enrich the number of ubiquitinated proteins.

3.3.1.iv Method Development and analysis of GX41 Results

In this chapter we have developed a protocol that has identified up to 400 sites of ubiquitination. It is the first report utilising the GX41 antibody to identify candidate

substrates of an E3 ligase, BRCA1 that does not necessarily have the downstream fate of prompt proteasomal degradation. Prior to this antibody, large-scale studies ranged from identifying 70 to 700 proteins that were thought to be ubiquitinated (Vasilescu et al. 2005; Meierhofer et al. 2008). Since the establishment of this antibody (Xu et al. 2010), there have been several studies that investigate the global ubiquitinome and have identified up to 20,000 ubiquitination sites (Wagner et al. 2011; Kim et al. 2011).

The development of the experimental pipeline is described in detail in Section 3.2.4.i. Experimental repeats were altered to maximise number of GlyGly peptides. The major changes were the use of an in-solution digest, increase in the amount of the starting material and increase in the amount of antibody used.

The pull-down of di-glycine modified peptides yielded a greater a number of hits with each subsequent iteration. The number of hits identified through Mascot and Pipeline were still relatively modest compared to the global studies which report identification of thousands of ubiquitination sites. However, this is to be somewhat expected given the use of proteasome inhibitors to enrich the sample with ubiquitinated species.

There were concerns, however, that there may be certain inherent biases against identifying di-glycine peptides through Mascot and Pipeline. Firstly, we are searching for individual peptides with a modified lysine, yet the Mascot analysis favours proteins with many contributing peptides, all of which may not be ubiquitinated. Thus a single di-glycine modified peptide may report a low score, or may not be reported at all, due to the protein not having sufficient peptides to contribute to having a score above threshold.

Inherent biases have been associated with databases used to characterize test proteins in the past due to the way the databases have been constructed, which required manual curation and validation. (Bell et al. 2009).

Furthermore, a proportion of the di-glycine modified peptides will have multiple missed cleavage sites, which may not correspond to the search parameters. The precursor ions for Gly-Gly peptides are also likely to be greater than the usual 2+. Finally, fragmentation of Gly-Gly peptides yields fewer b ions and y ions resulting in a lower score when matching the peptide spectra to the expected spectra, compared to unmodified proteins (Chicooree et al. 2013).

Consequently, the raw data was analysed through a de novo sequencing software called PEAKS de novo. This permitted the generation of amino acid sequences without the automated filtering and scoring that the database searching and matching process undergoes, allowing us to capture peptides that may be sitting just below the detection limit when analysed through Mascot or Pipeline.

3.3.2. The Biological Output from Proteomic Experiments

Having addressed the complications and obstacles faced when developing our experimental approaches, we shall address the: biological output.

3.3.2.i Cell lines and lysis procedure

Our initial approach to identify candidate substrates of BRCA1's E3 ligase activity is based upon comparing the differences in the ubiquitination profiles between cells with functional BRCA1 and ubiquitin ligase dead BRCA1. We are using a breast cancer epithelial cell line, HCC1937, which carries one BRCA1 null allele and a mutation (5382insC) in the second allele which results in a premature stop codon truncating the protein at residue 1829 and disrupting the BRCT domain (Tomlinson et al. 1998).

We utilised cell lines that had been transfected with either full length functional BRCA1 (HCC1937-WT) or with ubiquitin ligase dead BRCA1 through an Isoleucine to Alanine point mutation (HCC1937-I26A). This point mutation in the RING domain does

not significantly disrupt the 3D structure, but interferes with the interaction of BRCA1 with E2 conjugating enzymes such as UbcH5, eradicating its E3 ubiquitin ligase activity. (Brzovic et al. 2003b) The plasmid encoding WT and I26A BRCA1 is expressed under G418 selection. Western Blot of post-nuclear lysates validates the expression of BRCA1 within the WT and I26A cell lines.

3.3.2.ii Identification of poly and mono-ubiquitinated species

A significant advantage in using the UbiQapture kit was the ability to capture both poly and mono ubiquitinated species. A number of studies that attempt to identify ubiquitination substrates of E3 ligases employ an experimental design with the specific intention focusing on those that are targeted for proteosomal degradation. These approaches include the utilisation of proteasome inhibitors(Meierhofer et al. 2008), an immunoprecipitation specifically for poly-ubiquitinated substrates(Hjerpe et al. 2009) or identifying a pool of proteins that are enriched in cells overexpressing an ubiquitin ligase dead E3 isoform (Lai et al. 2011).

However, when designing the experiments, we wanted to capture both mono-ubiquitinated as poly-ubiquitinated species, including poly-ubiquitinated species that are not targeted for proteasomal degradation. There are two principal reasons guiding this decision. Firstly, some of the BRCA1 substrates that have been identified have shown to undergo mono-ubiquitination, such as Estrogen Receptor alpha(Ma et al. 2010), PCNA (Tian et al. 2013) and γ -tubulin (Starita et al. 2004). Secondly, we are particularly interested by BRCA1 activity at the cell membrane and cytosol. Another important RING E3 ubiquitin ligase found at these locations is c-Cbl, which is shown to mono-ubiquitinate its substrates as part of the ESCRT pathway. There may be some overlap in function and substrate between c-Cbl and BRCA1. Thus not excluding mono-ubiquitinated proteins is

an important consideration when searching for BRCA1 candidate substrates. Moreover, SAGB2 was recently shown to be a BRCA1 substrate for poly-ubiquitination, which is not targeted for proteasomal degradation (Song et al. 2011).

We have not grown cells in the presence of proteasome inhibitors, unlike the majority of large-scale ubiquitination studies. The addition of proteasome inhibitors stops the turnover of proteins by the ubiquitin proteasome system, which results in the accumulation of ubiquitinated species in the cell, as well as in an increase in the number of sites that are ubiquitinated. A disadvantage to this is that the cell undergoes proteasomal stress and behave abnormally, inhibiting growth, triggering apoptosis and interfering with the cell cycle (Han et al. 2009).

Given that BRCA1 may be ubiquitinating proteins at relatively low level, without the downstream consequence of proteasomal turnover, we did not use MG-132, as the ubiquitination signal of true potential mono-ubiquitinated candidates may be masked by the dramatic increase in the total number of ubiquitinated proteins. Moreover, we did not want to interfere with important cellular pathways as this could lead to the introduction of artefacts or false positives in our proteomics results.

These considerations mean that the protocols used in our experiments cannot be directly compared to the majority of published proteomic studies of ubiquitination. However, infrequent studies performed without proteasome inhibitors suggest that the number of targets and the reproducibility of results described in this chapter are consistent with current best published results (Udeshi et al. 2012; Lee et al. 2011).

3.3.2.iii GlyGly hits in UbiQapture experiment

The number of Gly-Gly tagged peptides identified in the UbiQapture experiment is relatively limited, with less than 20 peptides identified. These small numbers are also

observed in other large-scale ubiquitination proteomic studies (Lopitz-Otsoa et al. 2012; Meierhofer et al. 2008) using a protein-enrichment strategy, as opposed to the peptide-enrichment approach. Many proteins may only have a single site of ubiquitination; thus, the majority of the tryptic fragments generated from the digestion will not in fact contain a di-Glycine modified lysine residue. Furthermore, the di-glycine modified peptides will often contain several missed cleavage sites, and have poorer peptide spectrum match scores due to the nature of the fragmentation. Subsequently, these peptides may lie just below the detection threshold. Nonetheless, given the strong signal of eluted fractions in the Western Blot when probed with anti-ubiquitin antibody we can make the assertion that significant proportion of the hits are ubiquitinated under endogenous cellular conditions at steady state.

3.3.2.iv Functional relevance of WT or I26A enriched proteins

Proteins that have been significantly enriched within the WT sample represent a pool of proteins that are specifically ubiquitinated by BRCA1, given that the only difference between the two cell lines is the presence of functional BRCA1 E3 RING domain. The presence of these proteins suggests that either the proteins are ubiquitinated with a non-proteasome related downstream function or that they are targeted for proteasomal degradation but the substrates do not undergo rapid turnover and the rate of polyubiquitination and degradation is sufficiently low to be able to capture these proteins.

Proteins enriched in the I26A cells may indicate hits that are normally ubiquitinated by BRCA1 and are targeted for proteasomal degradation. The absence of BRCA1 results in the inability of these substrates to be directly ubiquitinated by BRCA1. Other E3 ubiquitin ligases may ubiquitinate these targets but at a lower rate leading to its enrichment in I26A cells, as substrates can be ubiquitinated by multiple E3 ubiquitin

ligases, competing for the same lysine or ubiquitinating at different sites (Brooks & Gu 2011).

We employed a relatively gentle lysis procedure as described in Section 2.3.6. It is therefore likely that a proportion of the proteins we have identified may not be directly ubiquitinated. These may be partners in a complex within which a single protein may be ubiquitinated. As with all pulldown experiments, there is also a chance that particularly abundant or sticky proteins may be present in our pull-down.

3.3.2.v Gene Ontology, Pathway Analysis and Interesting candidates

A number of interesting results are revealed through the gene ontology and KEGG pathway analysis. There were no obvious differences in the pathway analysis when comparing lists of all the proteins that had been identified in WT and I26A. However, when we focus the proteins that have been enriched in either WT (mean log₂ fold change > 0) or I26A (mean log₂ fold change < 0), we see a number of interesting differences.

Most notably, we consistently see an enrichment of cytoskeletal related proteins and pathways enriched in the cells expressing functional BRCA1. In the Biological Process and Molecular function Gene Ontology results, we see significantly more proteins in the WT cells for the categories actin binding cytoskeletal protein, non-motor actin binding protein, cytoskeletal protein and also a modest increase in the cell structure and motility categories. Furthermore, 3 of the KEGG categories enriched in the WT cells are related to motility.

This, coupled with the identification of some interesting hits with high log₂ mean fold change values, reinforces the newly emerging notion that BRCA1 has a role in regulating cell motility (Elisabeth D Coene et al. 2011; Yasmeen et al. 2008; Promkan et al. 2009). Loss of BRCA1 has been implicated in having a role in motility and invasion

but no clear mechanism has been ascribed to this process. Recent gene ontology analysis on microarray results focusing on BRCA1's function also shows an enrichment of genes involved in cell motility (Dacheux et al. 2013). The results of this chapter shed more light as to which pathways and proteins may be involved in regulating motility.

We see several pathways and processes that may be involved in regulating cell motility. Firstly, we identify the proteins moesin and ezrin enriched in WT. These two cytoskeletal related proteins are part of the ERM-family. They also contribute to the over-representation of the KEGG pathways leukocyte transendothelial migration and pathogenic E. Coli infection. On closer inspection of these pathways involves remodelling of the cell-cell contacts, cell adhesions and motility. Enrichment in these categories indicates an enrichment of cytoskeletal related proteins in the WT sample, pointing towards the increasing possibility that BRCA1 has a role to play in regulating cell motility. Moreover, these proteins have been implicated in carcinogenesis. Moesin has specifically been associated with breast carcinomas exhibiting aggressive, basal-like phenotypes(Charafe-Jauffret et al. 2007). On a molecular level, phosphorylation of ezrin regulates the invasive potential of breast carcinoma cells (Mak et al. 2012). The ERM family proteins have also been specifically linked to regulating cell spreading and motility through interactions with BRCA1(Elisabeth D Coene et al. 2011).

Other interesting proteins include IQGAP1, which has established roles in regulating cell-cell adhesion and cell migration(Johnson et al. 2009). The relative expression of IQGAP1 in breast and ovarian cancerous cell lines are higher than normal as verified by immunohistochemistry, Western Blot and qRT-PCR. Notably, the invasive breast cancer cell line MDA-MB-231 has a 3 fold increase in IQGAP1 relative to MCF-7 cell line, which is less invasive(Jadeski et al. 2008). A recent study demonstrated that

BRCA1 and beta-catenin directly interact (Li et al. 2010), suggesting this could be another attractive candidate to investigate further.

Interestingly, we also see a significant enrichment of proteins and pathways involved in the ribosome and translation, in the WT expressing cells. The role of BRCA1 in translation has not been explored in great detail, but recent research has shown that BRCA1 has been implicated in regulating translation (Dacheux et al. 2013). Specifically, BRCA1 has been shown to regulate translation through interaction with PABP1 (Dizin et al. 2006), which is also enriched in WT from our proteomic experiments, making it a possible candidate substrate for BRCA1 E3 ubiquitin ligase activity. BRCA1 may ubiquitinate PABP1 and potentially other ribosomal proteins to stabilise an mRNA binding complex or initiate a signalling cascade to modulate transcription. The large number of ribosomal proteins we see in our study may be due to several ribosomal proteins undergoing ubiquitination by BRCA1, or alternatively, a significant proportion of these may be partner proteins that are not directly ubiquitinated.

Other potentially interesting candidates have been previously associated with BRCA1. Expression levels of I26A enriched thrombospondin have been linked to BRCA1 expression levels (Grillo et al. n.d.). RNA Helicase A has been shown to directly interact with BRCA1 (Schlegel et al. 2003). Tubulin expression has been correlated with the expression of BRCA1 and associated with poor outcomes in malignant thymic epithelial tumours (Kaira et al. 2011). The two proteins are also observed to co-localise with at mitotic spindle as verified by confocal microscopy and co-immunoprecipitation experiments (Lotti et al. 2002).

3.3.2.vi Discussion of *gx41* results

Here we have identified a number of differences between WT and I26A and the di-glycine hits that are consistent with the UbiQapture findings that have been tabulated with their associated site of ubiquitination on Table 3.1.a and Table 3.1.b. The vast majority of studies have looked at the ubiquitination profiles of cells with the treatment of proteasome inhibitors, DUB inhibitors and DNA damaging agents, in order to induce ubiquitination and enrich sites of ubiquitination. We have not treated cells with any reagent that interferes with crucial cellular function; thus, we obtain a snapshot of the cells in their endogenous, healthy, *in-vitro* steady state. Although this comprises the total number of sites observed, it allows us to identify potentially identify more functionally relevant sites and avoid hits that are a consequence of the stress caused by aggregates of poly-ubiquitinated proteins that accumulate when proteasome function is inhibited.

Identifying sites of ubiquitination is still not an easy task – there were a number of hurdles to overcome and there are still some caveats to this study. Firstly, the ubiquitin remnant profiling approach cannot differentiate between ubiquitination and certain other post-translational modifications, specifically Neddylation and ISG15-ylation, which both result in di-glycine bonds on lysine residues. However, the proportion of these post-translational modifications is considered to be relatively small, with one report stating that only 6% of di-Gly peptides are derived from ISG15 and NEDD8 (Kim et al. 2011).

In addition, further sites could be identified through the use of DUB and proteasome inhibitors, as it has been shown that both of these factors, and particularly proteasome inhibitors, have a large effect on the ubiquitination landscape (Udeshi et al. 2012). However, there are biological advantages to not using these, namely not putting the cells under stress and masking functionally relevant hits.

More sites could be identified by pre-fractionating (Udeshi et al. 2012) the samples more and extending the run time. Furthermore, the ubiquitin remnant profiling can be undertaken quantitatively by coupling it to SILAC as carried out by Emanuelle et al.

Whilst we did three repeats of the experiment, each iteration had minor alterations to the protocol in order to try to maximise the number of di-Gly hits. This explains the increase in the number of di-Gly peptides identified with each successive repeat. For this reason, the peptides and proteins identified are not very reproducible, and thus are not being processed in the same way as the UbiQapture experiments. Moreover, it has been reported that reproducibility in ubiquitin remnant profiling experiments is relatively low, even for mass spectrometry experiments. In four repeats, Kim et al. report 44% of the peptides identified were unique to one repeat and 24% of the peptides were present in all four repeats (Kim et al. 2011).

3.3.2.vi Biological output from GX41 experiment

From the GX41 experiment we identify up to 400 peptides through both Mascot search and de novo sequencing using PEAKS. In every experiment we identify ubiquitin peptides as would be expected. We have annotated and labelled the hits that are consistent with the UbiQapture experiment as these are the hits which are of most interest as they represent potential candidates.

Within the subset of proteins identified within these experiments, there is a significant subset that is identified in both WT and I26A; however, with closer inspection, it appears that the peptides identified are different. This is the case with Myosin-9, which appears in both lists, but where the ubiquitinated peptides are different. These represent an important additional category of potentially significant hits, since the differential ubiquitination of the protein at distinct sites may be due to direct effect of the

presence/absence of BRCA1, and may have distinct functional consequences. We note that such subtle effects of BRCA1 could never be detected using the UbQuapture technique, since the protein is ubiquitinated in both cell types.

3.4 Conclusions

To further understand the molecular basis of hereditary breast and ovarian cancers, focus needs to be applied to protein-protein interactions, comparing proteomic profiles of healthy and cancerous cells. It is still not completely understood why patients with mutations in their BRCA1 gene have much more aggressive breast and ovarian cancers.

In this chapter, we have tried to address this problem by trying to identify the substrates of BRCA1 enzyme activity. We have done this by taking an unbiased, quantitative dual strategy proteomic approach to identify potential candidate substrates for BRCA1 E3 ubiquitin ligase activity. This is the first report of a systematic approach to identifying candidate BRCA1's extra-nuclear substrates.

We have established proteomic potential substrates through pull down of ubiquitinated proteins using an affinity column containing an immobilised ubiquitin binding domain. In addition, we have identified, on a peptide level, novel sites of ubiquitination and potential differences between cells expressing functional BRCA1 and ubiquitin ligase dead BRCA1.

This study can be utilised to gain insight into BRCA1's extra-nuclear functions by examining the candidate ubiquitin ligase substrates. Particularly promising and relatively unexplored areas are BRCA1's role in motility and its role in translation.

Moreover, this approach can be applied to studies looking to identify substrates of other E3 ubiquitin ligases, specifically when non-degradation ubiquitination is anticipated. The dual approach is extremely beneficial, as from a protein level, one can highlight enriched complexes and pathways which may be relevant to protein function on a more global scale. The additional layer of the peptide approach allows one to focus on specific sites and consider the regulation and modulation of an individual amino acid and how modification of said residue could affect protein function.

From a clinical perspective, this study generates the possibility of developing therapeutics for novel targets involved in hereditary breast and ovarian cancer. This chapter concludes on a candidate list of BRCA1 E3 substrates. These will need to be validated by further in vitro and in vivo studies, specifically determining the interaction of E2, E3 and substrate.

Particularly interesting from a therapeutic perspective, are the proteins involved in motility. Metastasis is the fatal step in tumourigenesis and, by understanding the interaction of BRCA1 with cytoskeletal regulating proteins, novel approaches can be taken to try to prevent and treat metastasis of BRCA1 tumours. The next step in doing this is to establish which of the candidate proteins are involved in motility and how their function affects cell migration and invasion.

Chapter 4 – Linking Candidate substrates to a motility phenotype using an in-silico approach

4.1 Introduction

There is a growing body of evidence that implicates BRCA1 in having a role in regulating motility (Coene et al. 2011; Yasmeen et al. 2008) (as discussed in Section 1.5.6). This is of particular interest when considering the biology of BRCA1 tumours in the context of invasion and metastasis, as there is a strong correlation between individuals with BRCA1 mutated breast tumours and shorter metastatic-free periods (Stoppa-Lyonnet et al. 2000). Moreover, a significant number of our candidates generated from the proteomic experiments from Chapter 3 are cytoskeletal and motility regulating proteins, which is consistent with the emerging notion that BRCA1 ubiquitin ligase activity plays a role in regulating motility.

4.1.1 BRCA1 candidate substrates – next step

In order to elucidate BRCA1's role in motility, we need to test the candidate substrates and investigate their potential function in motility. The most compelling and convincing candidates will be the proteins with a robust proteomic identification and experimental verification of having a role in motility. These candidates will be verified further in order to determine the interaction and relevance with BRCA1. This will subsequently increase our knowledge of the cell biology of BRCA1-mutated cancers with the long-term aim of developing therapeutics to target these pathways to prevent metastasis.

Several methods can be used to assess the role of individual proteins in cell motility. A logical first step would be to knock-down the expression of each protein by siRNA and look at the motility phenotype of the cell. Whilst doing this for a few proteins

is experimentally manageable, doing it for more than a handful of candidates very quickly becomes time and resource costly. Therefore we decided to utilise an *in silico* approach by taking advantage of an existing online database that summarises the result of over 192,000 siRNA experiments. Such a genome-wide approach is unbiased and has the advantage of giving information for every single gene and allows one to gain an understanding of the landscape and boundaries of a biological function such as cell motility.

4.1.2 The MitoCheck Database

The MitoCheck Database contains a series of time-lapse confocal movies of HeLa cells expressing GFP-tagged Histone 2B, which have systematically siRNA knocked down every gene in the human genome. These movies track the behaviour of single cells and were used primarily to investigate and discover novel genes involved in mitosis (Neumann et al. 2010).

4.1.2.i Organisation of the MitoCheck database

In order to knockdown 23,000 genes, Neumann et al established an siRNA library, created by Ambion based on ENSEMBL v27, with 2-3 independent sequences used to target each gene. Each siRNA sequence uniquely targeted and had 100% homology to the target gene. Live-cell time lapse microscopy assays were undertaken over 48 hours, with images taken every 30 minutes. Validation of knockdown on over a 1000 genes showed that 97% of the siRNAs knocked down of the relevant mRNA levels to less than 30% of the original levels. As the primary focus of the study was to investigate mitosis, frames were segmented and information regarding nuclear morphology was extracted.

Identification and categorisation of nuclear morphology was developed by machine-learning algorithm which initially required manual curation and validation. Each experiment was carried out on a 384 well plate, with positive controls of INCEP, COPBI

and KIF11 and negative control Scrambled to ensure cells were behaving as anticipated and ensured consistency between collaborating centres.

4.1.2.ii Outcomes of Genome-wide analysis

The primary motivation of this very comprehensive genome-wide study, was to investigate and further elucidate the proteins involved in regulating mitosis. To analyse mitosis, each nucleus was automatically assigned into one of 16 classes, based on their morphology and proliferation rate. This was amalgamated for each time-lapse movie which resulted in a phenotypic profile for every movie. Given that the cells were monitored for 48 hours, several iterations of mitosis were typically recorded.

The initial genome-wide screen categorised mitotically interesting hits based on unusual morphology such as poly-lobed or graped nuclei, and in addition, cells which exhibited delayed mitosis. 572 genes were identified as being significantly different to the control and were validated by a second screen with two additional siRNA knockdowns. These were investigated further through Gene Ontology analysis and the data was mined to identify the temporal order of mitotic phenotypes as well as other novel links and patterns.

Whilst the primary focus of the paper was mitosis, the authors also used the data to investigate motility and cell proliferation. The data obtained through segmentation and analysis of nuclear morphology was utilised to characterise the Euclidian distance travelled between the first and final frames for the nuclei. Moreover, the top scoring distance hits are reported in the paper's Supplementary Materials.

The MitoCheck database is an excellent resource and contains a remarkable amount of information. The raw data has been made publically available which makes it an incredibly useful tool for the field of cell biology, particularly those investigating the cell cycle, apoptosis and cell migration. We decided to utilise this database to investigate

the effects of the knockdown of motility in greater detail, given that there is also interphase information in the movies. We are more interested in the effect the siRNA knockdown has on cell speed, given that the Euclidian distance between the first and final frame may not always be able to capture erratic cell movement or hypermotility. Moreover, we would like to do this on the whole of the genome and identify whether any of our proteomic candidates exhibit increased speed phenotypes through this screen.

4.1.3 Aims of the Chapter

In this chapter our aim is to develop a new resource based on the MitoCheck database, and use this to assess which of our BRCA1 ubiquitination targets may have a role in regulating cell motility. In order to achieve this we develop an algorithm in collaboration with Dr Bogulsaw Obara to be able to identify and track cell motility from the MitoCheck time-lapse movies. After validating the methodology, we utilise the new migration database to perform a genome-wide analysis of the effect of knocking down individual genes by siRNA. Our three principal aims are:

- 1) Develop an automated pipeline and algorithm to track cell motility from the siRNA knockdown time-lapse movies available from the MitoCheck database
- 2) Establish a genome-wide database that contains motility data for every gene, annotated with gene ontology and KEGG pathway information.
- 3) Establish which of our candidates may be relevant in regulating cell motility based on their motility data from the MitoCheck database.

Ultimately we would like to use this tool to focus in on a handful of candidate proteins that have been identified through our proteomic experiments and appear to have a role in cellular motility. The long-term aim of this is to be able to help us establish the pathways and mechanisms that drive metastasis in BRCA1 mutated tumours.

We hypothesise that upon analysing motility after extracting cell speed data from MitoCheck database, there will be a proportion of genes that cause a reduction in cell motility when siRNA knocked down and conversely a proportion of genes that increase motility after knockdown. The majority of genes will not have an impact on cell speed and motility. We anticipate that there may be overlap in genes that increase motility after siRNA knockdown and proteins enriched in HCC1937 + WT BRCA1 cells and these are going to be the strongest candidates in the pursuit of identifying BRCA1 substrates involved in cell motility regulation.

4.2. Results

4.2.1 Developing an algorithm to measure cell tracking

In order to extract motility data from the MitoCheck time-lapse movies we developed an algorithm to detect and track the cells in each movie in collaboration with Dr Boguslaw Obara. The original MitoCheck movies were at a 10x magnification (NA 0.4), images taken every 30 minutes for 48 hours, with an average starting number of 67 cells (± 30). Primarily, nuclear morphologies were analysed based on segmentation step and analysing cell shape. However, Euclidian distance was calculated based on the difference in nuclear coordinates between the first and final frame. These values were used to report the top migration hits by Neumann et al. They also report a speed value for these top distance hits, based on the maximal difference in speed from the negative control between 2 consecutive frames.

When trying to establish our own approach to track cells there are multiple steps and obstacles that need to be considered and a general approach is displayed in Figure 4.1.A. The fundamental steps are converting the movie into frames, identifying the cells and tracking the same cells through all the frames, ensuring the cells maintain their assigned ID throughout all the frames. We want to be able to calculate the average speed of cells based on the distance moved by every nucleus between corresponding frames during interphase.

Figure 4.1.B shows the principal steps incorporated into our pipeline and algorithm to obtain speed measurements from the MitoCheck data. The movies were first downloaded, decompressed and converted into individual frames in Matlab. A local database of all movies in this form was maintained in Oxford and served to provide the input for our analytical algorithm.

The algorithm itself has three principal steps. The first step is identifying individual cells based on their nuclear fluorescence. As the videos only consist of fluorescence images, we make the assumption that there is one nucleus per cell. The nucleus is detected based upon its relative intensity compared to background. Images are then binarized and each cell/nucleus is eroded down to a single pixel, called a centroid. The centroid is determined for each frame. For this step, both a threshold value and kernel size needs to be established.

The next step is to trace the centroid from frame n to frame $n+1$. In order to do this, a search radius value, r_1 , needs to be established within which the algorithm should try to detect the centroid in frame $n+1$. Too small and the centroids are not linked between frames, whereas making it too large results in false-positive links between centroids from different cells in close proximity.

The following step involves merging broken traces together. Here a second radius value, r_2 , which is larger than r_1 , needs to be established. This represents the radius within which broken traces in frame n will try to be matched to broken traces starting in frame $n+1$.

Once traces are established, track lengths can be measured to calculate cell displacement and velocity, given that the time between frames is consistently 30 minutes. This can be used to identify hyper- and hypo-motile cells and their relation to our candidate list.

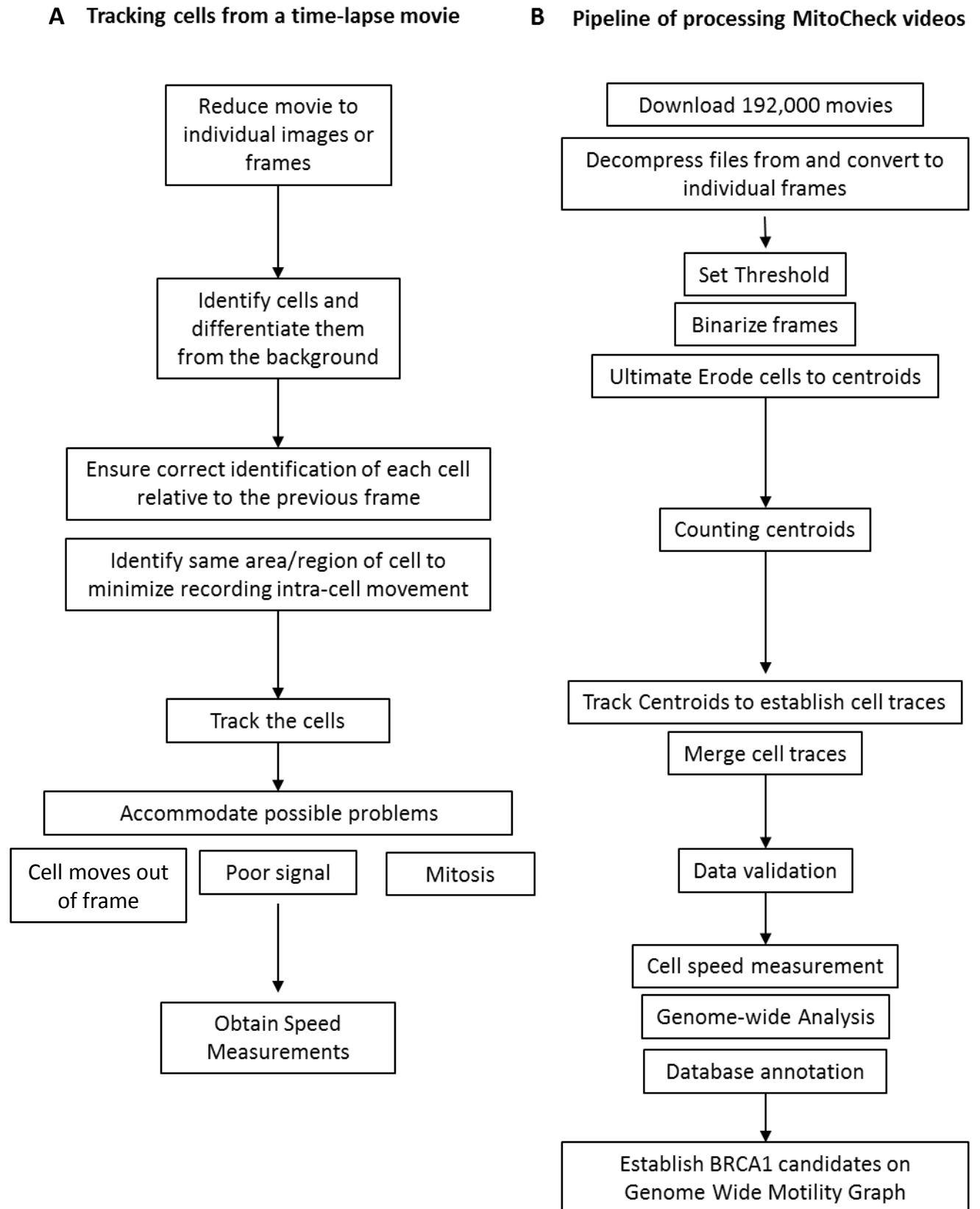


Figure 4.1 Schematic of analysing time-lapse movies: (A) General approach to track cells from a time-lapse movie. (B) Schematic of pipeline for processing videos from MitoCheck.

4.2.2 Establishing the variables for cell tracking algorithms

There are 3 principal steps to the algorithm, identification of the cells/centroids, joining the centroids between frames to form traces and merging broken traces. There are 5 main variables that needed to be optimised and established to ensure effective tracking of cells. These are the threshold, kernel size, tracing radius, merging radius and maximum number of frames to skip to merge traces.

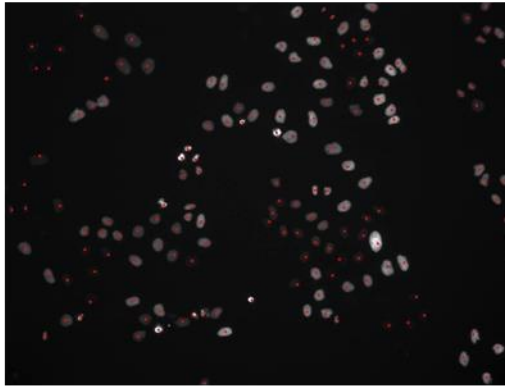
To establish the optimum kernel size and threshold we manually tested a series of different values for each variable for multiple videos. 8 movies were randomly selected based on a list of all the videos from proteins observed in the proteomic experiments. 2 frames were selected from each video, one from frames 1-45 and the other from frame 46-90. This was to ensure that neither photobleaching, nor the number of cells present in the frame would significantly affect our values. Cells from these frames were manually identified and utilised as the template or gold standard to which the varying outputs from the algorithm were compared.

Figure 4.2 shows an example of the process and analysis. Cells were marked in GIMP. From Figure 4.2.B we see the results of the comparison and the relative F-measure. F-measure is utilised to compare how similar two sets of data or clustering results are. As the F-measure value gets closer to 1, the greater the similarity of the two data-sets, with a value of 1 representing identical results. We ran the first step of the algorithm with a range of threshold and kernel size values and determined the F-measure value for each combination. Subsequently we could identify the best values for matching the algorithm output with the manual identification of cells. In the two examples shown in Figure 4.2.B, the F-measure values are represented as a heat map, red representing F-measure value close to 1 and blue close to 0. For the early frame, the optimum values were 13.5 and 0.025 for the kernel size and threshold giving an F-measure value of 0.97.

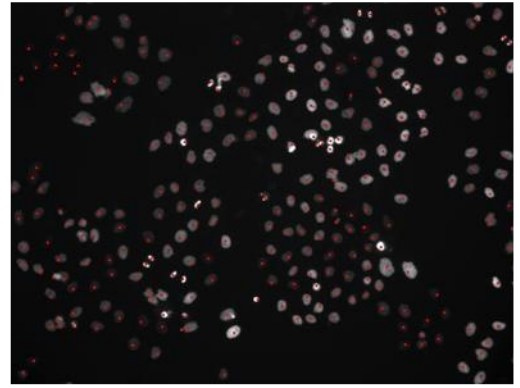
In the late frame, the optimum values were calculated to be 11.5 and 0.025, with an F-measure value of 0.96. The reduced kernel size in later frames could be due to more cells going through mitosis, which results in concentration of the nuclear signal and thus a smaller corresponding area.

In order to be able to run the algorithm on all frames of the movie it is necessary to choose a single value for these parameters. The optimum kernel size and threshold value was 12 and 0.02 respectively. In order to minimise false positives, we settled to input a threshold value of the mean plus one standard deviation, resulting in the final values of 12 and 0.025 for the kernel size and threshold.

The next variable r_1 , the first radius within which to search for the centroids between frames, is essentially an *in-silico* join-the-dots exercise. Again, there is an optimum value to be established. A radius that is too small will result in centroids not linking and many small broken traces; whereas too large may cause multiple centroids may be present within the search radius disrupting correct trace formation. To establish the optimum value, a visual iterative approach was used. Ideally one would want to take a similar approach to determining the kernel size and threshold, however manually tracking every cell in every frame for multiple movies was not practicable. Instead, varying radius values were inputted into the algorithm and 3D plots were generated, similar to the plot in Figure 4.3. These plots could be fully manipulated in 3D, permitting extensive manual examination, alongside the playing of the time-lapse movies to compare the efficacy of following centroids and forming correct traces when varying the radius value. A conservative value was assigned, as having a more stringent and smaller threshold would result in more broken tracks, yet fewer false positives. Moreover, the next step of the

A

KIF2A Video ID 7832 Frame 32



KIF2A Video ID 7832 Frame 71

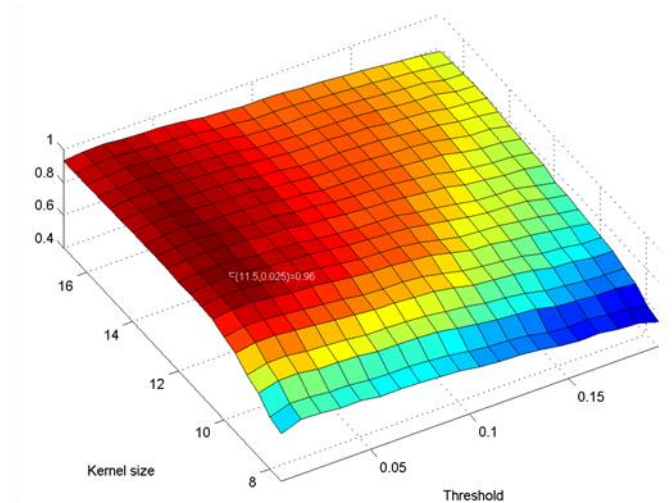
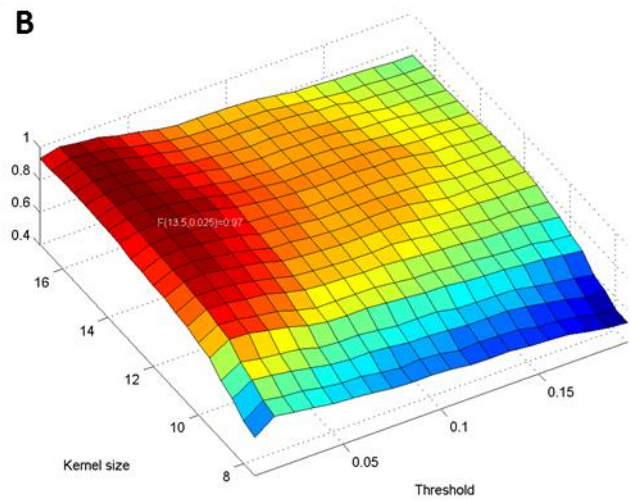
B

Figure 4.2: F-measure plots to determine optimum kernel size and threshold Algorithm was developed in Matlab in collaboration with Boguslaw Obara to track cells and measure cell motility. 8 videos were randomly selected based on a list of all the videos from proteins observed in the proteomic experiments. 2 frames were selected from each video, one from frames 1-45 and the other from frame 46-90. Here an example is shown of determining optimum kernel size and threshold. (A) First centroids were manually detected and marked on the selected frames. This was then used as the optimum template to which to compare the efficacy of the algorithm to detect centroids. Then varying values of kernel size and threshold were inputted into the algorithm in Matlab and to generate centroids for each frame. (B) F-measure statistics determined the values which best represented the match between manual detection and algorithm detection. Red represents a high F-measure value close to 1, thus a good match between the manual detection and algorithm, detection. Blue represents a low F-measure value, thus centroids are not being identified by the algorithm. The average optimum kernel size was 12 and the average optimum threshold was determined to be 0.020. We wanted to avoid false positives, thus implemented a more conservative threshold value of the mean plus one standard deviation equaling 0.025.

algorithm, involved merging tracks, thus this could accommodate for certain broken tracks. The optimum tracing radius value was established as being 8 pixels.

For the merging procedure, the same visual iterative approach was adopted to identify the optimum merging radius. The merging radius was determined as being 15. This could be significantly larger than the tracing radius, given that there are fewer centroids in corresponding frames which could potentially be detected and joined. The number of frames that could be skipped to maintain a trace was established as 2, which was relatively conservative, but necessary for good trace formation given the variable fluorescence and poor signal in many cells.

4.2.3 Validation of cell tracking algorithm

Multiple quality controls and thresholds were established to try optimise the efficacy of the algorithm to detect and track cell motility. After establishing a fixed value for all the variables as discussed in the previous section, a subset of 60 videos were analysed to assess whether the algorithm worked properly. Close inspection of the results suggested the algorithm worked in correctly identifying cells and tracking them. This was further verified through visual analysis of the 3D plots and comparing them to the time-lapse movies.

Subsequently the algorithm was run on all 192,006 time-lapse movies in the MitoCheck database in order to gain speed and distance values for every video and thus every gene in the human genome that was analysed by the MitoCheck consortium. In order to further validate the accurate identification and tracking of the algorithm, 10 random time-lapse movies were chosen from the entire dataset. From each movie, 3

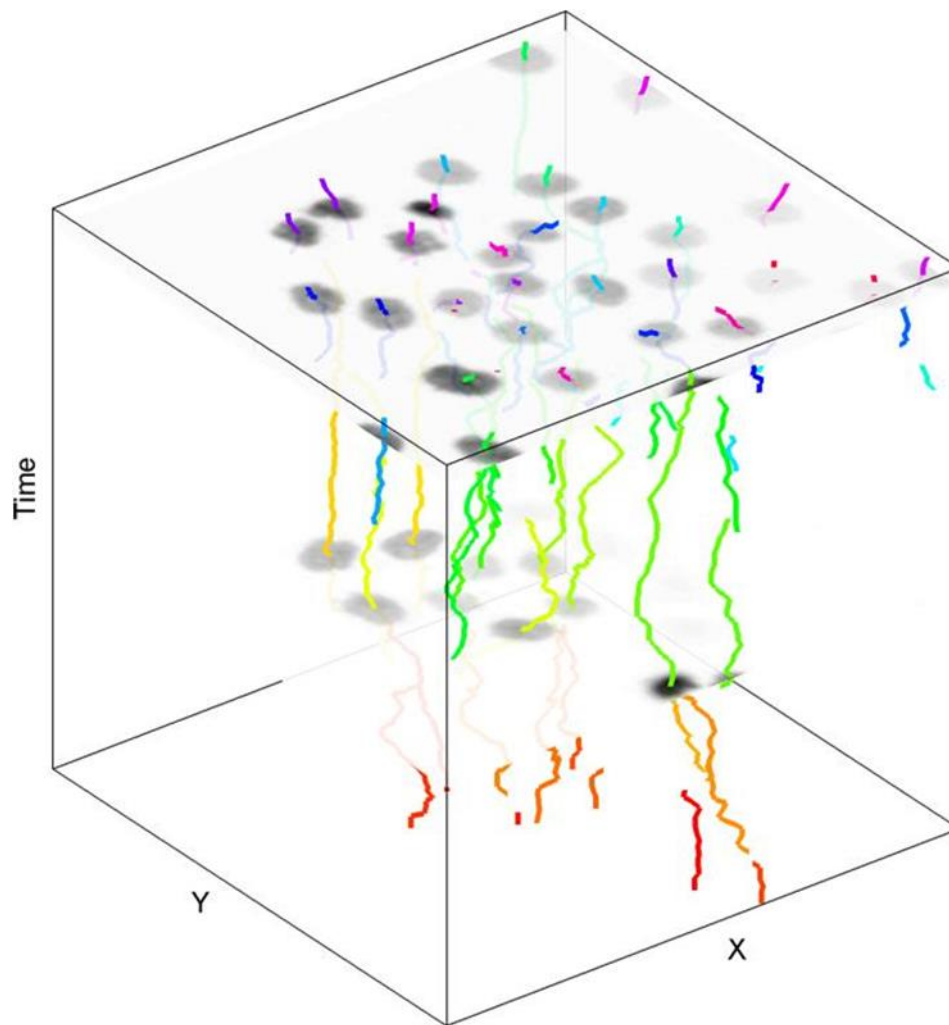


Figure 4.3: Visulisation of MitoCheck videos: After determining centroid positions, cells were tracked by forming traces between frames for every centroid. Optimum radius for track building between consecutive frames was determined through manual visual iterative process. In this figure we see the visualization of this process, with X and Y axes representing the coordinates from the frames and the z axis representing time. 2 frames from the time lapse movie have been transposed on to demonstrate cell position relative to trace. The colour of the traces represents the age of the trace, with red and orange traces originating from the early frames and blue and violet representing traces that are identified late on. In this example the final step of the algorithm, of merging traces has not been established.

frames were randomly assigned and manual identification of cells was undertaken using the same method used to establish the gold standard for threshold and kernel size optimisation. F-measure values were all above 0.9 for the 30 frames with manually identified centroids.

However, there were certain movies that demonstrated unusually high intermittent speed values. Figure 4.4 demonstrates the issue found with a subset of these videos, which was presumed to represent an unstable component within the setup of the confocal microscope resulted in erratic movement. The hallmark of this effect was that all cells within the field of view exhibited a co-ordinate movement profile, moving up to 20 pixels or 12.9µm within a single frame. Movies with these abnormal shifts were identified by setting a threshold of 5 pixels common movement. It transpired that only 194 time-lapse movies had this issue and they all shared the ID ValidMitosisSon384Plate07_03, suggesting that it was a single experiment at a specific location that had unexplained erratic movement. These movies were corrected as described in Section 2.5.7 and gave a useful insight into difficulties with high-throughput analysis. This further reinforces the effectiveness of our curation and validation approaches and can be considered by others when utilising the database.

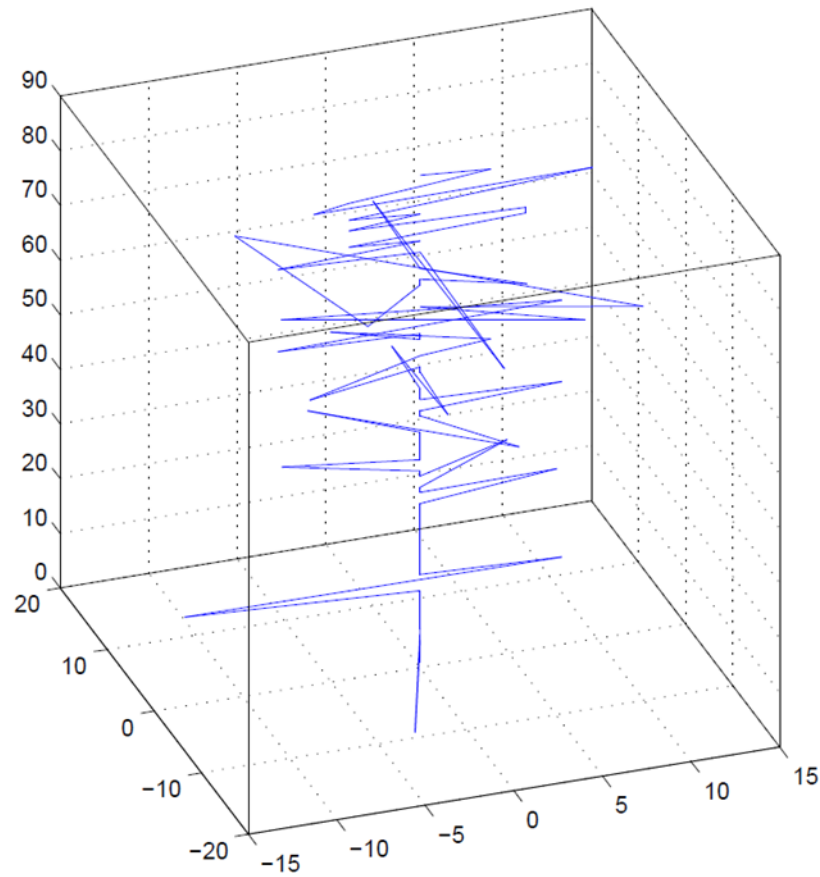


Figure 4.4 Validation and Frame Shifting: Initial inspection of the tracked videos revealed a subset of videos had extremely high speed values. With closer examination of the time lapse movies, it was established that there was an internal co-ordinated shift of the entire frame which gave all the cells a high speed value, with all cells moving up to 15 pixels between a single frame. It was identified that this was a subset of videos with the same location and date ID, indicating that perhaps the microscope stage for the single experiment had some unusual erratic movement. Above is an example of one of the videos and the internal shifts. All the movies subsequently underwent cross-correlation and auto-correlation analysis to correct for any noise and unwanted shift.

4.2.4 Genome-wide siRNA knockdown effect on motility

4.2.4.i Expectations for the results of the genome-wide analysis.

There are multiple possible outcomes from the genome-wide analysis of spontaneous motility from the MitoCheck database. When ranking genes based on their relative motility after siRNA knockdown against their speed value there are several feasible outcomes.

We would anticipate the most likely outcome is the possibility that there is a slope at the beginning of the graph, representing genes which when knocked down reduce motility. This is followed by a plateau, for the majority of genes knockdown of which has no significant effect on motility in the spontaneous single cell model. This plateau is followed by another slope, increasing in velocity, representing a subset of genes which when knocked down individually, result in an increase in cell motility. We would expect the genome-wide motility profile to produce a graph with these characteristics.

This is consistent with the notion that cells are vulnerable to the emergence of hypermotile phenotypes from the loss of function of single proteins, presumably representing regulators that are rate-limiting under our experimental conditions. An important consequence of this idea is that stochastic mutation during oncogenesis *in vivo* may create individual tumour cells that acquire a propensity to metastasise due to a loss of a rate-limiting regulatory function. A combination of such a loss of a regulatory function with changes in cell-cell adhesion, and expression of secreted enzymes affecting extracellular matrix properties would generate tumour cells that exhibit an invasive phenotype, causing an increase in metastatic behaviour.

Alternative possibilities could include an output of a straight line, with potentially a slight gradient. This would indicate that siRNA knockdown of genes had little to no effect on spontaneous cellular motility, for any of the genes, suggesting that it is a tightly

regulated complex process that is functioning under a redundant system. Another possibility is a decreased slope at the beginning of the graph, which plateaus. This would indicate knocking down a subset of genes results in a decrease in motility, possibly due to interfering with the fundamental functioning of the cell, resulting in a multitude of cellular functions slowing down, including cell motility. The lack of an increase in motility at the upper end would suggest that no single gene could increase motility when knocked down.

4.2.4.ii Results from Mitocheck Speed graph

All 192,006 time-lapse movies were processed using the algorithm and generating a mean speed and a mean distance for cells per movie. The raw output was for distance and speed was in pixels and pixels per frame respectively. Each pixel was 0.645µm and each image were taken every 30 minutes, thus distance values were converted to micrometers and speed values converted to nanometers per minute.

Whilst both speed and distance values were generated by the algorithm, we processed and mined the data using the speed values, principally because the mean speed value is independent of track length. Mean distance is the mean length of the tracks in the 3D representations and this could be affected by experimental issues such as weak fluorescence, poor signal to noise, unusual mitotic behaviour and erratic movement of the stage. A problematic example is the weak signal of a significant proportion of cells causes cells not to be identified in all the frames due to the proximity of the signal to the threshold. This would generate a large number of very small broken tracks, biasing the mean track length and reducing the overall mean. Conversely, speed is calculated for every trace formed between two frames and is thus independent from track length. This data could be mined further, particularly by tracking the heritability of mitotic or migration

phenotypes. Additionally, the nature of the motility could be segregated further into nuanced different phenotypes.

Whilst Figure 4.5 appears to be a very simple plot, it is the culmination of 2 years of work and represents to our knowledge the first reported complete genome-wide analysis of spontaneous cell motility in mammalian cells. The genes have been ranked based on their speed value after siRNA knockdown. The mean speed values range from $39\text{nm}\cdot\text{min}^{-1}$ to $78\text{nm}\cdot\text{min}^{-1}$ with the overall mean speed of HeLa cells being $52\text{nm}\cdot\text{min}^{-1}$, with a standard deviation of $3.66\text{ nm}\cdot\text{min}^{-1}$ and standard error of $0.027\text{ nm}\cdot\text{min}^{-1}$. The profile is as predicted, with the vast majority of genes not having a considerable impact on motility and 90% of the genes having a corresponding speed of 46.5 to $58.5\text{nm}\cdot\text{min}^{-1}$ when knocked down by siRNA. The first 5% of genes have a corresponding speed value which is significantly reduced compared to the mean. This reduction could be due to the knockdown affecting cell motility directly, or interfering with the general metabolic functioning of the cell, which could reduce movement or causes the reduction of all cellular processes. Conversely, the 5% at the other end of the graph represent hits which have an increased cell speed when these genes are knocked down. These represent genes that may have a regulating role in cell motility and the knockdown of these proteins result in the cell exhibiting a hypermotile phenotype.

These values are the mean speed values based on multiple siRNAs each with multiple time-lapse movies associated with it. Subsequently, there is a significant amount of variation in the raw outputs for the movies.

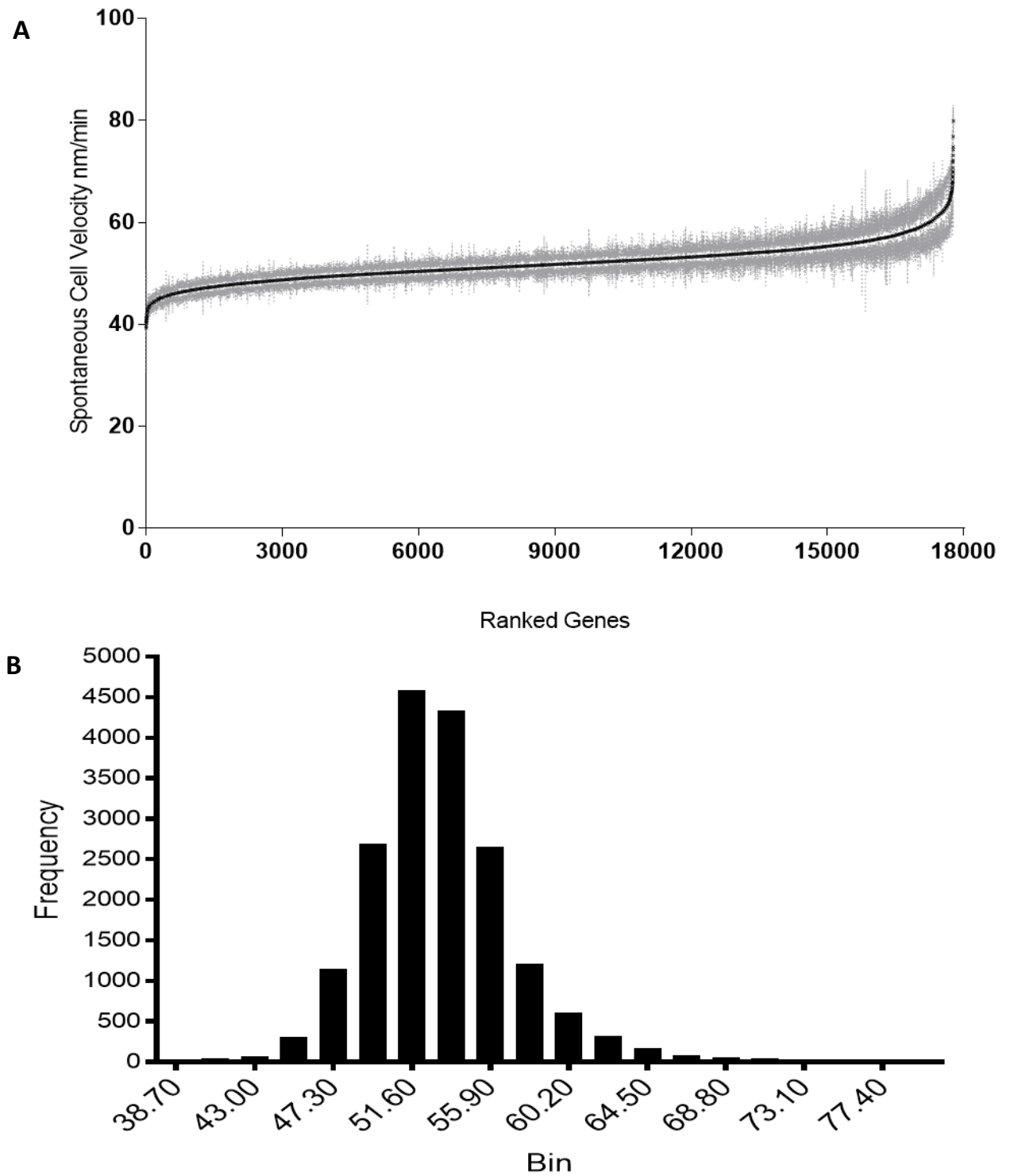


Figure 4.5 Genome-Wide Analysis of Cell Motility : (A) Genes are plotted with their mean spontaneous cell velocity when the gene is knocked down by siRNA. The genes have been ranked along the x-axis. The standard error has also been plotted for all the genes. (B) The associated histogram to display the large number of hits near the average of plot A.

Different siRNAs will have differential efficacies of knocking down the gene of interest. Moreover, siRNAs may have off-target effects that one cannot accommodate for or easily investigate in a large-scale analysis. Finally on a cellular level, cells exhibit natural variation in their behaviours and motility is not a perfectly consistent behaviour, even in untransfected cells.

We carried out a two tailed student's t-test on the dataset, to establish which hits have a significantly different mean to the overall mean of 52nm.min^{-1} based on the hits' mean and standard deviation. As one would expect, the number of statistically significant hits increases at the extremes of the distribution. There are 976 hits above the mean and 2145 below the mean which have a p-value less than 0.05 as shown in Figure 4.6. The threshold $p < 0.05$ is a relatively stringent threshold in this instance given the inherent variability one would expect with means based on many movies with different siRNAs. Interestingly, there are fewer hits above the mean, suggesting that it is harder and more variable to consistently increase cell motility. Whereas, knocking down a gene and causing a reduction in motility generates a mean with a tighter standard deviation which is more reproducible.

A comparison that we undertook was to compare the reported MitoCheck migration hits and place them on our profile. They report the top 300 hits that had the highest distance value, as calculated by the Euclidian distance between the cell position in the first and final frame. Sorting this by speed, we plotted the top 20 hits on our speed map (not shown) and established that the majority of these hits according to the MitoCheck list have an increased speed value relative to the mean and a significant proportion (70%) are within the top 5% of genes, exhibiting a hypermotile phenotype. This is an encouraging result, suggesting there is some overlap in our top hits based on results generated from a different algorithm in a list sorted based on a separate metric

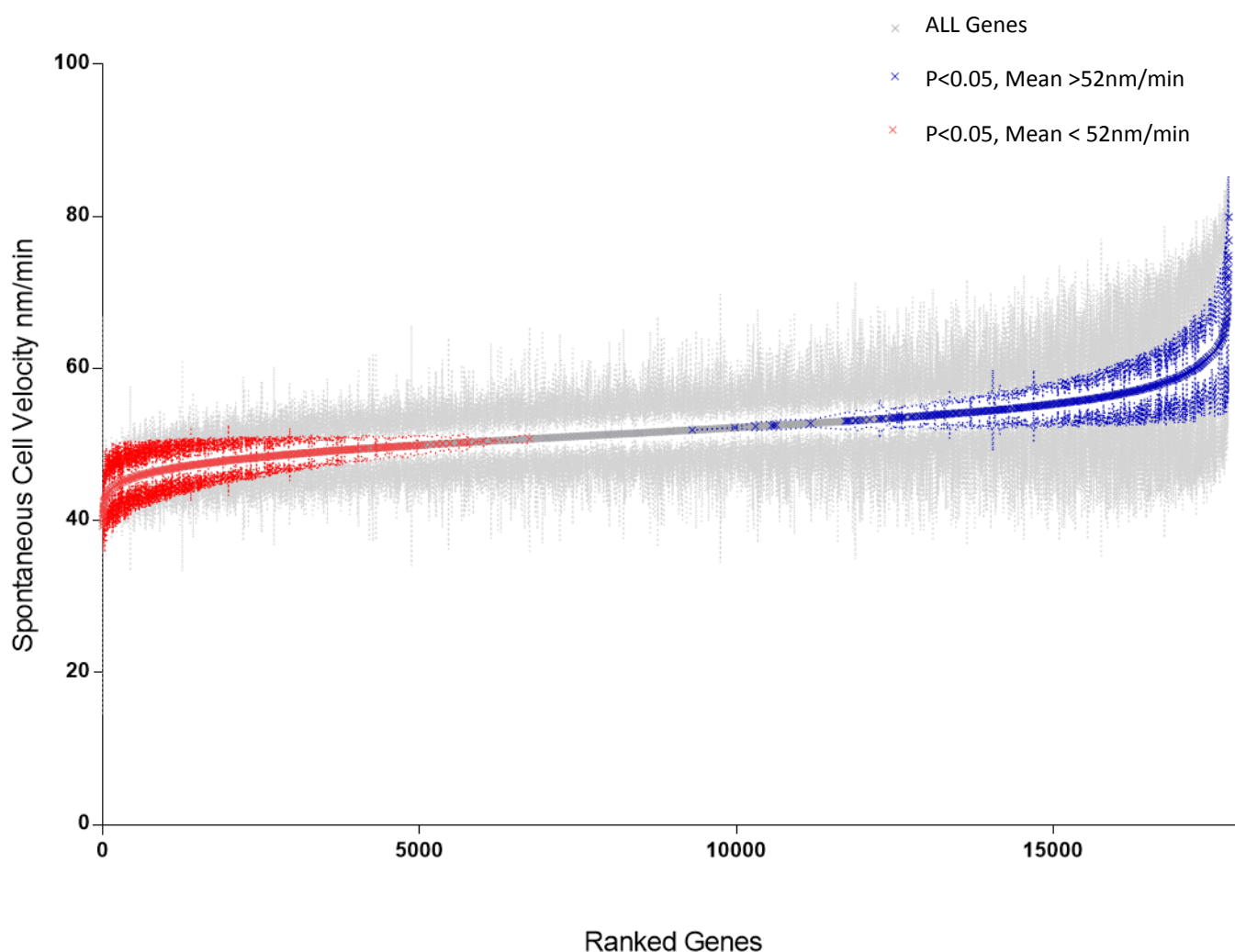


Figure 4.6 Genome-Wide Analysis of Cell Motility with significant hits: Genes are plotted with their mean spontaneous cell velocity when the gene is knocked down by siRNA. The genes have been ranked along the x-axis. The standard error has also been plotted for all the genes. The t-test was performed to identify which hits were statistically significantly different from the mean cell velocity of 52nm.min⁻¹, based upon the mean and standard deviation for each hit. The red hits are those with a p-value less than 0.05 which fall below the mean, whereas the blue hits are those with a p-value less than 0.05 which have a mean cell velocity greater than the overall mean. The background standard error has also been plotted, which is quite variable, due to the effects of the differing siRNAs used to knock down the gene. There are more than 13,000 hits which do not have a significant p-value associated with it.

(distance moved over 48 hours) and goes towards validating our results.

4.2.5 Establishing the Motility Database - Linking Gene Ontologies and KEGG

Pathways to the Motility Database

The Motility Database was curated and annotated further in order to provide more information, functionality and make it easier to search. Figure 4.6 shows two screenshots demonstrating the different components of the motility database. Firstly the database has been annotated to include SwissProt IDs and Accession numbers, Ensembl IDs, RefSeq IDs, and the Gene Description allowing the database to be searched by using a variety of gene nomenclature systems. There is also the siRNA IDs which refer to the siRNA used to knockdown every gene based on the original Ambion siRNA library which utilised the ENSEMBL build 27 to target every gene used by Neumann et al.

The database has also been annotated to include information regarding Gene Ontology terms and KEGG pathways associated with every single gene. This was generated using the Ensembl IDs inputted into the gene ID converter tool (Alibés et al. 2007). The output of these categories are based on KEGG version 27 and Gene Ontology from Ensembl version 49.36k.

Associated with each gene is the mean speed, standard deviation and standard error of the mean speed. The database also consists of information regarding the mean track length and standard deviation. As shown in figure 4.7, each row represents a gene with all the information stated above. The first columns consist of the Gene IDs, pathway terms, description and mean speed as described above. After this there is specific information regarding each individual time-lapse movie that contributes to the mean value, consisting of the mean speed, track lengths and standard deviations for the cells in the movie, as well as the siRNA ID used to knock down the gene of interest.

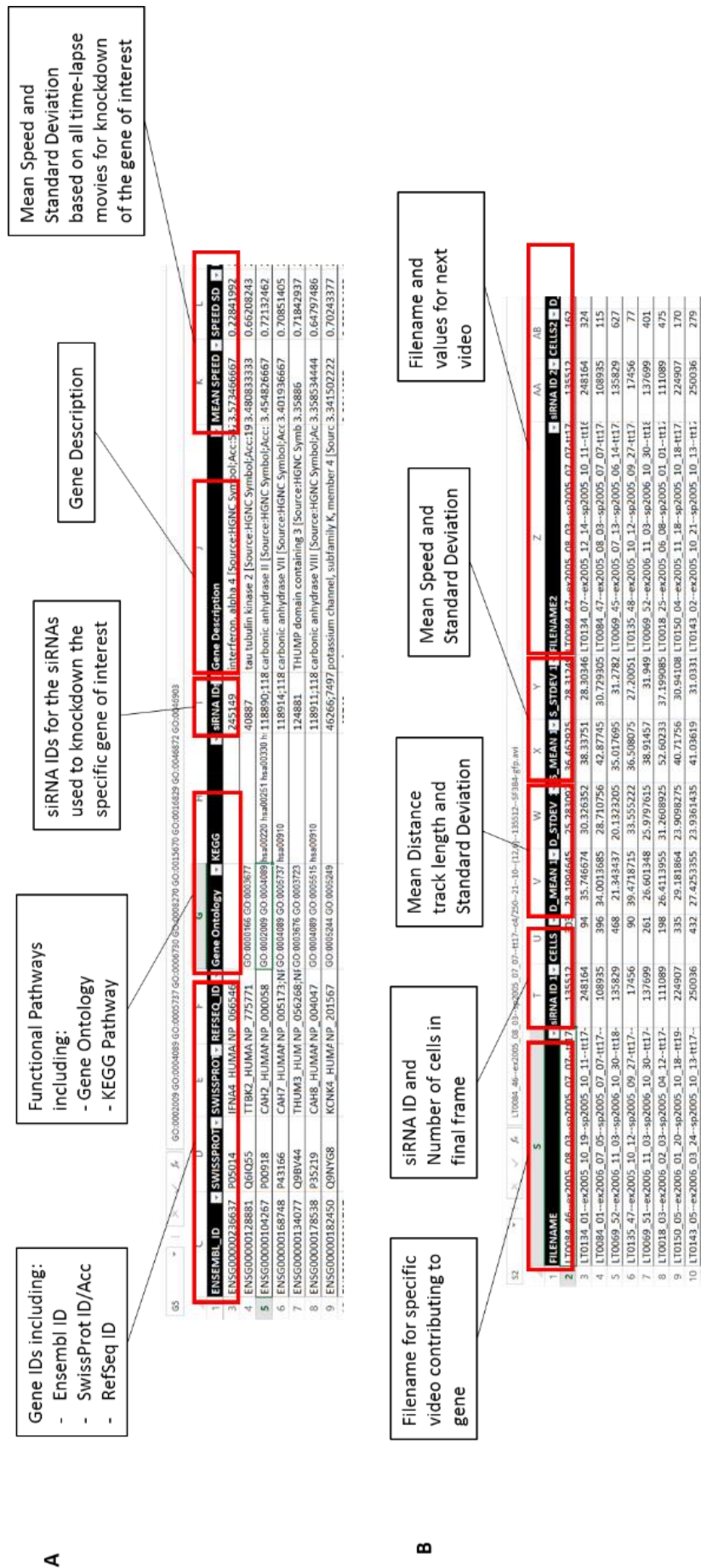


Figure 4.7 Genome-Wide Motility Database: (A) First section of genome-wide motility database consists of overall features for each gene. Each row represents one gene. There are multiple variables associated with it including different Gene IDs, KEGG and GO Pathways to which the gene belongs to, siRNA IDs used to knockdown the gene of interest and the overall mean speed and standard deviation when the gene is knocked down by the different siRNAs. (B) Continuation along the row for each gene is the specific details regarding each time-lapse movie associated with the gene from the Mitocheck database. Here there is the filename of the specific movie, the siRNA used to knock-down the gene and the mean speed and distance of the tracks. This then continues for every single movie associated with the gene of interest which varies from 2 to 16.

NAME	SIZE	Normalised Enrichment Score	NOM p-val	FDR q-val
Nitrogen Metabolism	21	3.18	0.0000	0.000
Gluconeogenesis	51	3.09	0.0000	0.000
Nitrogen Metabolism (HSA00910)	24	3.06	0.0000	0.000
Glycolysis	51	3.05	0.0000	0.001
Glycolysis and Gluconeogenesis (HSA00010)	62	2.88	0.0000	0.001
Renin Angiotensin System (HSA04614)	17	2.69	0.0000	0.006
Pyruvate Metabolism (HSA00620)	42	2.65	0.0000	0.008
Matrix Metalloproteases	29	2.62	0.0000	0.008
Histidine Metabolism	23	2.61	0.0000	0.008
Pyruvate Metabolism	37	2.55	0.0000	0.011
Cell Cycle (HSA04110)	103	2.47	0.0000	0.018
Fatty Acid Metabolism (HSA00071)	44	2.43	0.0015	0.019
MAPK Signalling Pathway (HSA04010)	246	2.41	0.0012	0.021
Beta Alanine Metabolism	27	2.40	0.0000	0.021
Propanoate Metabolism	31	2.39	0.0000	0.020

Table 4.1 Gene Set Enrichment Analysis: (A) Top 20 hits from the Gene Set Enrichment Analysis based on the Normalised Enrichment Score (NES). These represent categories, families and pathways of genes that are enriched in the proportion of genes that when knocked down have an increased motility.

4.2.6 Gene Set Enrichment Analysis of KEGG pathways

We now have a complete database with speed data associated with each of the genes we could identify through NCBI gene conversion 17773 genes. This can be searched to examine hypotheses based on prior knowledge of specific candidates, and inspected manually to identify possible links between biologically interesting candidates and a motility phenotype. However, the most general unbiased approach would be to test for the presence of sets of genes which were associated with increased or reduced motility. The Gene Set Enrichment Analysis tool can be utilised to analyse a list of genes and identify gene sets (categories, families or pathways of a genes) that are enriched or overrepresented at the top or bottom of a list of genes with associated intensity values (Subramanian et al. 2005; Mootha et al. 2003).

We inputted the MitoCheck Gene IDs and speed values and processed them through the GSEA as described in Methods Section 2.5.6. This allowed us to gain an insight into whether there are particular pathways, when perturbed result in an increase or decrease in motility.

The output from GSEA involves a number of important metrics. The Enrichment score represents the degree to which a gene is overrepresented at either end of the list, with high positive scores signifying enrichment at the top of the list and a low negative score indicating enrichment at the bottom of the list. This Enrichment Score is calculated by walking down the list of genes and increasing the “running sum” score every time a gene from a gene-set is present and decreasing the score if it is not within the gene set. The maximum deviation from zero is calculated as being the enrichment score. The normalised enrichment score is accounting for the different size gene sets and is calculated by dividing the enrichment score by the mean enrichment score for all permutations of the dataset.

The nominal p-value determines the statistical significance of the enrichment of a particular gene set and the FDR represents the possibility of the presence of the hit being a false positive, often set in GSEA experiments as 25%.

Table 4.1.A shows the top GSEA categories with the highest Normalised Enrichment Score, (NES) with a nominal p-value below 0.05. Knocking down genes in these pathways or families result in hypermotility. We unexpectedly see a large number of metabolic pathways which are enriched at the top of the list. In particular, Nitrogen metabolism has the best NES. This category is unusually mainly represented by the increased number of carbonic anhydrases, proteins that convert bicarbonate to carbon dioxide and water. (This association of the carbonic anhydrases with nitrogen metabolism may be due to the interplay between the pathways in cyanobacteria). Similarly glycolysis, gluconeogenesis and pyruvate metabolism all score highly as well.

4.2.7 The Carbonic Anhydrase family

Gene Set Enrichment Analysis revealed that the nitrogen metabolism KEGG pathway is overrepresented in the genes associated with a hypermotile phenotype. This was mainly due to the carbonic anhydrases, which have high speed values. This could be due to the carbonic anhydrase enzyme activity being important in regulating motility. To test for this we treated HeLa cells with a general carbonic anhydrase inhibitor, Acetazolamide, to see whether it affects the motility of the cells.

Cells were treated with a serial dilution of acetazolamide ranging from 100 μM to 1 μM . The hypothesis was that if carbonic anhydrase enzyme activity is important in motility, one should be able to recreate the hypermotile phenotype by inhibiting enzyme function, through a competitive inhibitory mechanism as opposed to the genetic knock down of the protein. Thus one would expect a dose dependent motility increase with increasing amounts of Acetazolamide.

In order to test this hypothesis, cells were treated with Acetazolamide for 48 hours before measuring speed by single cell motility assays, as described in Chapter 2.6.1. Figure 4.8.A shows the single cell motility decrease as the concentration of acetazolamide increases, in what appears to be a dose dependent manner. The maximum single cell motility is the DMSO control, with 0 μM Acetazolamide with a mean speed of $0.08 \mu\text{m} \cdot \text{min}^{-1}$. This reduces to a minimum of $0.0051 \mu\text{m} \cdot \text{min}^{-1}$ at 25 μM of Acetazolamide, with a mean speed of $0.55 \mu\text{m} \cdot \text{min}^{-1}$ when treated with 100 μM Acetazolamide. This decrease may be due to cell toxicity effects of Acetazolamide, interfering with regular metabolic functioning of the cell, resulting in cells moving less. These results are initially contrary to the speed results calculated from the MitoCheck database, as knocking down several of the Carbonic Anhydrases resulted in a hypermotile phenotype. Thus, the hypothesis that loss of carbonic anhydrase enzyme activity confers hypermotility is not

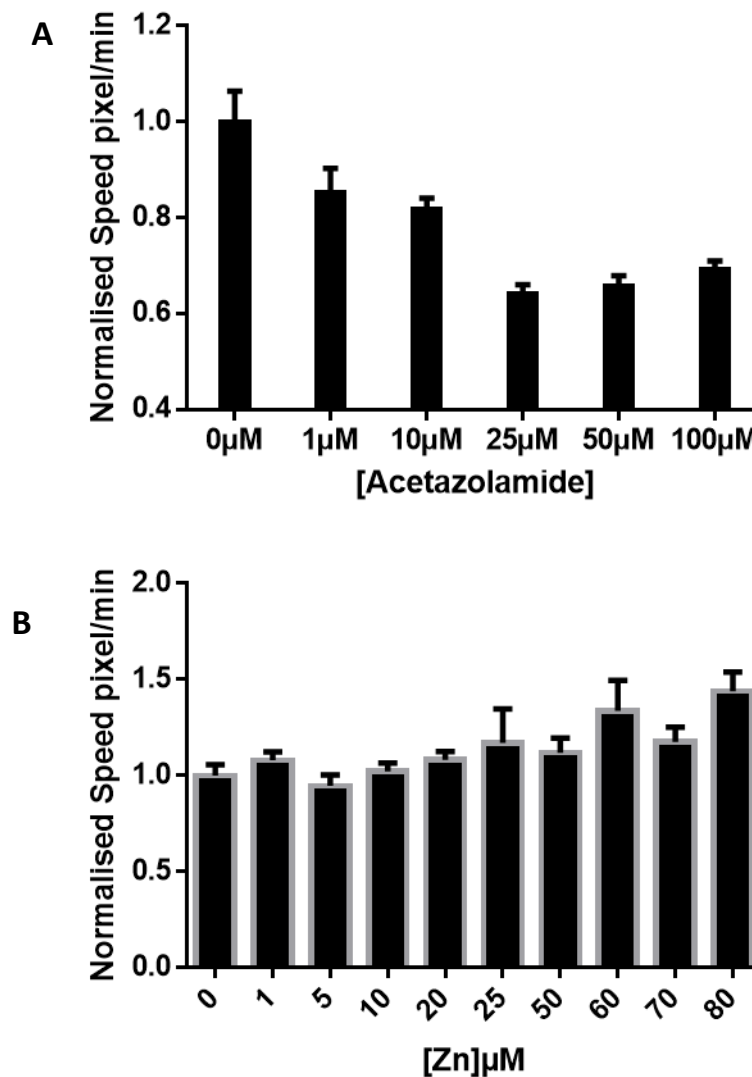


Figure 4.8 The effect of Acetazolamide and Zinc Sulphate on HeLa cell motility: Single cell motility was measured through recording time-lapse movies by confocal microscopy 24 hours after seeding HeLa cells at low density. (A) Graph showing the mean speed of HeLa cells with varying concentrations of the carbonic anhydrase inhibitor, Acetazolamide. Single cell motility decreases as the concentration of Acetazolamide increases. P-values < 0.05 for all concentrations against the control $0\mu\text{M}$ Acetazolamide by student t-test. (B) Graph showing mean speed of HeLa cells after being treated with varying concentrations of Zinc Sulphate. There is an overall trend that the mean speed increases as the zinc concentration increases. P-values < 0.05 for $60\mu\text{M}$ and $80\mu\text{M}$ zinc against control of $0\mu\text{M}$ zinc, based on student t-test.

supported. What, then, could be the explanation for the carbonic anhydrase family appearing so strongly in the GSEA result?

An alternative explanation for this could be that it is an increase in intracellular zinc concentration and not the enzymatic activity of the carbonic anhydrases that has an effect on cell motility. Most of the Carbonic Anhydrases have an active site that contains a zinc ion, thus knocking down any member of the carbonic anhydrase family may increase the concentration of free zinc in the cell. Such an increase in zinc concentration may affect signalling pathways and subsequently have an effect on cell motility.

In order to test this hypothesis we set out to measure motility in cells exposed to varying levels of extracellular zinc, which have been shown to be reflected in changes in intracellular zinc concentrations (Taylor et al. 2012). Varying concentrations of zinc sulphate were prepared in water and cells were treated with relevant concentrations for 48 hours. Single cell motility was recorded and measure as described in Chapter 2.6.1.

Results show that there is an overall trend that increasing the cellular concentration of zinc ions results in an increase in cell motility. This is indicative that intracellular levels of zinc may have an effect on regulating cell motility, which is consistent with the notion that knockdown of carbonic anhydrase may result in an increased level of intracellular zinc, subsequently effecting cell motility. While this hypothesis requires further testing, these preliminary results demonstrate the potential use and interest of the MitoCheck motility database, as it contains a wealth of information which should be independently mined or utilised for a wide range of cell biological studies.

4.2.8 BRCA1 Candidates linked to the MitoCheck Database

In Chapter 3 we identified a number of potential candidate proteins that may be BRCA1 E3 ubiquitin ligase substrates. To determine which of these may be implicated in cell motility, we developed an algorithm to track cells and measure cellular speed. This

gave us a genome-wide motility profile, demonstrating the motility of cells when every gene is individually knocked down.

Tables 4.2.a and 4.2.b display information about our candidates generated from Chapter 3, coupled with their relative speed values calculated from analysing the MitoCheck database. Table 4.2.a contains all the hits that have a mean speed value greater than the average. Within this subset there is a number of interesting candidates, as further demonstrated by Figure 4.9 which displays the candidates on the speed graph. In particular, kinesin, ezrin and fermitin are hits that have established links in motility (E D Coene et al. 2011; Ren et al. 2012; Shi & Wu 2008; Loyo et al. 2011), appear in our proteomic screen and have an increased speed values of 56.14, 54.55 and 55.04nm.min⁻¹ respectively. Moesin, which was a very robust proteomic hit, which is known to have a role in motility, has a relatively modest speed value of 51.75nm.min⁻¹.

Table 4.2.b shows candidates with a reduced speed value relative to the mean. Within this category, there are also several hits that are of significant interest, given their established role in motility and their occurrence in the output of the proteomic study (chapter 3). These include delta-catenin, cofilin-1, and coatomer subunit beta with mean speed values of 44.95, 48.2 and 40.66 nm.min⁻¹ respectively.

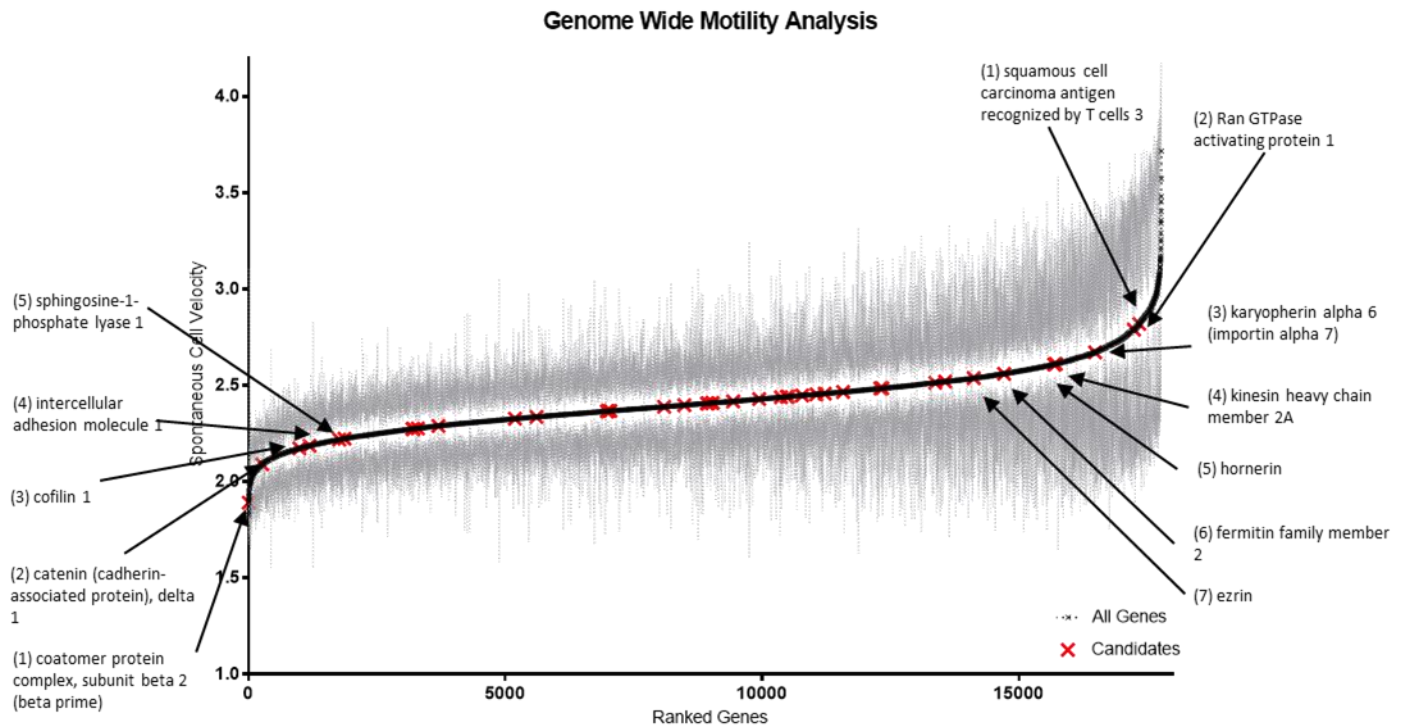


Figure 4.9 Candidates overlaid on genome-wide motility graph: The genome-wide motility graph with candidates overlaid demonstrates that the possible substrates are distributed throughout the genome-wide motility profile. Some of the candidates which are motility outliers are labelled on the graph.

Rank	UniProt ID	Gene Description	Mitocheck Speed nm/min	Proteomics (UbiQ) Category	Proteomics (UbiQ) Mean Log2 Fold Change	Proteomics (GX41) Peptide
17754	P35579	myosin, heavy chain 9, non-muscle	67.69	1:2	-0.23	
17342	Q15020	squamous cell carcinoma antigen recognized by T cells 3	60.62	2:0	3.00	
17242	P46060	Ran GTPase activating protein 1	59.98	1:0	5.58	
17218	Q14103	heterogeneous nuclear ribonucleoprotein D	59.90	2:1	0.25	K.FGEVVDCTLK(+114.04)LDPIGRSR
16481	O60684	karyopherin alpha 6 (importin alpha 7)	57.45	1:0	5.21	
15718	O00139	kinesin heavy chain member 2A	56.14	1:0	4.93	
15687	Q86V23	hornerin	56.10	1:0	2.58	
14709	Q96AC1	fermitin family member 2	55.04	1:0	3.51	
14120	P15311	ezrin	54.55	2:1	0.33	EAQDDLVTIK(+114.04)TK
9437	O00410	importin 5	51.96	1:0	2.58	
9218	P08240	signal recognition particle receptor (docking protein)	51.88	0:1	-2.77	GTGGMFGMLK(+114.04)GLVGSS
9035	O43707	actinin, alpha 4	51.79	2:0	1.45	
9029	Q9UM54	myosin VI	51.79	1:0	3.47	
9014	P21980	transglutaminase 2	51.78	1:2	-1.73	ILGEPK(+114.04)QR
8950	P26038	moesin	51.75	3:0	2.91	LLPQRVLEQHLINK
8925	P60981	destrin (actin depolymerizing factor)	51.74	1:0	3.56	
8487	P07437	tubulin, beta class I	51.54	0:3	-0.84	
8094	P46940	IQ motif containing GTPase activating protein 1	51.37	2:1	0.78	SLNIK(+114.04)TDPVDIYS
7840	Q9NQC3	reticulon 4	51.26	0:1	-7.24	LPSDTEK(+114.04)EDS
7331	P22102	phosphoribosylglycinamide formyltransferase	51.02	1:0	4.20	NLIESMQINGSLKNGSLTNHFEKK(+114.04)
7228	P06748	nucleophosmin (nucleolar phosphoprotein B23, numatrin)	50.98	0:2	-4.96	SIRDTPAK(+114.04)NAQS
7181	Q99460	proteasome (prosome, macropain) 26S subunit, non-ATPase, 1	50.95	0:1	-1.36	KEEK(+114.04)EK
7129	P26373	ribosomal protein L13	50.93	2:1	0.53	YHTK(+114.04)VR
7045	Q04637	eukaryotic translation initiation factor 4 gamma, 1	50.89	2:0	1.72	
7001	P53618	coatomer protein complex, subunit beta 1	50.87	1:0	5.76	EAADPLASK(+114.04)LINK
6976	Q96EY7	Pentatricopeptide repeat domain 3	50.86	3:0	3.49	
6975	Q01518	CAP, adenylate cyclase-associated protein 1 (yeast)	50.86	1:0	0.17	K(+114.04)PGPVVKEMNDAAAFYTNRVLYEYK(+114.04)
6967	P30050	ribosomal protein L12	50.86	0:1	-2.00	IGPLGLSPK(+114.04)
6265	P27824	calnexin	50.54	1:0	4.67	QTSQMEYK(+114.04)KTDAPQ
6089	Q29RF7	PDS5, regulator of cohesion maintenance, homolog A (S. cerevisiae)	50.47	1:0	0.25	LFQSK(+114.04)

Table 4.2.a BRCA1 substrateCandidate Table with MitoCheck and Motility Data: Data has been ranked on the basis of cell speed from the MitoCheck output. Some additional candidates have been considered on the basis of their MitoCheck speed. Here the faster hits are shown.

Rank	UniProt ID	Gene Description	Mitocheck Speed nm/min	Category	Proteomics (UbiQ) Mean Log2 Fold Change	Proteomics (GX41) Peptide
5825	P10809	heat shock 60kDa protein 1 (chaperonin)	50.34	0:1	-1.11	DVK(+114.04)FGADARALMLQGVLDLADAVAVTMGPG
5707	Q72406	myosin, heavy chain 14, non-muscle	50.29	0:1	-4.80	LTAEIWK
5651	P04406	glyceraldehyde-3-phosphate dehydrogenase	50.26	0:3	-0.72	VVDLMAHMASK(+114.04)
5605	P08238	heat shock protein 90kDa alpha (cytosolic), class B member 1	50.23	3:0	2.23	
5380	O60568	procollagen-lysine, 2-oxoglutarate 5-dioxygenase 3	50.13	0:1	-2.80	K(+114.04)FVQSGSL
5187	Q14258	tripartite motif containing 25	50.03	3:0	2.57	
4768	Q15083	WD repeat domain 1	49.80	1:0	3.67	WDSGSSVGELTGHK(+114.04)
4670	Q10567	adaptor-related protein in complex 1, beta 1 subunit	49.75	0:1	-0.56	VFFVK(+114.04)
4273	Q9UHB9	signal recognition particle 68kDa	49.55	1:2	-0.60	TLNFK(+114.04)MGNH
4232	Q96RT1	erbB2 interacting protein	49.53	0:1	-1.43	IYDILSDNGPOQPSTTVKITSVDGK(+114.04)
3913	P27816	microtubule-associated protein 4	49.35	0:1	-2.99	Multiple peptides
3697	O00629	karyopherin alpha 4 (importin alpha 3)	49.24	1:0	6.62	
3401	P35268	ribosomal protein L22	49.03	2:0	6.54	KLVVYK(+114.04)GGKK
3360	O95782	adaptor-related protein in complex 2, alpha 1 subunit	49.00	1:0	0.42	KK(+114.04)GPGAGSALDDGR
3297	P50402	emerin	48.96	1:0	6.29	
3264	Q9P2E9	ribosome binding protein 1 homolog 180kDa	48.94	0:1	-0.30	K(+114.04)QLVAREQET
3207	P52294	karyopherin alpha 1 (importin alpha 5)	48.90	2:0	1.96	
3204	P62249	ribosomal protein S16	48.90	3:0	4.45	
3195	P50570	dynamitin 2	48.89	1:0	2.51	
2505	Q562R1	Beta-actin-like protein 2 OS	48.36	0:2	-7.78	
2312	P53621	coatamer protein complex, subunit alpha	48.20	1:2	-0.59	LNDLIQRLQCYQLTVVGK(+114.04)FEE
2174	Q9Y678	coatamer protein complex, subunit gamma	48.10	0:2	-1.41	ATFYUNVLEOK(+114.04)
1867	P62899	ribosomal protein L31	47.80	2:0	4.98	
1768	O95470	sphingosine-1-phosphate lyase 1	47.69	2:1	0.70	
1188	P05362	intercellular adhesion molecule 1	47.01	1:0	4.32	
991	P23528	cofilin 1 (non-muscle)	46.71	1:0	1.76	SKMIYASSK(+114.04)DAIKK
871	P40227	chaperonin containing TCP1, subunit 6A (zeta 1)	46.52	1:0	1.76	Multiple peptides
856	Q9NT15	PDS5, regulator of cohesion maintenance, homolog B	46.49	1:0	1.80	SALLPVLHHKSK(+114.04)
304	P60709	Actin, cytoplasmic 1 OS	45.12	0:3	-3.63	
266	O60716	catenin (cadherin-associated protein), delta 1	44.95	2:0	2.18	K(+114.04)PLEE
125	P49368	chaperonin containing TCP1, subunit 3 (gamma)	44.13	0:1	-0.86	P.RIVLDSSLEYK(+114.04)G
99	Q00610	clathrin, heavy chain (Hc)	43.92	1:0	0.85	Multiple peptides
6	P35606	coatamer protein complex, subunit beta 2 (beta prime)	40.66	1:0	1.83	

Table 4.2.b BRCA1 Candidate Table with MitoCheck and Motility Data: Data has been ranked on the basis of cell speed from the MitoCheck output. Some additional candidates have been considered on the basis of their MitoCheck speed. Here the slower hits are shown.

4.3 Discussion

In this chapter we report the successful development of an algorithm in collaboration with Boguslaw Obara that can track all cells in images from a large public database of time-lapse movies collected by the MitoCheck consortium. In addition we have used this algorithm on over 192,000 time-lapse movies so as to establish a genome-wide motility database, which has been curated to contain functional and pathway analysis data. Finally we have linked our BRCA1 candidates generated from Chapter 3 to the genome-wide motility graph, in order to determine hits which may be involved in motility.

4.3.1 Developing an algorithm and tracking cells

The algorithm to track cells from the MitoCheck database was developed in collaboration with and written in Matlab by Dr. Boguslaw Obara. Cell motility is a highly complex, fundamental behaviour and is important in a range of biological processes such as embryogenesis, neurone development, wound healing and metastasis. In the past 10 years, with the improvement of confocal microscopes and emergence of significantly more time-lapse movies, automated cell tracking has become an extremely important tool to study these biological processes. Multiple software and plugins are available to track objects, using such image analysis platforms as Imaris, Metamorph and ImageJ. However, we decided to develop our own algorithm given the large-scale nature of the data, the specific form of the movies (i.e. nuclear fluorescence rather than phase contrast images for cell identification) and the fact that most of the existing software tools were only semi-automated, requiring a 'point and click' identification of cells, at least in the first frame.

The algorithm was able to effectively identify cells by optimising the threshold and the initial kernel size based on a subset of time-lapse movies as demonstrated in Figure 4.2. Cells were tracked through identifying the optimum radius within which to search for

the centroids between frame n and frame $n+1$. Finally tracks were merged to join broken tracks, based on another optimised radius as described in Section 4.2.2. A series of iterative selection processes was used to optimise the values for the parameters used by the algorithm, and the method was validated at each step by reference to data obtained by manual identification and tracking from random samples of images and movies.

The algorithm was designed to be flexible and open ended, and the internal dataset of cells and tracks from all movies is now available for more complex analysis of subtle movement behaviour in the future. For example, the version of the algorithm used for the analysis described here does not track and establish relationships between parent-daughter cells after mitosis, and so there is either two new tracks that appear in the frame after mitosis. Alternatively if one of the centroids is within the search radius, an extended track will be present alongside a newly formed track. This additional feature would allow for further analysis including tracking heritability of nuclear, migratory and mitotic phenotypes. Since the bifurcation of a track defines mitosis, it is also possible to measure speed of each cell in relation to an estimate of stage of the cell cycle; the start of a daughter track represents early G1 and the end of a track that is immediately followed by a second bifurcation represents the end of G2.

The MitoCheck data is a superb resource, but the experimental design was not focussed on interphase motility. Thus, there are inevitable trade-offs related to limitations in the available image data. For example, the movies are focused on nuclear identification via GFP-tagged histone H2B, without the capturing of phase images. Consequently, we are reliant on using nuclear movement as a proxy for cellular movement; this chapter demonstrates that this does not prevent analysis of motility phenotypes. However, there are a significant number of cytoplasmic-specific motility features, such as formations of pseudopodia and filopodia, which cannot be detected in the existing data. For this reason,

we are unable to identify potentially relevant hits that affect these aspects of cell movement. However, tracking the nucleus give us a reproducible, consistent means to measure motility.

4.3.2 Genome-wide analysis

Having tried and tested the algorithm and confirming that it successfully managed to track cells, we proceeded to analyse all 192,006 movies in the MitoCheck database. The outcome of this can be captured in a very simple figure (Figure 4.5) which summarises a vast amount of information. It is the genome-wide analysis for cell motility in HeLa cells. The genes are ranked by speed and demonstrate that the majority of genes do not have a direct effect on motility, but a small proportion result in a significant decrease in cell motility and another subset when knocked down result in increased motility. A number of interesting genes that have previously implicated in motility and invasion appear in the top 0.7% hypermotile section of the graph including Myosin-9 (Even-Ram et al. 2007), several members of the Mitogen Activated Protein Kinases family, several integrin proteins and GAP junction proteins. Integrin $\beta 5$, has a high speed value of 66.81nm.min⁻¹ and is in the top 0.15% of hypermotile hits. Genetic knockdown of integrin $\beta 5$ has previously been linked with spreading and motility, albeit a reduced motility as measured by wound healing and a decrease in spreading (Bianchi-Smiraglia et al. 2013).

Validation was undertaken through several methods. Firstly, a random selection of frames from randomly selected movies were chosen, manually analysed to identify centroids and then compared to the algorithm centroid identification. Secondly a subset of movies which had unusually high motility were analysed further and discovered to have erratic movement due to an unstable component. These movies were corrected for the

internal shifts. In addition, the original paper by Neumann et al. had supplementary information that included top migration hits. We sorted this table to identify their top speed hits and overlay the results on our genome-wide analysis. Whilst not a perfect match, all the hits have an increased motility in our database and 12 of the hits appear in the top 10% of fastest genes, suggesting there is a correlation between the two sets of results. This database contains motility information in the form of mean speed and displacement of cells for 17,773 genes.

There may be a significant subset of proteins with potentially established roles in motility, which do not have a particularly high or low speed value. This is because the average is a mean of the time-lapse movies for siRNA used to knockdown the gene of interest. Thus efficiency of knockdown will vary between siRNAs and furthermore there may be off-target effects which could have indirect phenotypic effects. In addition, there are many proteins that have been reported to be important in motility and their overexpression is correlated with a metastatic phenotype. The knockdown of the proteins may not be sufficient to elicit a phenotype and is dependent on the genetic context of the cell. But given that HeLa cells are the most used cell line, this database can be a useful resource to gain an insight into whether one's gene(s) of interest have a potential role in motility.

To our knowledge, this is one of the first examples of a genome-wide siRNA study in mammalian cells to assess motility and link it to biological functional categories and pathways. Several genome-wide siRNA studies have been undertaken to examine other features such as in bacteria (Girgis et al. 2007), drosophila (Wang et al. 2006) and *C. elegans* (Cram et al. 2006). There have been a few other studies looking at knockdown and motility. A comprehensive study by Simpson et al. explores the effect of knocking down 1081 genes involved in wound healing (Simpson et al. 2008). Smolen et al.

undertook a genome-wide shRNA knockdown approach, similar to our method, but the migration assay used was a 3D invasion assay using the Boyden chamber (Smolen et al. 2010). This study found 31 candidates which are involved with regulating motility, whose knockdown is associated with an increased motility. When comparing these hits to our study, we find that the majority of hits also have an increased motility in our study and a number of them are in the top 5% of genes that exhibit a hypermotile phenotype, such as NF κ B inhibitor α , RhoA and Tumour protein D52. This is another encouraging aspect finding that a third of their hits have a speed value in the top 10% of hits with the highest speed values in our study. Conversely, we also identified candidate genes that showed hypermotility on knockdown that were not identified in the Smolen study. Several explanations are possible, ranging from the different cell types and genetic backgrounds used, to the more complex three dimensional nature of the motility task in a Boyden chamber assay, which may obscure motility phenotypes visible during spontaneous migration on a planar surface.

4.3.3 Mining the database for interesting hits

We successfully annotated the motility database to contain Gene Ontology and KEGG pathway information that was available for every gene. Moreover, we identified several sets or families of genes that were overrepresented in the hypermotile and hypomotile groups of genes through Gene Set Enrichment Analysis. The results from this were surprising, with a large number of metabolic processes, such as glycolysis, gluconeogenesis and amino acid metabolism enriched in the hypermotile section of the graph. However, there were some more recognisable pathways such as the MAPK signalling pathway and matrix-metalloproteases, which are both associated with motility. The MAPK family are involved in signalling and knockdown of family members have

been associated with increased motility and an invasive phenotype (Klemke 1997). The enrichment of the MMP family is more difficult to explain, as overexpression of MMPs is associated with invasion and motility by the degradation of the extracellular matrix, carving paths through which cells can more easily traverse through. Knocking these genes down in cells positioned in a 3D context would be likely to decrease motility through an ECM but in MitoCheck movies, cells are seeded from solution onto a 2D surface with collagen, upon which cells can move relatively freely. Thus this may be due to secondary signalling effects, or a role for surface MMPs in the turnover of adhesion molecules at the training edge of moving cells.

Pathways enriched in the hypomotile section of the graph include Regulation of the cytoskeleton by Rho GTPase, Cell Adhesion molecule Activity and Ribosome families (data not shown). Disruption of the cytoskeleton and cell adhesion make biological sense given that they are required for efficient cell motility and adhesion, thus knocking down members that positively regulate formation of filopodia/pseudopodia/actin fibres would result in a reduced ability for the cell to move.

In order to demonstrate the utility of our data for identification of pathways involved in cell motility by gene set enrichment, we selected a very significant but wholly unexpected result for further study. We specifically looked at the Carbonic Anhydrase family that significantly contributed to the high GSEA normalised enrichment score to see if we could further establish this unreported set of genes related to motility. If this effect had been limited to CA 2 and CA 9 we would have concluded that the result reflected the known involvement of both of these genes in tumour cell migration phenotypes (Fiaschi et al. 2013; Cianchi et al. 2010). However, the GSEA result identifies eight CA enzymes with different subcellular and tissue distributions as all having an association with motility. Whilst it seemed unlikely that inhibition of carbonic anhydrase activity in any one of these

enzymes would elicit the same phenotype, we tested the hypothesis that this was the case using a potent general inhibitor of carbonic anhydrases. We showed that inhibiting the enzyme activity using the inhibitor Acetazolamide, resulted in a decrease in motility, the opposite of what was observed when knocking down carbonic anhydrase genes. We showed that the increased motility could be due to an increased concentration in intracellular zinc, due to most of the carbonic anhydrases coordinating zinc ions in their active sites. There is a positive correlation of increasing zinc ions to cultured cells and increased motility. There have recently been a few papers that link an increase of intracellular zinc to metastasis (Hogstrand et al. 2013; Taylor et al. 2012), and whilst this is still a novel association, our data would support this hypothesis.

4.3.5 Linking Candidates to Motility

The principal aim of this chapter was to link our candidate proteins generated in Chapter 3 to the MitoCheck database. We successfully managed to overlay the majority of the candidates on the MitoCheck motility graph and identify the most interesting hits. We are particularly interested in the candidate hits that have hypermotile or hypomotile phenotypes, as these hits may shed light on pathways or substrates that are responsible for regulating motility through a BRCA1 mediated pathway.

Genes which result in an increased cellular motility when knocked down are of interest, particular in relation to the corresponding proteins that had been found to be enriched in the cells expressing functional BRCA1 from the proteomic study.

A pathway that would be consistent with these group of genes/proteins would be that BRCA1 interacts with and ubiquitinates protein X, thus the proteomic enrichment. This ubiquitination process has a stabilising or a downstream signalling effect, which has a role to play in regulating cell motility and effectively acting as a 'break'. Subsequent

knockdown of Protein X, would result in the inability of BRCA1 to stabilise or signal through this protein, resulting in a disruption in the regulation of motility and thus increasing overall cell motility. This is one example of a possible pathway which could be tested and validated further through more specific knockdown and migration assays, mutagenesis experiments and ubiquitination assays.

Similarly, proteins that were enriched in WT BRCA1 expressing cells from the proteomic experiments which have a corresponding hypomotile phenotype when the corresponding gene is knocked down could be functionally significant. Particularly interesting candidates which had an increased speed value relative to the mean are ezrin, fermitin-2 and kinesin, as all of these proteins have been associated with roles in motility and cancer. At the other end, delta-catenin, coatamer subunit beta 2 and cofilin all look like interesting candidates due to their reduced motility.

Ezrin is a particularly appealing candidate, given that it was initially established as being linked to BRCA1 in the regulation of cell motility (E D Coene et al. 2011). Moreover, it has established links to metastasis of breast tumours. Some studies suggest that ezrin overexpression may be associated with invasion, contrary to the upregulation observed here, but given the nature of the MitoCheck assay, dysregulation of motility is worth investigating further. Moreover, the influence of ezrin on motility may be context dependent, particularly due to its two differing conformational states. Thus ezrin makes for an exciting prospect.

Fermitin-2, also known as kindlin-2 is a newly established focal adhesion protein associated with integrin activation. Studies have linked dysregulation of fermitin-2 with breast cancers (Zhao et al. 2013; Yu et al. 2013). In particular, a recent study demonstrated that its expression level was reduced in fibrosarcoma cells (Shi & Wu 2008). Furthermore overexpression reduced its invasive capability and *in vitro* studies

demonstrated knockdown of fermitin-2 increased invasiveness based on Matrigel assays. These results make fermitin-2 a particularly fascinating candidate given it is consistent with the results from our MitoCheck motility screen.

Kinesin is a molecular motor protein involved in the transport of cargo along microtubules. Moreover it is associated with mitosis and the depolarisation of microtubules at centrosome, which makes it of particular interest to the original MitoCheck screen (Goshima & Vale 2005). Interestingly, one study showed epigenetic silencing of KIF1A associated with nasopharyngeal which may be of relevance when considering kinesin in the context of malignancy and metastasis (Loyo et al. 2011).

Coatomer subunit beta 2 is associated with Golgi non-clathrin coated vesicles important in retrograde Golgi to endoplasmic reticulum trafficking. Whilst it has no direct links with cell motility or cancer, vesicle trafficking in general is important in the delivery and movement of chemo attractants that can influence cell motility (Kriebel et al. 2008). Cofilin also has strong links with breast cancer metastasis (Wang et al. 2007) and loss of cofilin promotes a mesenchymal-type movement but loss of amoeboid movement, of particular relevance to the EMT (Sidani et al. 2007).

Delta-catenin is part of the catenin family, and has been linked to regulating NF-kB. In addition, dysregulation of delta-catenin has been associated with breast cancer (Schackmann et al. 2013).

These are just a handful of candidates which have an up or down regulated motility relative to the mean, which have relevant roles in motility or cell signalling and trafficking and further investigation may reveal more about their functions.

4.4 Conclusions and Future Directions

In this Chapter we have established candidates that exhibit interesting motility features when their genes are knocked down, through development of an algorithm that can track cells and doing genome-wide analysis of the MitoCheck database. More importantly, we have linked the generated candidates from the proteomic experiments in Chapter 3. Thus we have a list of proteins that are shown to be reproducibly ubiquitinated in cells expressing functional BRCA1 and that have interesting motility profiles when the corresponding gene is knocked down. These could shed light on pathways that are important in BRCA1-mediated motility and go towards explaining why patients with BRCA1 mutations develop aggressive highly metastatic tumours. Both of these studies are large-scale and unbiased, which is of particular significance when trying to identify novel pathways not yet known to have a role in BRCA1-mediated motility.

The next step is to select the most interesting candidates based on their proteomic enrichment in and motility profiles and undertake further migration assays to ascertain their possible relevance in BRCA1 mediated motility. Knockdown of candidates in cells expressing functional BRCA1 and mutated BRCA1 and assessing the motility phenotypes is a logical first step in order to try answer the question.

Of course, the assays in the MitoCheck database are carried out on a 2D plastic surface and thus the phenotypes may not be replicated *in vivo*. Particularly when considering metastasis, many of the steps involved with the initial invasion behaviour is the loss of cell-cell contacts that cannot be observed in these single cell motility assays. Thus one would be hesitant in immediately claiming all the motility-upregulated proteins cause cancer. Nevertheless, a change in this fundamental process of spontaneous motility is one of the first steps in metastasis and thus these proteins are still of interest and deserve

further investigation as they may uncover some motility biology that would not immediately be obvious from clinical data.

Aside from the BRCA1 narrative, we have developed a curated genome-wide siRNA knockdown motility database for HeLa cells that can be used as a resource by others, interested in cancer, developmental biology and cell migration. This work goes beyond the scope of the BRCA1 project and can be a useful resource for anyone studying cell motility.

Chapter 5: Establishing BRCA1's role in Cell Motility

5.1 Introduction

The process of metastasis is complex, involving multiple steps including loss of cell-cell adhesion, increase in cellular motility, secretion of matrix-metalloproteases, intravasation, transport through the circulatory system, and colonisation in a new microenvironment.

Understanding metastasis is vital in the goal to find therapeutics to treat cancers and improving mortality rates, given that 90% of cancer-related deaths are due to metastasis (Hanahan & Weinberg 2011). Treating a primary tumour is feasible through surgical intervention, regional irradiation or systemic chemotherapy, but treatment of a metastasized tumour is substantially more difficult, with a poor prognosis associated with metastasized breast cancer; less than 50% survival rate within the first year after diagnosis (Figueroa et al. 1993).

The initial fundamental step to the metastatic cascade is the ability for a cell or a group of cells to become motile and invasive. Cells dislodging from the primary tumour and invading the surrounding tissue, requires an increase in cell motility and disruption of cell-cell adhesions. There are three established modes of cell motility; collective motility, mesenchymal motility and amoeboid motility, all reported to be involved in tumourigenesis. A cell can switch between these modes. Mesenchymal motility is associated with more fibroblastic features, such as the formation of filopodia and lamellipodia and having an obvious cell polarity. Amoeboid migrating cells have a more morphologically flexible behaviour, with the ability to manoeuvre through extracellular matrix without the requirement of secreting matrix-metalloproteases (Sahai 2007).

5.1.1 BRCA1 role in invasion and metastasis in breast and ovarian cancer

As discussed in Section 1.2.1, patients with BRCA1 mutations are particularly susceptible to developing aggressively metastasizing breast and ovarian cancers. Moreover, BRCA1 expression is absent or decreased in 30-40% of cases of sporadic breast cancers (Wei et al. 2005; Esteller et al. 2000), probably via transcriptional repression involving MTA1 (Molli et al 2008) or HOXA9 (Gilbert et al 2010). Recent observations correlate reduced expression of BRCA1 with metastatic disease and poor clinical outcome in other cancers, including prostatic, colorectal, and pancreatic carcinomas, and uterine leiomyosarcoma (discussed in Bednarz et al 2010).

BRCA1 has established roles in DNA damage repair, transcriptional regulation and cell cycle regulation and these are all important functions in maintaining a healthy cell. Mutations of BRCA1 can interfere with these functions and whilst genetic instability is associated with malleable cancerous phenotypes, there is a growing body of evidence that suggests that BRCA1 may have more direct roles in preventing metastasis through regulation of motility. This has been inferred from studies focusing on the EMT associated transcription factors slug and snail and its inverse relationship with BRCA1 expression levels (Wu et al. 2012) and also from its localisation and interactions at the cell membrane (Coene et al. 2011). This could help understand the clinical observation of the associations of BRCA1 mutations with aggressive, highly metastasizing cancers in patients.

5.1.2 Selection of candidates based on proteomics and MitoCheck results

Whilst our proteomic and bioinformatics screens each generated a large number of interesting hits, we are predominantly interested in the intersection of these two sets of hits; candidates that are BRCA1 ubiquitination targets and also play a role in regulating

cell migration. It was clear that within the scope of the project we could only effectively analyse a handful of these. The candidates were selected based on overall proteomic score, MitoCheck speed and any established links in the literature which demonstrates an association with cell motility, cancer or invasion. Section 4.3.5 identified the 4 proteins delta-catenin, moesin, ezrin and fermitin-2 and discussed their relative merits. Based upon this analysis we decided to start by investigating these candidates through motility assays.

5.1.3 *In vitro* methods of analysing motility

Cell motility and invasion have been studied for many years and multiple *in vitro* and *in vivo* assays have been developed to better understand these processes. Investigating cell motility *in vivo* is an essential step in the pipeline of developing therapeutics to disrupt tumourigenesis. However, they can be time and money consuming making it is difficult to assess the relevance of more than a few candidates. Thus it is a step best suited later on in the pipeline, after *in vitro* experiments have narrowed down the list of candidates and produced results that confirm the importance of the relevant candidates.

In vitro 3D assays traditionally utilise a 3D scaffold to mimic the extracellular matrix, such as Matrigel, in systems such as a transwell chamber. These assays are useful in determining cellular movement in a 3D system, which is closer to an *in vivo* context. However, many of these are endpoint assays, recording cell position and motility at a particular snapshot in time.

Whilst 2D *in vitro* assays are less physiologically relevant, they are highly useful as an initial assessment of cell-cell and cell-substrate interaction. One of the most established 2D motility assays is the single cell motility assay, which is the method we used to analyse the MitoCheck database (Neumann et al. 2010). This tracks the fundamental locomotion

capabilities of a cell and can be used to differentiate the basal level of movement of a cell, in the absence of a specific chemoattractant stimulus for directed motion.

Another useful technique, which highlights additional features in cell motility compared to the single cell assay, is the wound healing (scratch) assay. This involves the making of a wound in a confluent monolayer of cells using a pipette tip and using time-lapse confocal microscopy to record the closure of the wound (Liang et al. 2007; Simpson et al. 2008; Coene et al. 2011). This assay can be utilised to investigate properties of cell-cell adhesion, cell-substrate and collective motility. It also provides regular observation of the cell motility and wound closure and allows unbiased recording of all the cells in your field of view.

In this chapter, we used 2D *in vitro* assays to investigate BRCA1-mediated cell motility, with the intention of honing in on specific relevant candidates to suggest future 3D *in vivo* studies.

5.1.4 Aims of the Chapter

In this chapter our aim is to gain a better understanding of BRCA1-mediated motility, by investigating the involvement of the best candidates from Chapters 3 and 4. For this purpose, 4 approaches were used:

- 1) Identify motility differences between HCC1937 + WT BRCA1 cells and HCC1937 + I26A BRCA1 cells.
- 2) Establish a microscopy pipeline to be able to assess motility and spreading metrics.
- 3) Investigate the top candidate(s) and establish whether they may have a role in motility in these cell lines.
- 4) Propose a model for BRCA1-mediated regulation of motility.

Our hypothesis is that HCC1937 +WT BRCA1 cells will be less motile, more spread, and maintain stronger adhesions to neighbouring cells, thus exhibiting less metastatic characteristics, than HCC1937 + I26A BRCA1 cells. We anticipate that our selected candidates may affect motility when siRNA knocked down *in vitro*. Moreover, we would expect there to be a differential effect on motility dependent on the BRCA1 genotype of the cell, further strengthening the case for said candidate to be a substrate of BRCA1 involved in regulating cell motility.

Our results show differences in motility profiles between cells HCC1937 cells expressing either functional or ubiquitin ligase dead BRCA1. Moreover, our candidate screen revealed multiple candidates that could affect motility in our single cell assay. We focus on ezrin and investigate the effect of knockdown of the protein on cell spreading, single cell motility and wound healing. Results show that ezrin affects cell motility and we propose a model for BRCA1's interaction with the ezrin and the downstream regulation of motility.

5.2 Results

5.2.1 Single Cell Motility assay of HCC1937 cell lines

In order to establish if any differences in cell motility exist between HCC1937 + WT BRCA1 and HCC1937 + I26A BRCA1 cells, and also to identify putative candidates and pathways involved in regulating motility, we needed to establish assays to assess cellular speed. We used the spontaneous single-cell motility approach as our assay to assess motility, as this is effectively the same experimental setup that generated the MitoCheck movies. In addition, Coene et al. also used this single cell assay to establish differences in cell lines expressing functional and mutant BRCA1.

Our first task was to optimise this assay to identify differences between cell lines so that it can be utilised in a high-throughput setup in order to test multiple candidates. There are multiple factors that need to be considered when designing live-cell microscopy experiments and there is typically a balance that needs to be achieved between obtaining the highest resolution images and ensuring the conditions or cellular behaviour do not dramatically alter image quality, for example by photobleaching. In this experiment, a first consideration was ensuring the conditions were as close to our tissue culturing conditions as possible, specifically 37°C and 5% CO₂. Whilst the confocal microscope has a sealed chamber and heating device to ensure a constant temperature, the chamber with a regulated CO₂ flow could not accommodate multi-well plates for automated multi-position timelapse, thus making it difficult to ensure a constant delivery of 5% CO₂ to all the wells equally. Different approaches were undertaken to try overcome this problem, including using media that maintains the correct pH without requiring CO₂. This had unwanted

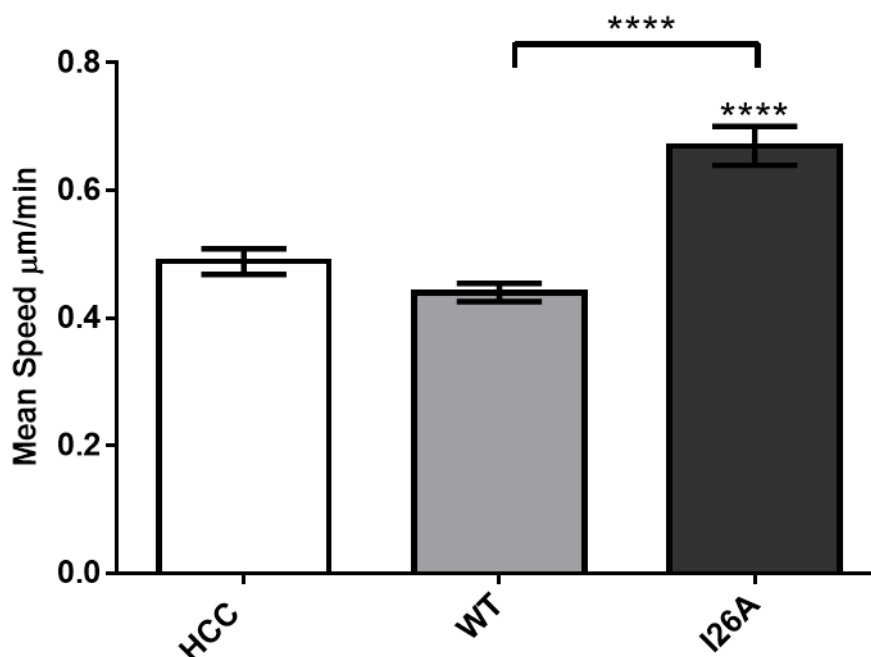


Figure 5.1 Single cell motility assays on HCC1937: (A) HCC1937, HCC1937 + WT BRCA1 and HCC1937 + I26A BRCA1 were seeded and cultured for 24 hours and motility was measured by confocal microscopy. The results are the amalgamated data from 3 repeat single cell motility assays showing mean speed ($\pm$ SEM, $170 < n < 205$). Transfecting HCC1937 cells with WT BRCA1 cause a reduction in single cell motility, whereas transfecting with ubiquitin ligase dead BRCA1 results in a significant increase in cell motility. The notable significant difference is between speed of cells with functional WT BRCA1 and ubiquitin ligase dead I26A BRCA1, with a p -value < 0.0001 . These results are consistent with the motility results from Coene et al. 2011.

secondary effects (Lepe-Zuniga et al. 1987), with cells producing large amount of exosomes, which compromised image quality. In addition, numerous reports have surfaced of HEPES toxicity during live-cell imaging experiments, presumably due to increased free radical formation by the synthetic buffer under the illumination conditions required to visualize fluorescent probes.

In the optimum approach, cells were incubated in 37°C and 5% CO₂ equilibrated media and wells were filled to the top and the plate was sealed with parafilm to ensure minimum gas exchange. Cells proliferated at the same rate as cells cultured in individual MatTek dishes (which could receive a constant CO₂ supply) for 48 hours (data not shown).

Another consideration was being able to ensure that the cells could be tracked in a semi-automated manner to be able to obtain speed and distance measurements with the intent to use for high-throughput multi-well experiments. Manual tracking is not feasible for high cell numbers over a long period of time, thus having the end user in mind is crucial to the design of the experiment. It is difficult to systematically identify cells from phase-contrast images, due to differential intensities of dark and light rings at the cell boundaries. Multiple studies have been undertaken to develop algorithms and software to track cells from phase contrast time-lapse movies (Ambühl et al. 2012; Debeir et al. 2005), but Hand et al. concluded that tracking cells from fluorescence-based images is more efficient and accurate (Hand et al. 2009). Consequently, we treated cells with DRAQ-5 a nuclear fluorescent marker, to allow semi-automated tracking.

Figure 5.1 shows the comparison of the three HCC1937 cell lines. The experiment was setup as described in Section 2.6.1. The results show that HCC1937 cells have a mean speed of $0.49 \mu\text{m}.\text{min}^{-1}$, whereas this reduces to $0.44 \mu\text{m}.\text{min}^{-1}$ for HCC1937 + WT BRCA1. HCC1937 + I26A BRCA1 cells, (with no ubiquitin ligase activity) have an increase in the mean cell speed to $0.67 \mu\text{m}.\text{min}^{-1}$. This profile is consistent with the results Coene et al. report, confirming again the BRCA1 ubiquitin ligase activity is important in regulating cell motility.

5.2.2 Wound healing assay of HCC1937 cell lines

We proceeded to test the HCC1937 cell lines through another well-established motility experiment, the monolayer wound healing assay. The time-lapse movies show that the HCC1937 parent cell line closed the wound within 18 hours after creating it with a p10 tip. The cells exhibit erratic behaviour with leading edge being relatively disordered and individual cells migrating into the wound. A proportion of cells apparently lose cell-

cell adhesions, and migrate over the top of other cells. In addition, many of the cells exhibiting a hypermotile phenotype, do not necessarily migrate into the wound, but rather, move randomly throughout the monolayer, moving over and above adjacent cells. HCC1937+ WT BRCA1 cells, recall with a functional ubiquitin ligase domain, also filled the wound within 18 hours, but have a uniform leading edge (although the figures show time points prior to the close to demonstrate phenotypes). Cells appear to maintain a tight monolayer with the majority exhibiting good cell-cell adhesions. Very few cells migrate independently into the wound relative to the parent cell line. HCC1937 + I26A BRCA1 show a similar but even more exacerbated version of wound closure to the parent HCC1937 cell line. Again, many individual cells have a hypermotile phenotype, migrating in and out of the wound, independent to other cells.

Figure 5.2 shows two snapshots of the time-lapse movies; one frame at the start of the assay and one 10 hours later, with sample cell tracks highlighted to demonstrate the differences. HCC1937 + WT BRCA1 cells move in a relatively linear direction perpendicular to the wound edge and towards the wound with little lateral movement, exhibiting good cell-cell adhesion. In contrast, the HCC1937 + I26A BRCA1 cells have more variable cell movements, some moving independently into the wound, others moving back and forth with a more random directionality. The rate of closure between the three cell lines is not significantly different, but the mode of closure is markedly different. This behaviour may not be captured by certain assays, such as the Boyden chamber, which rely on observing number of cells at a specific snapshot. We tested the use of an automated capacitance measurement system that generates a continuous signal (xCelligence system), but found the results of biological repeats of a single cell type too variable to be reliably used (data not shown).

Figure 5.3 shows the results of quantitating the disorder of the leading edge. These results were obtained through use of a MetaMorph journal developed by Mr Nick White as described in Section 2.6.7. The perimeter variability index (PVI) represents the sum of the absolute change of the leading edge perimeter between every single frame and used as a proxy for the uniformity of the leading edge. This makes the assumption that a ‘uniform’ leading edge is one tending towards a straight line. The greater the PVI, the longer the leading edge perimeter and thus a more disordered leading edge. We see the I26A cell line PVI is 63.1 and is the most disordered, concordant with the visual representatives in Figure 5.2. Then the HCC PVI value is 42.3 and finally the WT is the least disordered, with a PVI of 29.9.

This behaviour suggests that functional BRCA1 is important in regulating cell-cell adhesion and uniform motility. Disruption of the ubiquitin ligase domain causes a loss of cell-cell contact and increases individual cell motility. This is indicative of important features of a cell exhibiting a metastatic phenotype, as one of the first steps is to dissociate from the primary tumour, to have an increase in motility and to invade surrounding tissue. Cells with I26A BRCA1 show a reduced cell-cell adhesion and increased individual migration, suggesting in an *in vivo* context, they may be more effective in invading surrounding tissue.

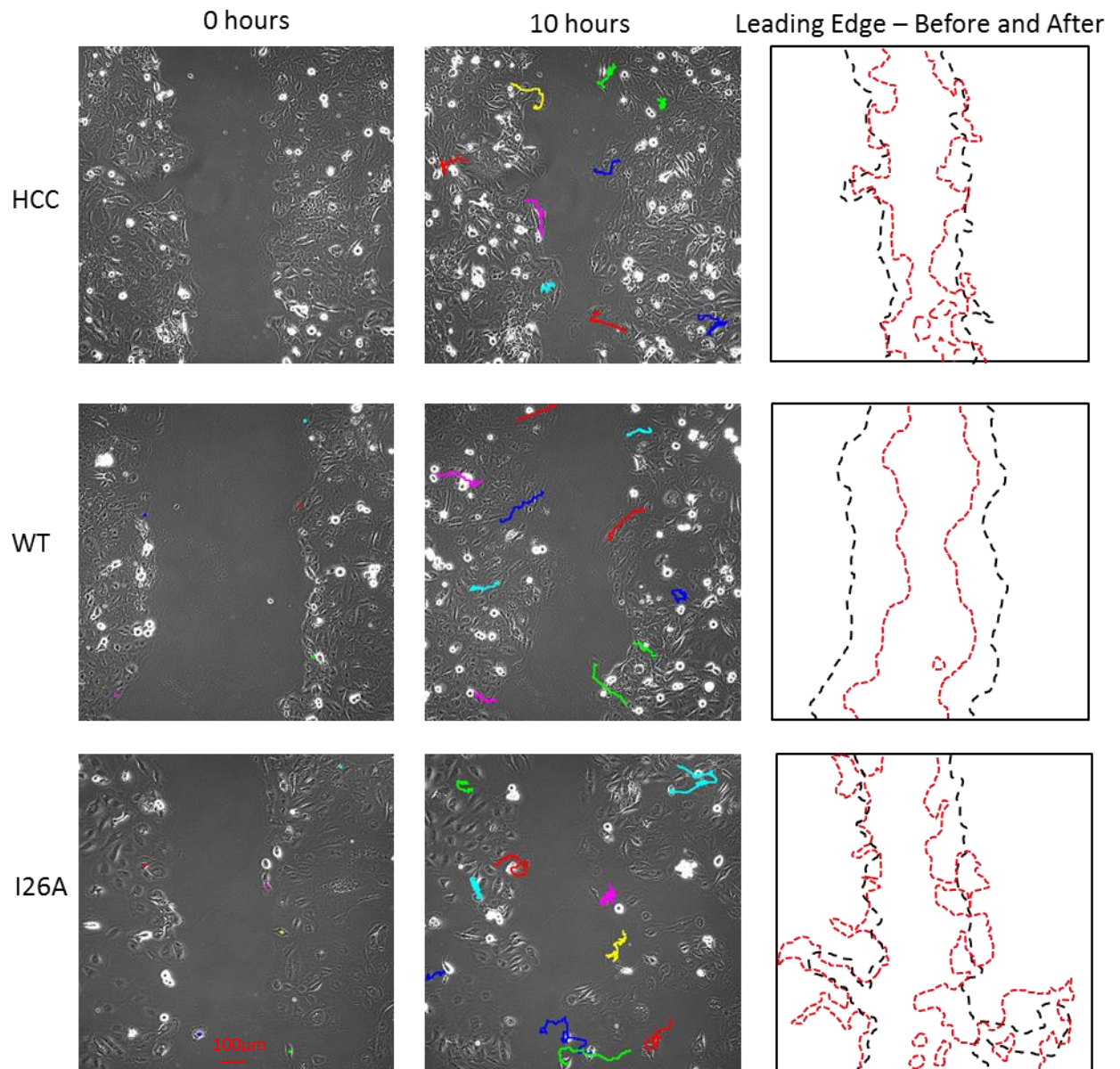


Figure 5.2 Wound healing experiments on HCC1937 cell lines: HCC1937, HCC1937 + WT BRCA1 and HCC1937 + I26A BRCA1 were seeded and cultured for 24 hours until a confluent monolayer was established. Wounds were made using a p10 pipette tip and cells were washed gently 3 times with media. Wound healing assays were run for 12-24 hours until wound closure, with images taken every 10 minutes. Results show representative wound healing images. HCC1937 parent cell lines and ubiquitin ligase dead I26A cell line having a disordered leading edge, with cells moving into the wound and the leading edge becoming more disordered with time, indicative of an abnormal motility phenotype. Rescue WT BRCA1 has a more uniform leading edge with fewer gaps and breaks. 10 cells, 6 from the leading edge, 4 from within the monolayer, have been manually tracked to illustrate typical cell behaviour. The final column illustrates the differences in leading edge, the black line representing the starting frame and red line representing the frame after 10 hours.

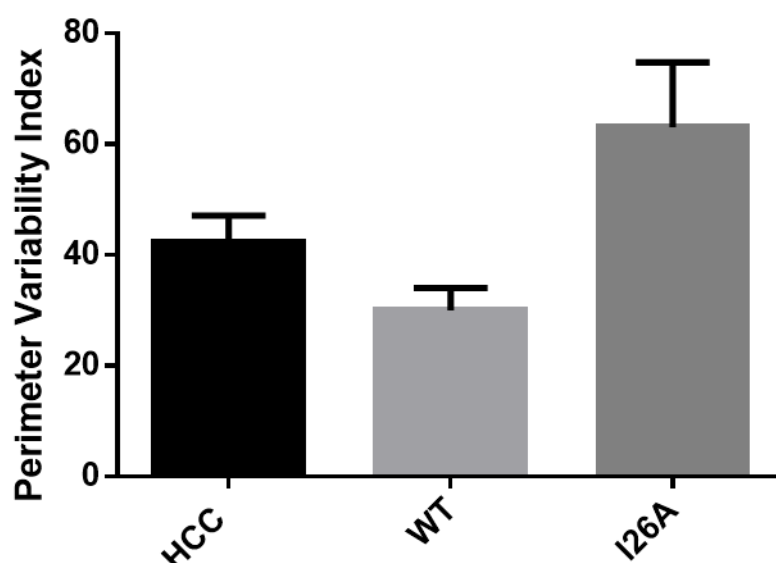


Figure 5.3 Leading edge perimeter analysis of wound healing experiments: Based on the wound healing experiments the disorder of the leading edge was calculated using a MetaMorph journal developed by Mr Nick White. The disorder or the amount the leading edge varied between frames was calculated by measuring the perimeter in every frame and quantitating the sum of the absolute difference as demonstrated by the Perimeter variability index (PVI). The higher the PVI, the more disordered the leading edge of the wound. Here we quantitate and compare the difference in PVI. As suggested by the time-lapse movies and stills shown in Figure 5.2, the leading edge is most disordered in I26A, then HCC and most ordered and uniform in WT. Student t-test gives p-values < 0.05 for all three comparisons.

5.2.3 siRNA knockdown of top candidate BRCA1 substrates

Based on high scoring in the proteomic experiments (Chapter 3) and on the high motility (speed value) in the MitoCheck analysis (Chapter 4), ezrin, moesin, fermitin-2 and delta-catenin were chosen for further investigation. These four selected candidates were individually knocked down as described in Section 2.6.4. Before each test experiment, the optimum concentration of siRNA to knockdown each candidate was first determined by serial dilution of the specific siRNA (from 20 nM to 1nM) A reduction of ezrin, moesin, fermitin-2 expression as the concentration of siRNA increases (data not shown).

Ezrin, moesin and fermitin-2 all showed significant reduction in protein level 48 hours after transfection with 20 nM siRNA, whilst the beta-actin levels remained constant (Figure 5.4.A). This indicates that there was good knockdown of protein levels for these candidates under these conditions. Experiments were repeated for delta-catenin, but due to the poor quality of the commercial antibody, delta-catenin could not be detected by western blot, and thus knockdown could not be biochemically confirmed.

5.2.4 Single cell motility on multiple candidates

We then investigated the effect of the knockdown on single cell motility. The candidate proteins were knocked by siRNA for 48 hours, which was based on the conditions optimised in Section 5.2.3, before single cell motility assay were performed. The results of the single cell motility show that the siRNA knock down of all of our candidates had an effect on motility (Figure 5.4.B). Knockdown of moesin resulted in a significant increase in motility for both WT and I26A cell lines relative to the scrambled siRNA controls. Similarly, treatment of cells with siRNA targeting delta-catenin results in an increased single cell motility. However, due to poor antibody detection, we could not confirm the knockdown of delta-catenin, thus further experiments, through qRT PCR or immunofluorescence would need to be carried out to validate knockdown. siRNA knockdown of Fermitin-2 results in a modest increase in single cell motility.

Whilst we have generated data from the MitoCheck database regarding the effect of siRNA knockdown of proteins on cell motility, one cannot assume that our experiments would lead to identical results. Firstly, we are using HCC1937 breast cancer cells as opposed to HeLa cells, thus the genetic context will be different. Moreover, whilst in principle we are doing the same assay, there are multiple subtle differences to the setup of the experiment, which will contribute to differences. Finally, the MitoCheck database

amalgamates the mean speed, based on multiple siRNAs, which will have varying knockdown efficacy. In addition the siRNAs were different to the ones used for the MitoCheck assays.

Nevertheless, for fermitin-2 the increase in motility is consistent with the analysis of the MitoCheck database. However, the p-values show that the speed increase in HCC1937 + WT BRCA1 cells is of statistical significance, whereas it is not in the HCC1937 + I26A BRCA1 cells. From the MitoCheck database, cell speed of moesin knockdown HeLa cells is close to the mean speed of the whole genome, whereas in HCC1937 cells, there was an increase of 2 fold.

Interestingly, contrary to the MitoCheck result, knockdown of ezrin results in a decreased mean velocity for both cell lines relative to the control and reaches statistical significance for the I26A cell line, whereas the WT cell line single cell motility is comparable between ezrin and scrambled treated. Altogether, these single cell motility experiments gives some results which are consistent with the MitoCheck database and others which are contrary to the MitoCheck results.

The difference in single cell motility of the scrambled siRNA treated WT and I26A cells is not significantly different, although I26A still has a greater mean speed value. This may be due to the relatively low number of cells of I26A particularly for this assay. The density of adherent cells was limited and more fields of view could not be taken as the time it takes to cycle through the positions is close to 2 minutes, the interval time between frames.

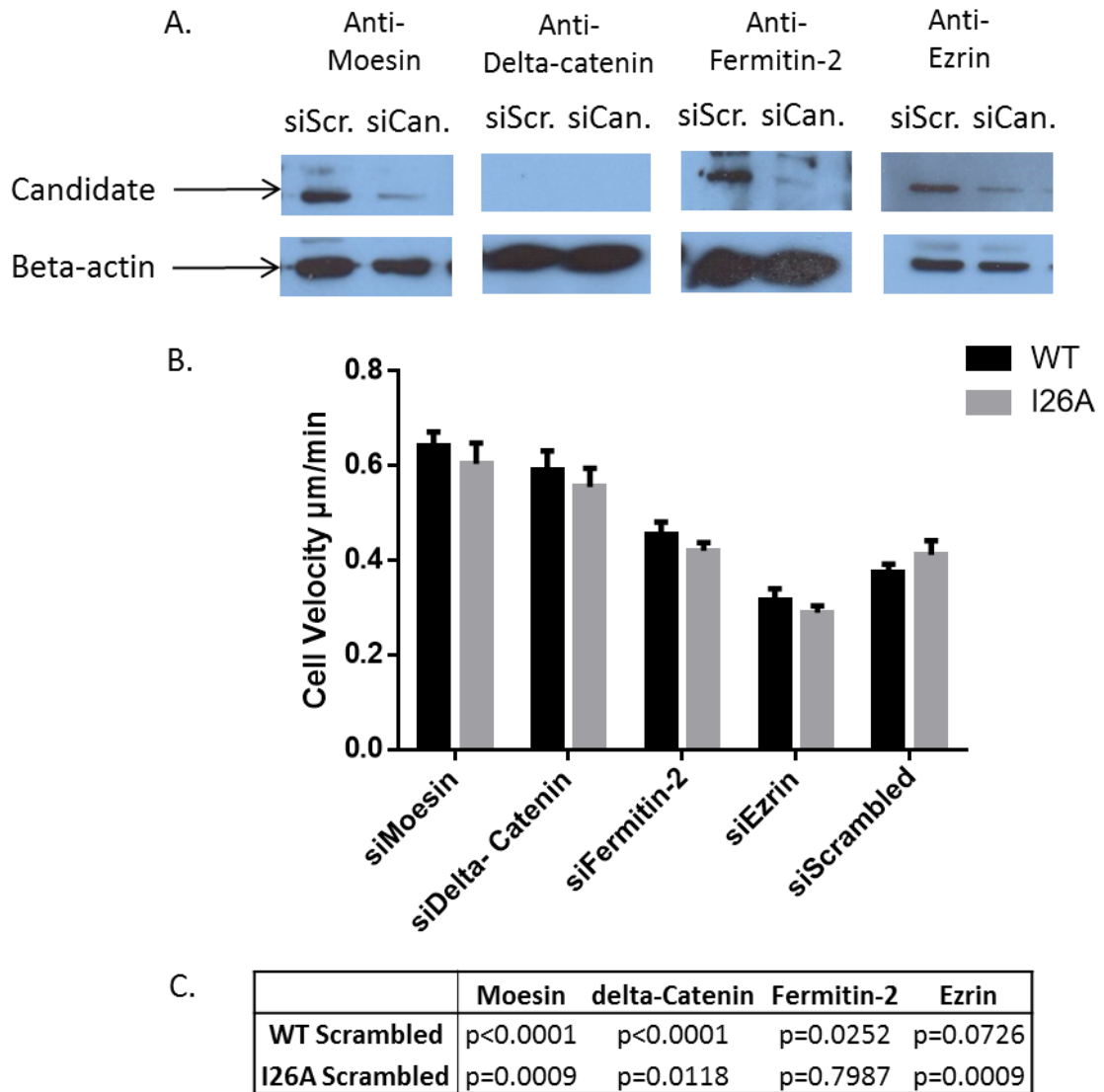


Figure 5.4 Knockdown of candidates in HCC1937 cells and single cell motility assay: (A) Western Blots of candidates after siRNA treatment in HCC1937 cells after 48 hours. Serial dilution of siRNA was used from 20nM to 1nM, here only the 20nM siRNA of the candidates and scrambled is shown, as this concentration 48 hours after treatment gave the optimum knockdown. Scrambled siRNA, 'Scr', shows endogenous protein level for each candidate, and 'Can' shows the level of the candidate protein when treated with candidate siRNA. Moesin, fermitin-2 and ezrin show good knockdown of protein, delta-catenin could not be detected. **(B)** HCC1937 cells were treated with siRNA under optimised conditions. After 24 hours, cells were trypsinised and resuspended in 24 well plates at single-cell density. Cells were cultured for 24 hours and then treated with nuclear marker DRAQ5 and confocal time-lapse movies recorded for 2 hours to assess single cell motility. siRNA knockdown of moesin, delta catenin and fermitin-2 all resulted in an increase to cell motility in WT and I26A cells. Whereas ezrin knockdown rescued the usually observed hypermotile phenotype in I26A and reduced the single cell velocity in WT. **(C)** Table of p-values calculated by Student's t-test. Each p-value is between the control motility (WT scrambled or I26A scrambled) and the motility when the candidate is knocked down in the same corresponding cell line. There was no significant difference between WT and I26A for any of the siRNAs.

5.2.5 The effect of ezrin knockdown on cell spreading, cell motility and wound healing

5.2.5.i Single cell motility and wound healing pipeline

As ezrin siRNA knockdown was the only condition, in both HCC1937 + WT BRCA1 and HCC1937 + I26A BRCA1, decreasing cell motility when compared to the control scrambled siRNA, we investigated it further. Moreover, ezrin has been substantially linked to invasion and metastasis specifically in breast (Sarrió et al. 2006; Elliott et al. 2005; Bruce et al. 2007) and ovarian carcinomas (Moilanen et al. 2003; Köbel et al. 2006). In addition, the I26A-specific reduction in cell motility suggests that there may be a difference between the WT and I26A cell lines and its relation to ezrin and thus ezrin's effect on cell motility may not be independent to BRCA1.

Three phenotypic behaviours were assessed; single cell motility, wound healing closure and cell spreading. In addition we wanted to identify the relative knockdown in the cells when assessing this behaviour. As evident from Figure 5.4.A, the knockdown of ezrin, moesin and fermitin-2 in HCC1937 + WT BRCA1 cells was successful but not 100% efficient. It is likely that cells may have varying degrees of protein synthesis knockdown and a proportion of cells is likely to have little to no targeted protein synthesis knockdown, due to failure of cellular internalisation of the siRNA product. This variation may lead to altering effects on the phenotype, which could have an effect on the overall average values, thus we took into account these variations, through validation of knockdown at the single cell level by immunofluorescence.

The experimental pipeline we developed is demonstrated as a schematic in Figure 5.4. Cells were grown and transfected as described in section 2.2.5. 24 hours after transfection, cells were resuspended to single cell density and cell spreading was

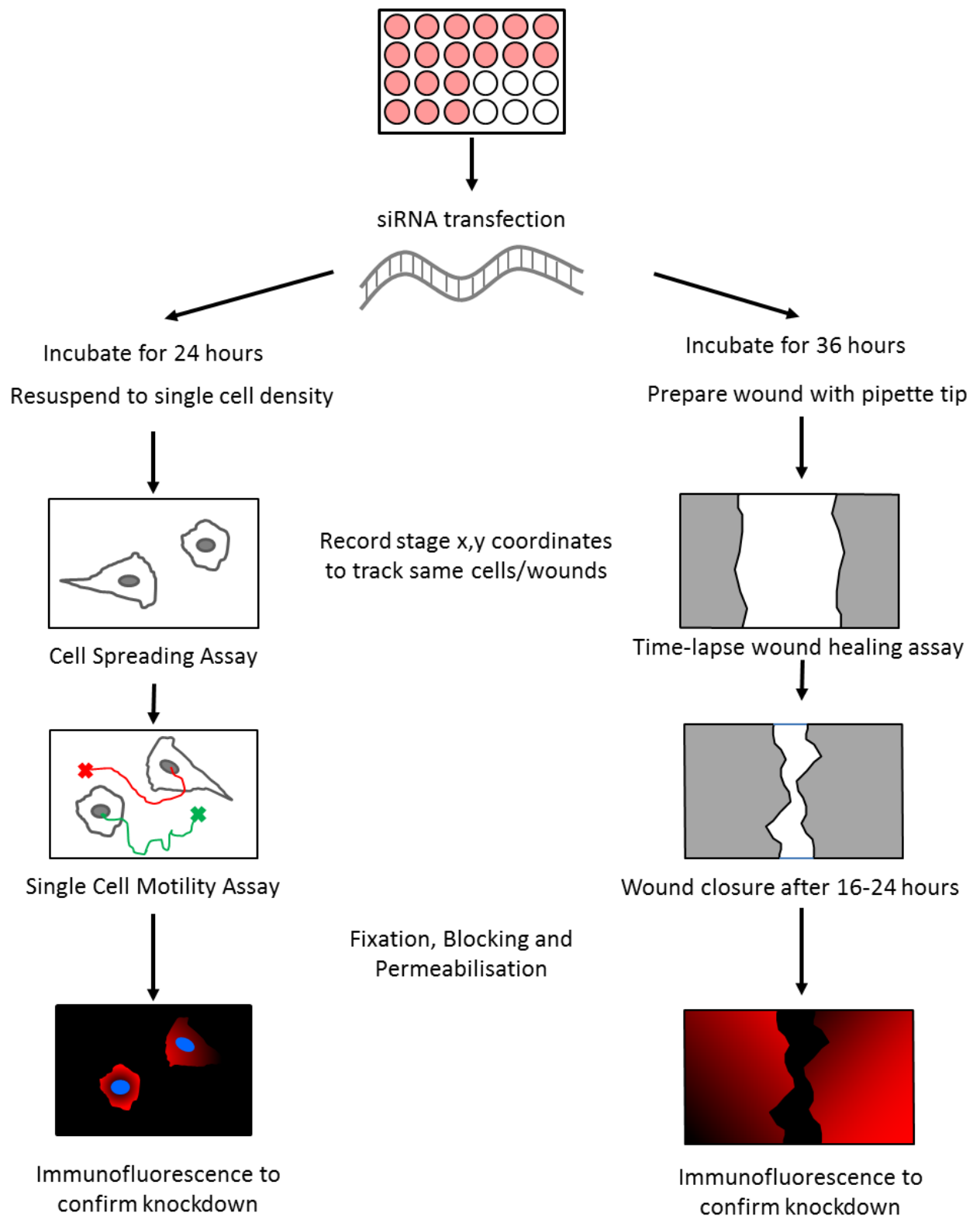


Figure 5.5 Schematic of single cell motility, spreading and wound healing pipeline:

Experimental pipeline was developed to carry out multiple assays and track the same cells and positions for every assay, at multiple positions, for multiple treatments.

Measuring cell spreading, single cell motility and immunofluorescence of the same exact same cells was one pipeline. Wound healing and immunofluorescence of our candidate(s) at the same position was another pipeline.

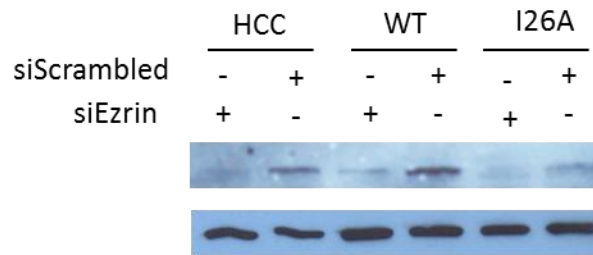


Figure 5.6 siRNA knockdown of ezrin after 24 hours: HCC1937 cell lines were knocked down with siRNA ezrin and lysed after 24 hours to establish the amount of ezrin. Efficient knockdown of ezrin is validated by western blot, probing nitrocellulose with anti-ezrin. This verifies ezrin knockdown at the start of the wound healing assay and for the spreading assay.

measured by phase contrast confocal microscopy as described in Section 2.6.3. Multiple positions were marked in the software to save the x,y coordinates to obtain at least 3 different fields of view per well. Cells were incubated for a further 21 hours, before undertaking the single cell motility assay at the same marked positions as defined for the spreading assay.

Cells were fixed and permeabilised for immunofluorescence at the same defined x,y positions. This protocol allowed us to obtain spreading, motility and a relative amount of protein based on immunofluorescence for multiples cells.

A similar pipeline was used to examine wound healing. Ezrin was knocked down by siRNA, cells were grown to a confluent monolayer during 36 hours post transfection, wounds were created and time-lapse movies were taken by confocal microscopy. Cells were fixed, permeablised and immunofluorescence was carried out on the cells. The microscope stage coordinates were kept constant for all the sites between the wound healing and immunofluorescence assay to ensure correct identification of the wounds and cells.

5.2.5.ii Effect of ezrin knockdown on Cell spreading

The first step in the Single Cell motility pipeline was to investigate the rate of spread of the different cell lines when treated with siRNA ezrin or siRNA scrambled. This was measured as described in Section 2.6.1. Cell Spread Index is the difference in total cell area (measured in pixels) between 30 minutes and 3 hours after seeding the cells after siRNA knockdown.

From Figure 5.6A, it is evident that the HCC1937 + I26A BRCA1 cell line spreads significantly less (44% reduction in spreading) than the WT cell line. This suggests that HCC1937 + WT BRCA1 cells are able to establish and maintain cell substrate adhesions more effectively than the ubiquitin ligase dead HCC1937 + I26A BRCA1 cell line. HCC1937 cells have an even lower mean spread compared to HCC1937 + WT BRCA1 and HCC1937 + I26A BRCA1 cell types. These results are consistent with Coene et al. where it was observed that CHO cells expressing WT BRCA1 had a greater spread compared to CHO cells expressing high levels of a dominant negative BRCA1 C-terminal domain protein lacking ubiquitin ligase activity.

Interestingly knockdown of ezrin results in a decrease in cell spread to a basal cell spread index level of 1500 pixels in both cell lines suggesting that ezrin is important for regulating cell spreading. Knockdown of ezrin was verified by immunofluorescence. Relative ezrin concentration was quantitated by immunofluorescence after the single cell motility assay. Cells were fixed, probed with a mouse anti-ezrin antibody and with an anti-mouse secondary antibody conjugated to the fluorophore Cy3, and the nucleic acids were stained with DAPI, as a nuclear label for cell counting and position registration. Immunofluorescence was quantitated by calculating the Corrected Total Cell Fluorescence (CTCF) in Image J (Burgess et al. 2010) (See Section 2.6.6). As Figure 5.7.B and C show,

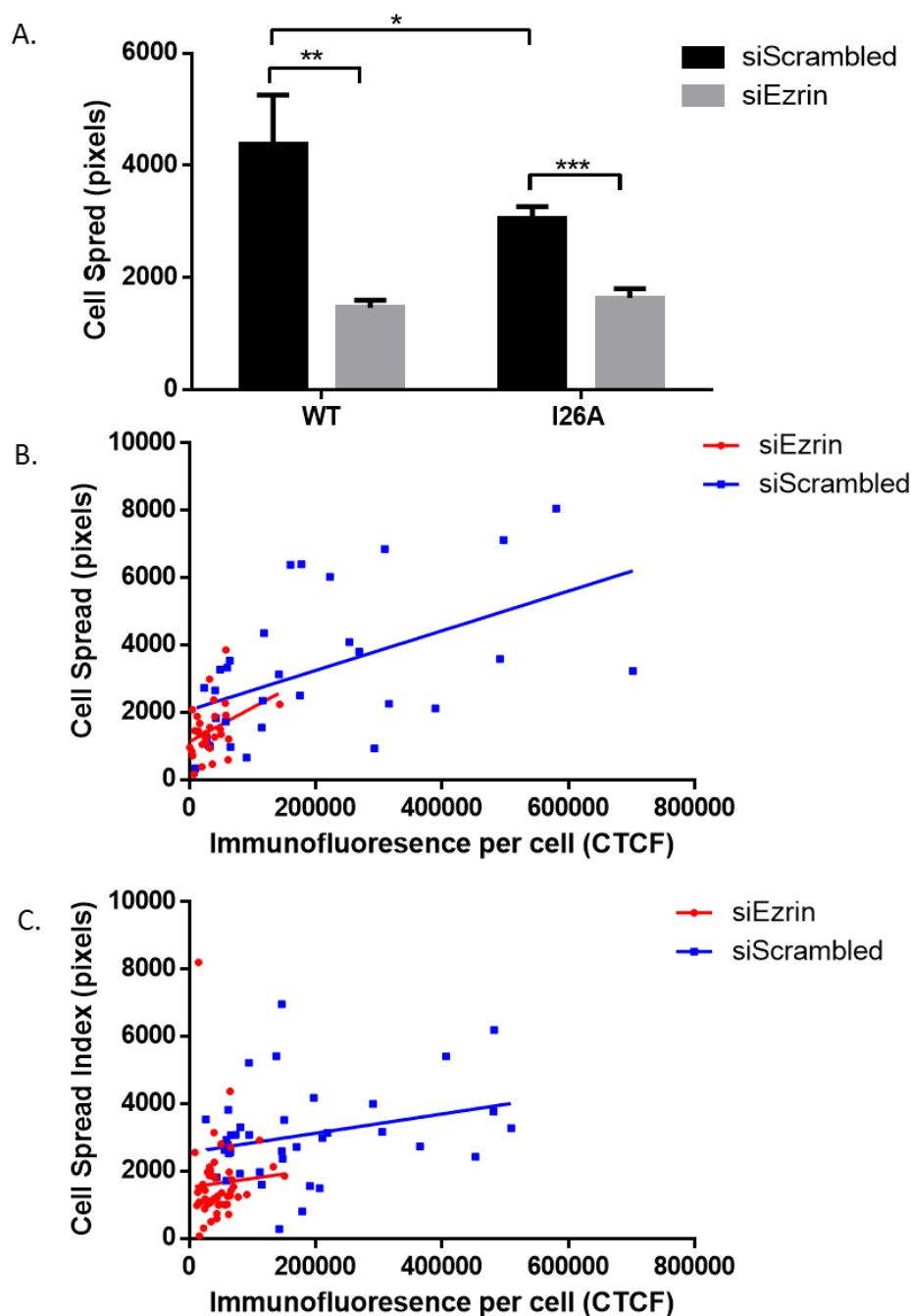


Figure 5.7 The effect of ezrin knockdown on cell spreading: (A) Spreading assay was carried out 24 hours after siRNA transfection. WT and I26A cells treated with scrambled siRNA show a significant difference ($p < 0.05$) with the mean spread of cells being 44% greater in WT cells. Ezrin siRNA treatment of both cell lines significantly reduces the mean cell spread to a similar level of approximately 1500 pixels). (B) WT and (C) I26A cell spread plot against the immunofluorescence of Cy3, which is conjugated to Anti-mouse used to detect the anti-ezrin antibody. Plot (B) has 2 points with greater cell spread not shown in the plot. For both WT and I26A, treated with siEzrin and siScrambled, there is a positive correlation between ezrin concentration and cell spread.

the mean CTCF per cell was reduced by 71-83% when both the WT and I26A cell lines were treated with siRNA ezrin.

We also plotted cell spread against ezrin amount based on IF as determined by immunofluorescence, to determine if there is a correlation between spread and IF. For the WT cell line treated with both ezrin and scrambled siRNA, there is a positive linear correlation between Corrected Total Cell Fluorescence and cell spread. Whilst the R^2 values are relatively modest, 0.26 and 0.13 for treatment with scrambled and ezrin siRNA respectively, both reach statistical significance with P-values less than 0.05, showing the slopes are significantly non-zero. Interestingly, the p-value is 10 fold greater in the siEzrin treated than in the siScrambled treated cells, which may suggest that there may be a threshold effect for ezrin's role in cell spreading. Above a threshold the amount of ezrin concentration can affect the rate of cell spreading, but below this threshold, ezrin no longer influences cell spread. Ultimately, the R^2 values are too small and we could not explicitly say there was a strong correlation between cell spread and immunofluorescence, but a greater number of cells may help establish a more clear relationship.

5.2.5.iii Effect of ezrin knockdown on single cell motility in HCC1937 cells

The single cell motility screen on four of the candidates yielded striking results. In particular, knockdown of ezrin resulted in the reduction of cell motility in both WT and I26A cell lines. Subsequently, ezrin was chosen to be investigated further. The initial screen was the result of only a single run. Thus we repeated the siRNA knockdown and single cell motility assay another 3 times, but incorporating it into the experimental pipeline described in Section 5.2.5 in order to include immunofluorescence data to quantify relative knockdown of ezrin.

The results shown in Figure 5.7A of ezrin knockdown was the same as the primary screen, which resulted in a significant decrease in single cell motility in HCC1937 + I26A BRCA1 cells with a 30.1% decrease in cell speed. HCC1937 + WT BRCA1 cells also had

a slight decrease in single cell motility, but this did not reach statistical significance. We also see the difference in cell speed between WT and I26A as observed in the first comparison between the three cell lines observed in Figure 5.1, however the differences are not as pronounced as the results from Figure 5.1.

Plotting cell speed against immunofluorescence generates some useful results (Figure 5.8 B and C). Similarly to the cell spread assay, it demonstrates that the knockdown of ezrin results in a decrease in relative ezrin level, and the motility values cluster to a slightly lower cell speed value, more visible in the HCC1937 + I26A BRCA1 graph. There is a positive correlation between the relative amount of ezrin and cell speed, however the Pearson's correlation statistic gives modest scores, with r values of 0.27 and 0.37 for WT ezrin and WT scrambled treated, and 0.195 and 0.258 for I26A ezrin and I26A scrambled, respectively. Only the WT scrambled treated cells gave a line of best fit which reached significance to determine the slope was not 0. Only the WT scrambled treated cells gave a line of best fit which reached significance for a no-zero slope. Similarly to the cell spread data in Section 5.2.5.ii, we cannot draw any strong conclusions regarding correlation of cell speed and ezrin immunofluorescence at the individual level, but again, further repetitions and data on more cells may establish a more clear relationship. These results clearly indicate the power of this correlative single cell knockdown and phenotype measurement for both the spreading and motility assays. Using this approach in experiments designed to cover a wider range of protein knockdown (by titration of siRNA concentration, for example), together with the collection of multi-parameter data on a larger number of cells would offer a means to investigate subtle differences and correlations between metrics measurable by confocal microscopy.

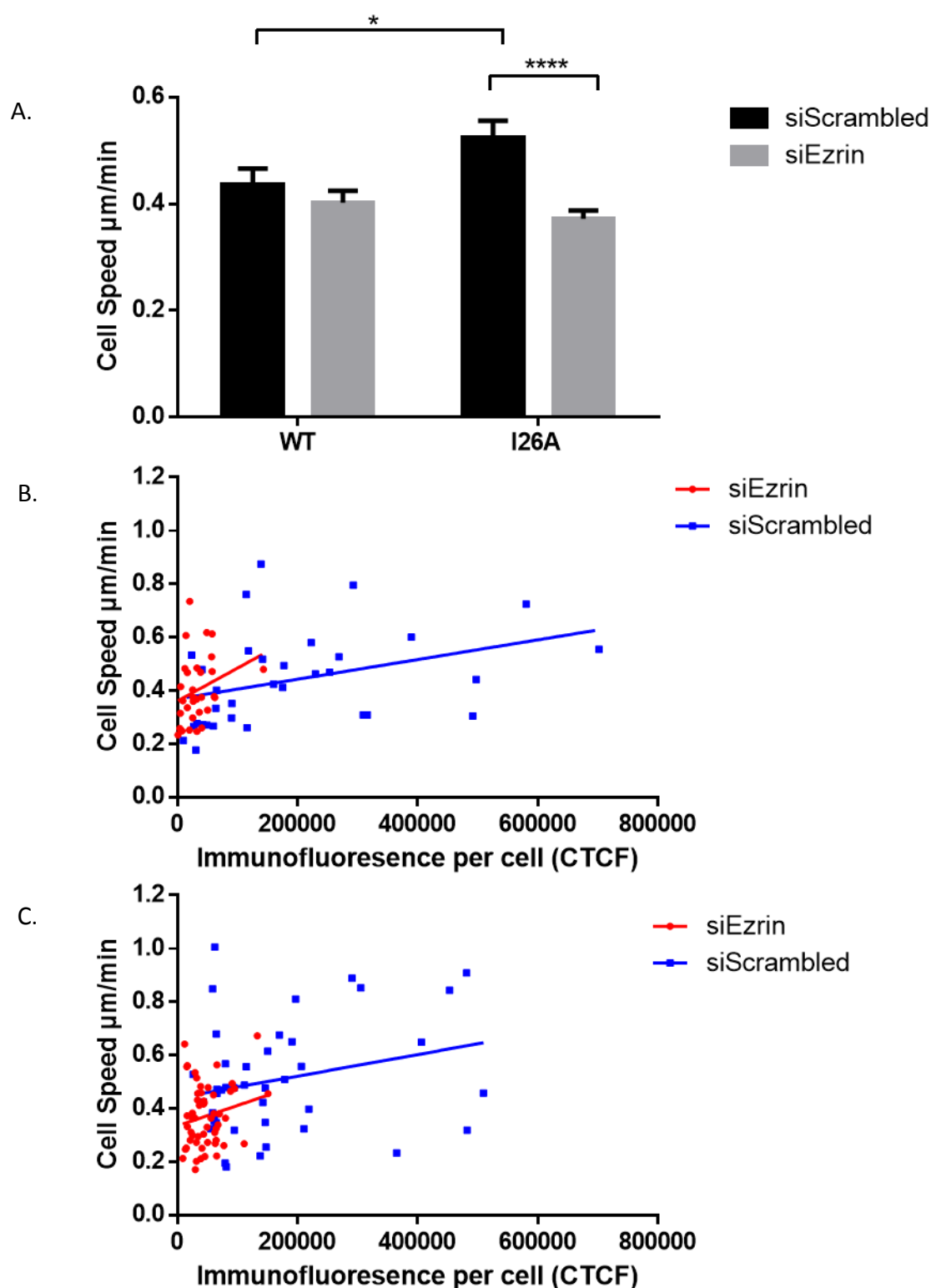


Figure 5.8 The effect of ezrin knockdown on single cell motility: (A) Single cell motility was observed 24 hours after spreading assay, 48 hours after siRNA transfection. WT and I26A cells treated with scrambled siRNA show a significant difference ($p < 0.05$) as is consistent with previous experiments. Ezrin siRNA significantly ($p < 0.0001$) reduces the mean cell motility of I26A cells to a similar level of motility of WT cells treated with siEzrin or siScrambled. (B) WT and (C) I26A cell speed plot against the immunofluorescence of Cy3, which is conjugated to Anti-mouse used to detect the anti-ezrin antibody. For both WT and I26A, treated with siEzrin and siScrambled, there is a positive correlation between ezrin concentration and single cell speed.

5.2.5.iv Effect of ezrin knockdown on wound healing in HCC1937 cell lines

Figure 5.9 shows representative effects of knocking down ezrin on HCC1937 + WT BRCA1 and + I26A BRCA1. After the final frame (which varied dependent on the nature of the wounds, typically after 60 frames, 10 hours), cells were fixed, blocked, permeabilised and probed with a mouse anti-ezrin antibody. Then the cells were incubated with a secondary anti-mouse antibody conjugated to the fluorophore Cy3 and the cells were mounted with Mowiol containing DAPI. Detecting ezrin by immunofluorescence after the wound healing assay validates effective knockdown of the protein and also demonstrates the effect of knockdown on the population, allowing identification of cells which may have normal endogenous level of protein. This approach replicates that of the single cell assays described above, but has the added advantage that the effect an altered phenotype in one cell has on the surrounding cells can also be measured.

When WT and I26A cells are treated with scrambled siRNA as a control, they maintain their phenotypes described in Section 5.2.2. HCC1937 + WT BRCA1 cells maintain a uniform leading edge with tight cell-cell adhesions within the monolayer. The wound repair occurs with the leading edge gradually progressing, with many of those cells initially at the wound edge staying at the front throughout wound closure. HCC1937 + I26A BRCA1 cells exhibit much more erratic behaviour, with a more disordered leading edge. Many cells lose their cell-cell adhesions and migrate independently into the wound. The example tracks show the directionality of HCC1937 + I26A BRCA1 cells is more variable, with cells changing directions more and often turning back on themselves. Conversely, HCC1937 + WT BRCA1 cells which start at the leading edge move in a relatively linear fashion, generally perpendicular to the wound. Moreover, HCC1937 + WT BRCA1 cells within the monolayer, away from the wound edge, exhibit less directional migration (for scrambled siRNA treated cells). The immunofluorescence also

starkly demonstrates the difference between the cell lines, with a complete breakdown of the leading edge and numerous isolated cells within the wound in the I26A cell line.

Ezrin knockdown in HCC1937 + WT BRCA1 cells show cells exhibiting a similar phenotype to HCC1937 + WT BRCA1 cells treated with scrambled siRNA, with an intact uniform leading edge. Wound closure is slightly reduced and individual cells do not migrate as far, compared to the control. Thus, in the presence of functional BRCA1, the reduction in ezrin expression does not cause an abnormal motility phenotype in this assay.

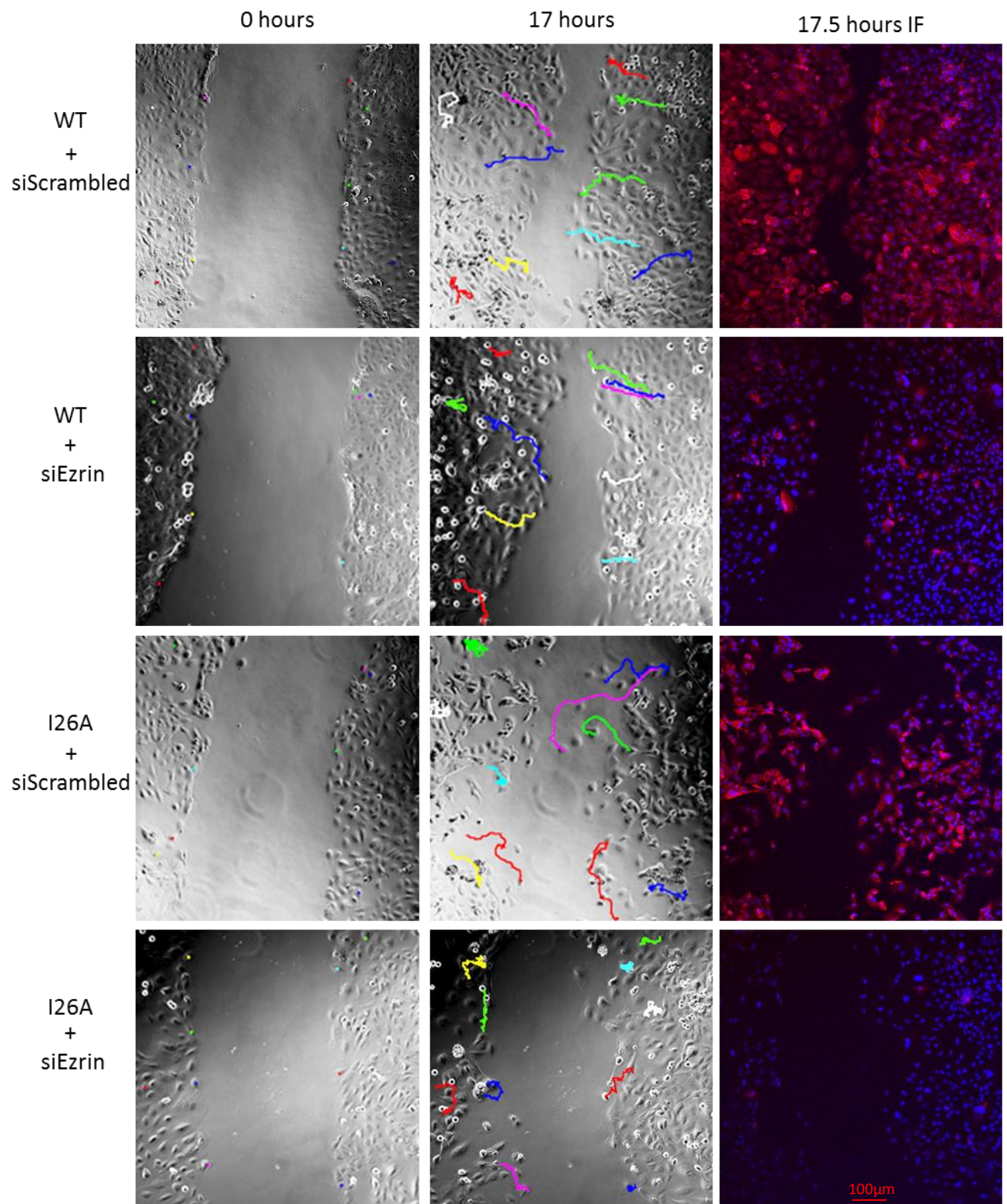
In contrast, HCC1937 + I26A BRCA1 cell monolayers transfected with ezrin siRNA exhibit a dramatic change in evolution of a wound. There is a more uniform leading edge after the 17 hours of wound repair and there are fewer isolated cells migrating independently into the wound. Cell traces are shorter and less erratic in directionality. This suggests that knocking down ezrin in the HCC1937 cell line rescues the WT phenotype, indicating that ezrin may be involved in BRCA1-mediated motility. The disruption to the leading edge after 17 hours in the experiment shown, (in the bottom left corner) was due to a single particularly large cytoplasmic cell, often observed when culturing the HCC1937 cells, which rounded up and appeared to go through apoptosis after 16 hours (not shown). This event lead to their being a less uniform leading edge at this site.

Quantitation of the disorder by normalised PVI (nPVI - normalised to the number of frames analysed because the number of frames varied between repeats) is consistent with this observation, as we see a reduction in the mean nPVI when cells treated with ezrin siRNA. However, the stark visual difference observed between I26A cell line when treated with ezrin and scrambled siRNA does not reach statistical significance based on the nPVI. This could be due to several reasons. Firstly the settings of the journal were set to be as similar as possible to the approach used by Coene et al. to facilitate comparisons.

The I26A ezrin siRNA result may be underestimating the severity of the phenotype. One of the characteristic phenotypes of the motility disorder in the wound healing experiments is the migration of free cells into the wound independently. .

However, this is not picked up by the algorithm, as it exclusively focuses on the large wound as an object which generates the perimeter values. Looking at the time-lapse movies there is an average of 14 cells that are in the wound separate from the leading edge after 10 hours in the HCC1937 + I26A BRCA1 scrambled siRNA treated, compared to an average of 4 cells in the I26A ezrin siRNA treated. The algorithm has been set for comparability between conditions, which it achieves, but cannot take account of this extreme phenotype of free migrating individual cells. It then remains possible that this extreme isolated motility phenotype is the behaviour that is due to the positive effect on motility of unmodified ezrin (which will be maximal in I26A scrambled siRNA experiments, where there is no BRCA1 ubiquitination and no reduction in ezrin expression). This is reinforced by the observation of long range cell autonomous movement in the HCC1937 + I26A BRCA1 scrambled or untreated wound healing experiments (Fig. 5.3 and Fig. 5.10).

Figure 5.9 The effect of ezrin knockdown on wound healing and immunofluorescence (following page): Cells were prepared according to the wound healing pipeline. Cells were treated with siRNA and after 24 hours wound healing assay was carried out. Results show that the wound at the start of the assay, 17 hours later and the immunofluorescence taken at the end of the assay. 10 cell tracks are overlaid to illustrate cell movement. WT and I26A treated with scrambled siRNA behave similarly to untreated WT and I26A cells as described in Section 5.2.2, WT cells having a uniform leading edge and ordered monolayer and I26A cells exhibiting more erratic cell movement, with cells moving into the wound. Treating WT cells with ezrin siRNA does not significantly effect the leading edge or closure of the wound, with cells still maintaining a uniform leading edge and exhibiting good cell-cell contact. Conversely, treating I26A cells with ezrin reverts the hypermotile phenotype, with the wound exhibiting a more ordered leading edge and fewer detached isolated cells. Cells move a shorter distance and the wound closure resembles WT wound closure. Effective knockdown was verified through immunofluorescence, with scrambled siRNA treated cells showing substantial intracellular ezrin concentration as determined by Cy3 signal. Whereas ezrin siRNA treated cells exhibited much lower levels of fluorescence, confirming ezrin knockdown.



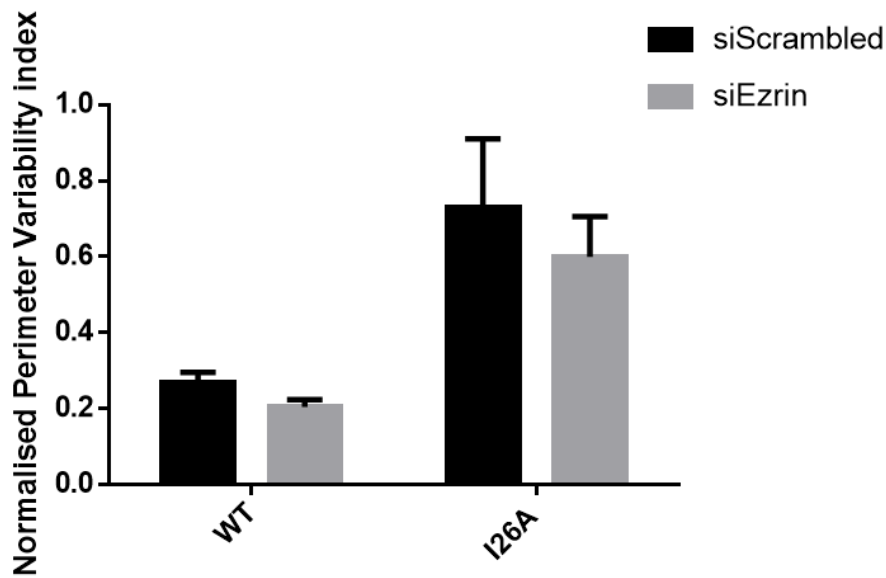


Figure 5.10 Leading edge perimeter analysis of ezrin siRNA wound healing experiments: Here we quantify the uniformity of the leading edge based on the perimeter variability index, for each of the set of wound healing movies with ezrin and scrambled siRNA treatment. We have a normalised value, as repeats had varying number of frames due to wound closure, so the PVI is divided by the number of frames. The average PVI ($n=6$) show that scrambled siRNA treatment WT and I26A show similar profiles and stark differences as seen in Figure 5.3. When cells are treated with siRNA ezrin, there is a reduction in the nPVI for both WT and I26A, suggesting the leading edge gets more ordered. This is consistent with the visualisation in Figure 5.9. However, the reduction of nPVI after ezrin siRNA treatment does not reach significance based on student's t-test.

5.3 Discussion

5.3.1 BRCA1 Ubiquitin ligase domain associated with motility and spreading.

Our results confirm that BRCA1 is involved in cell motility, cell spreading and the maintenance of cell-cell interactions. HCC1937 + I26A BRCA1 cells, which do not have a functional ubiquitin ligase domain, are not able to control these processes, with all of them being disrupted, specifically suggesting that BRCA1's RING domain is crucial to BRCA1-mediated reduction of motility. We found that the area of cell spread is greater in HCC1937 + WT BRCA1 cells compared to cells HCC1937 + I26A BRCA1. This suggests that the presence of BRCA1 ubiquitin ligase activity is important in forming and maintaining focal adhesion sites and possibly the lack of the ubiquitin ligase domain increases the rate of turnover of the focal adhesion sites. This observation is consistent with the spreading results from Coene et al. Given that single cell motility is typically upregulated in HCC1937 + I26A BRCA1 cells, may be moving more quickly, due to the more rapid turnover of cell-cell/focal adhesion sites. The increase of cell motility is consistent with the reduction of cell spreading, as cells have become polarised, do not spread in all directions and thus occupy a smaller area (Ware et al. 1998).

When considering our candidates via the initial motility screen, we observed surprisingly, (Figure 5.4) results which were not entirely consistent with the output from the MitoCheck database. Moesin had a modest speed close to the mean speed through the MitoCheck results, whereas it was the most upregulated speed when treated with siRNA in our experimental approach. Delta-catenin also had an increased motility in our screen as opposed to the MitoCheck results. Fermitin-2 could not be assessed due to the inability to identify effective knockdown. Lastly, ezrin had a modest increased speed from the MitoCheck, but a slight reduction in cell speed in our screen. These results could be for several reasons. Firstly, we are using a different cell type HCC1937 compared to HeLa

cells, which can have a gross impact on a number of phenotypes. Moreover, our approach has been designed to assess motility, as opposed to the MitoCheck results which are designed to assess mitosis. As a result the sensitivities of the two sets of experiments may be different. In addition, we are assessing single cell motility using DRAQ-5, which can be toxic to the cell after long periods of time, again impacting on the phenotype.

5.3.1.i Experimental design

We developed and tested an experimental pipeline to determine multiple data points for each cell during a time-course experiment (cell spreading over time, cell movement over time and target protein values for each cell). Whilst our graphs show positive correlations between cell spreading and ezrin immunofluorescence, the statistics are just below the threshold to be significant. However, this pipeline can be utilised to investigate other candidates and trying to undertake the experiment with more data points would be beneficial in order to achieve statistical significance. In this instance it was particularly difficult as we were trying to measure single cell spreading and motility, which requires a low seeding density. However, with more repeats one would be able to investigate more closely correlations between protein concentration and cell locomotion. One could even plot 3D graphs with each point representing multiple points of data. Whilst we tried this, the numbers were not great enough to deduce any sort of cloud, correlation or pattern within our dataset, at least by visual inspection. In addition, our ezrin knockdown was particularly successful, managing to knockdown approximately 80% of the protein, so a sensible next step would be to carry out the same experiment with a serial dilution to get a range of ezrin knockdown values and hopefully get a better spread of the data.

5.3.1.ii Potential Model of the association of ezrin with BRCA1

From the previous chapters we identified several proteins that appeared in our proteomic screen for differential ubiquitination of candidates in the presence and absence of functional BRCA1. We then identified whether any of these proteins have possible links to motility behaviour by assessing their relative speed of cells in which these targets are selectively knocked down based on the MitoCheck database. This allowed us to draw up a list of interesting biologically plausible proteins for further investigation, specifically ezrin, Moesin, fermitin-2 and delta-catenin.

We decided to investigate ezrin further due to it having substantial links to invasion and metastasis specifically in breast (Sarrió et al. 2006; Elliott et al. 2005; Bruce et al. 2007) and ovarian carcinomas (Moilanen et al. 2003; Köbel et al. 2006). In addition, the I26A-specific reduction in cell motility suggests that there may be a difference between the WT and I26A cell lines and its relation to ezrin and thus ezrin's effect on cell motility may not be independent to BRCA1.

Ezrin has two conformational states, an open, active state, which crosslinks the plasma membrane indirectly through the integrin receptor to the actin cytoskeleton and the closed state which involves an intra-molecular bond between the N-terminal FERM domain and the C-terminus. This state 'closes' the protein and prevents interaction with F-actin and the cell membrane. These two states are in a dynamic equilibrium.

Multiple post-translational modifications of ezrin have been identified and shown to be of functional relevance. Phosphorylation of ezrin at Tyr567 (Fievet et al. 2004) and Tyr 477 (Mak et al. 2012) is an important factor in determining the conformation of the protein, pushing it into the active state. This subsequently increases motility and invasion and has an effect on spreading and adherent junction formation (Khanna et al. 2004). Overexpression of ezrin is also associated with internalisation of E-cadherin which is a

well-established step in the EMT metastatic cascade (Li et al. 2008). In addition, ezrin has been shown to be ubiquitinated by WWP1. This has been linked to stabilisation of downstream factors involved in motility (Zaarour et al. 2012). Thus it follows that ezrin can be regulated by post-translational modification, including ubiquitination, and BRCA1 mediated ubiquitination of ezrin may be another regulatory process involved. Specifically in this instance, the ezrin site of ubiquitination is found in the linker region between the FERM domain and C-terminal-actin binding domain. This could be of important structural significance as the presence of an ubiquitin moiety within the linker region could impact the steric forces and the equilibrium between the open or closed state.

From our proteomic screen, ezrin was detected as being ubiquitinated in the presence of BRCA1. From the ubiquitin-remnant profiling approach, an ezrin peptide was detected, with the modified lysine observed to be in the linker region between the two FERM and C-terminal domain. In addition, Coene et al. identified a direct protein-protein interaction between BRCA1 and ezrin through pull-down experiments, protein domain overlay blotting, mass spectrometry and validation by confocal microscopy. Results from this chapter suggest that ezrin is important in regulating spreading, as knockdown of ezrin results in a decrease in the rate of cell spreading relative to cells with functional BRCA1 and ubiquitin ligase dead BRCA1. Other studies have also suggested that ezrin is required for efficient cell spreading (Ross et al. 2011)

Furthermore, knockdown of ezrin reduced single cell motility to a basal level comparable to single cell motility of cells expressing functional BRCA1. Wound healing results demonstrate that I26A wounds typically exhibit hypermotile, erratic behaviour, with a loss of cell-cell contact and increased motility. This phenotype is reversed when ezrin is knocked down.

All this is consistent with the notion that BRCA1 has a role in mediating cell motility and one of the ways it exerts this function may be through the regulation of ezrin. Whilst, phosphorylation of ezrin results in the open active state, increases cell motility and invasive characteristics, BRCA1 may have an opposing effect. BRCA1 mediated ubiquitination of ezrin may promote closure of ezrin to its inactive state due to a conformational change which may prevent interaction with PIP2, or kinases such as PKC and Src.

The model we propose is illustrated in Figure 5.11. BRCA1 may promote the closed state of ezrin through ubiquitination. In this model, knockdown of BRCA1, or an ubiquitin ligase dead BRCA1 would result in an increase in cell motility, due to the inability of BRCA1 to ubiquitinate ezrin and encourage inactivation. More ezrin proteins may have an open active confirmation, promoting cell motility. Phenotypically, this is what we observe in HCC1937 + I26A BRCA1 cells. Knocking down ezrin, may not have as much of a severe effect on motility, as the crosslinker is no longer present to bind the plasma membrane to the actin cytoskeleton, rendering ezrin-mediated motility ineffective. This is what we observe with the basal level of motility with HCC1937 + WT BRCA1 cells and knockdown of ezrin in both WT and I26A.

An interesting set of follow-up experiments to test this model would be to undertake overexpression of ezrin and BRCA1 and see their relative effects on motility. Increasing BRCA1, according to the model would promote ezrin ubiquitination and shift the balance towards increasing the pool of inactive ezrin. Overexpression of ezrin on the other hand would disrupt the balance and increase the amount of active ezrin, due to saturation of BRCA1. This in turn would increase cell motility, loss of cell-cell contact and invasive behaviour and this is what is observed clinically, that ezrin overexpression is important for metastatic behaviour.

There are likely to be many more effects that influence ezrin-mediated tumourigenesis, as ezrin is now being observed, similar to BRCA1, as a protein which acts as a platform upon which other complexes can form and interact. Thus ezrin has been implicated in other cellular functions such as cell signalling and cell metabolism (Ren et al. 2012). Depending on the stage of the metastatic cascade, there may be differential advantages to a tumour cell to have a greater concentration of active or inactive ezrin, due to the number of functions. This is supported by the fact that there are some contradictory results in the literature. On the one hand, active ezrin localises to the cell membrane and it is associated with motility and invasion whereas inactive ezrin localises to the cytoplasm, suggesting that ezrin at the cell membrane is responsible for a hypermotile phenotype. Conversely, the histopathology of patient samples from metastatic cells exhibit a more diffuse, cytoplasmic staining of ezrin associated with the inactive form (Sarrió et al. 2006). Thus what is important is the relative proportion of the active/inactive form of ezrin at different stages of metastasis. We suspect that at the initial stages of the metastatic cascade, it is advantageous for a cancer cell to have a greater proportion of active ezrin, in order to lose cell-cell contact, increase cell motility. BRCA1 may be important in determining and controlling the flux of the ezrin states, by ubiquitinating ezrin and shifting the equilibrium towards inactive ezrin.

Here we have described a possible model for BRCA1 engaging with ezrin-mediated motility pathway. However, it should be stated that BRCA1 may affect cell migration through multiple proteins. As observed in the nucleus, BRCA1 acts as a scaffold upon which complexes form and undertake different functions, thus it is not unreasonable to expect that depending on the partner proteins and localisation, BRCA1

may ubiquitinate a number of proteins which are involved in cell motility and the next informative one to consider would be moesin, which may have an antagonistic function relative to ezrin based on the single cell motility results in Section 5.2.4.

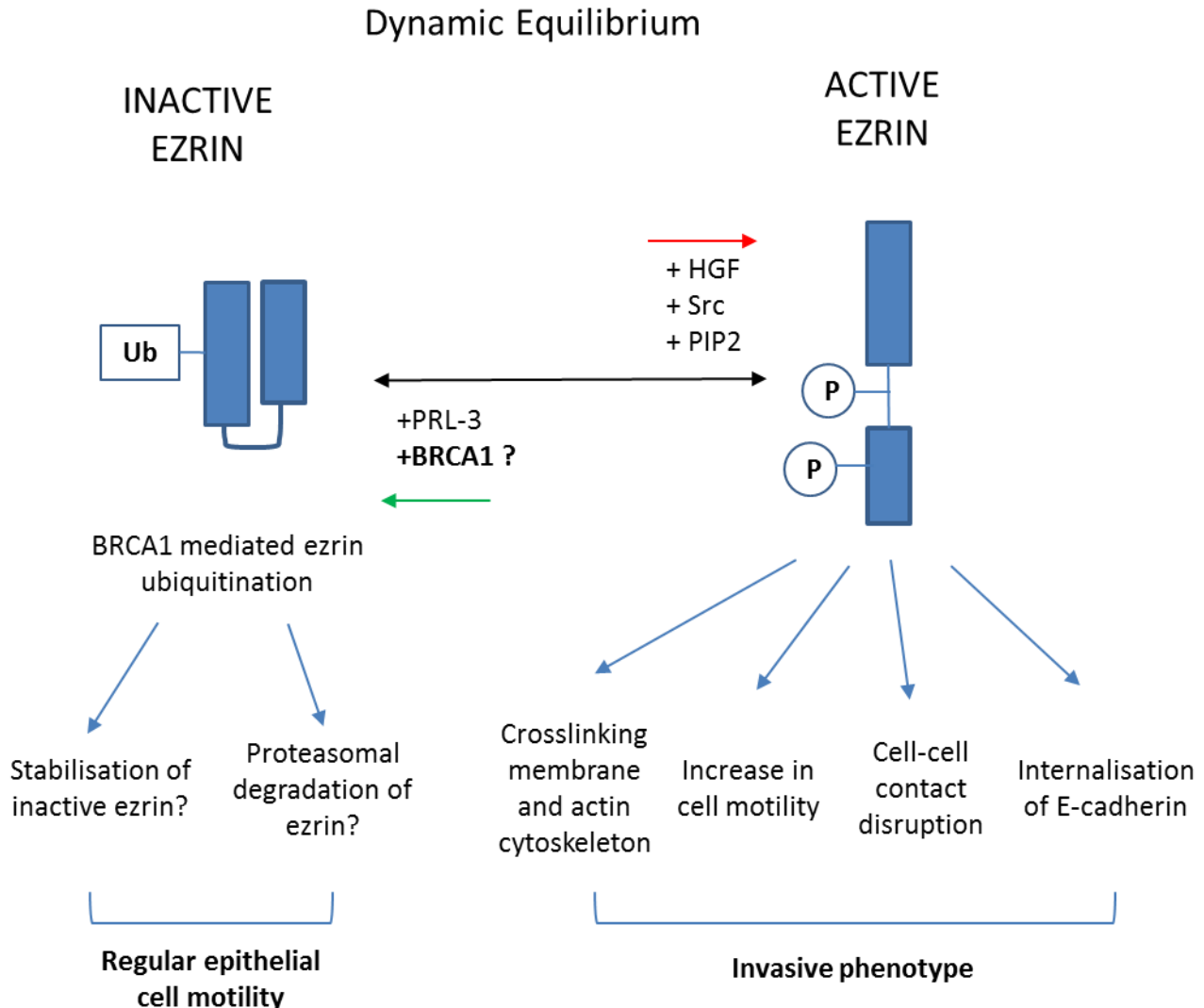


Figure 5.11 Possible model for BRCA1 regulation of motility through ezrin: Given our experimental results and the literature, we propose the following model. Ezrin has 2 states, active and inactive, which is in a dynamic equilibrium. Phosphorylation and binding to PIP2 activates ezrin. This is also stimulated by growth factors such as HGF. The active form crosslinks the cell membrane to F-actin and an increased flux of active ezrin is associated with increased motility and loss of cell-cell adhesions. We propose that BRCA1 ubiquitinates ezrin, which has the effect of inactivating ezrin, possibly through stabilisation of ezrin in its inactive form or possibly targeting ezrin for proteasomal degradation, thereby regulating cell motility.

5.4 Conclusions and Future Directions

Here we have confirmed and shown that the BRCA1 ubiquitin ligase activity is important in regulating motility and spreading in a 2D *in vitro* context, which potentially has important consequences for understanding tumourigenesis of BRCA1 mutated tumours. In addition, ezrin has shown to be an interesting candidate and results are consistent with it being a BRCA1 ubiquitin ligase target important in BRCA1-mediated motility. Further biochemical experiments need to be undertaken to validate this, such as *in vitro* ubiquitination assays using purified proteins. Confirming BRCA1 ubiquitinates ezrin, should be followed by site-directed mutagenesis experiments, with ezrin that cannot be ubiquitinated due to K to R mutations. Transfecting K->R mutants in HCC1937 cell lines +/- functional BRCA1 ubiquitin ligase activity will be important in further understanding the role of this pathway.

Confirming results within other cell lines, such as the ovarian cancer BRCA1-null cell line UWB.1.289, and the BRCA1-expressing breast cancer cell line MCF-7 would strengthen the case. My preliminary results of the ezrin knockdown experiments on MCF-7 cells expressing WT and I26A BRCA1 generates similar motility profiles, with I26A cells exhibiting an increased motility which is reversed after knockdown of ezrin.

After confirming these hits *in vitro* in other cell lines and 3D models, the next step would be to investigate this pathway in mice with the aim to develop therapeutics which may address the nuanced balance of BRCA1-directed ubiquitinated species and their relative levels of ubiquitination.

Chapter 6: Discussion

Breast and ovarian cancers with BRCA1 mutations are typically characterised by being highly aggressive, with a propensity to metastasize, and are associated with poor prognoses (Moller et al. 2007; Lakhani et al. 2000). As a result, considerable efforts have gone into understanding the molecular mechanisms underlying the phenotype of these tumours. Genetic instability is one of the hallmarks of cancer and the loss of BRCA1 disrupts multiple DNA damage response pathways, increasing the risk of mutation and subsequent selection that generates the somatic variability that contributes to further tumour growth and heterogeneity in response to therapy. However, novel pathways associated with BRCA1 may shed light on the aggressive metastatic nature of BRCA1-mutated tumours.

In this thesis we have attempted to uncover some of BRCA1's functions within the cell and in particular to further understand BRCA1's specific role in regulating motility. Several studies have demonstrated BRCA1 has a role in regulating cell motility (Yasmeen et al. 2008; Promkan et al. 2009) and moreover that the E3 ubiquitin ligase activity is important in controlling this behaviour (E D Coene et al. 2011). Our starting point was to pose the question, what are the extra-nuclear BRCA1 E3 ubiquitin ligase substrates, and are any of these associated with regulating cell motility?

To address this question we have undertaken two large-scale, system-wide (and thus unbiased by prior expectation) studies that between them generated an enormous amount of data. Our proteomic comparison of enriched ubiquitinated species between cell lines expressing functional BRCA1 and ubiquitin ligase dead BRCA1 identified over 50 proteins which were deemed to be attractive candidates based on their reproducible enrichment, and their known biology. In addition we identified several sites of

ubiquitination using the ubiquitin remnant profiling approach. Within the candidate list, a large subset of proteins was cytoskeletal related and another group was linked to translation, as indicated by pathway analysis. The former was of particular interest when considering BRCA1 in the context of motility. Whilst we did not experimentally confirm the direct interaction between BRCA1 and the 55 candidates, these results can hopefully be utilised by the BRCA1 research community to identify or re-establish candidates. These results further contribute to the aim of establishing a BRCA1 ‘ubiquitome’, a quest that is becoming more realistic with the use of state of the art mass spectrometers and ubiquitin remnant profiling (Kim et al. 2011; Emanuele et al. 2011).

We wanted to address which if any of these candidate ubiquitination targets had a role to play in cell motility in order to focus in on a possible pathway relevant to BRCA1-mediated motility. Instead of running *in vitro* motility experiments for each of the candidates, we adopted an *in silico* approach using re-analysis of data in a pre-existing online database. The MitoCheck database is a fantastic resource consisting of 192,000 time-lapse movies of HeLa cells with every gene systematically siRNA knocked down (Neumann et al. 2010). We analysed these movies to determine the speed of cell movement after siRNA knockdown and subsequently produced a genome-wide motility database, which contains speed and distance information for HeLa cells after knockdown of every single gene. In addition we curated this database to contain gene ontology and KEGG pathway information for each of the genes. From the results, we looked for overlap between the two datasets in order to establish which of our ubiquitination candidates also exhibited an abnormal motility phenotype after siRNA knockdown. This approach enable us to prioritise a subset of proteins to explore further, namely ezrin, moesin, fermitin-2 and delta-catenin. These proteins are all associated with cytoskeletal regulation and have been implicated in tumourigenesis (Bruce et al. 2007; Charafe-Jauffret et al. 2007; Schackmann

et al. 2013; Shen et al. 2012). In addition, analysis of the motility database through Gene Set Enrichment Analysis allowed us to uncover families that up-regulate motility when knocked down. Unexpectedly, the top scoring gene set was the carbonic anhydrase family. Whilst disruption of individual carbonic anhydrases have been linked to motility (Fiaschi et al. 2013), it has yet to be reported that the knockdown of multiple members of the family increase cell motility. We suspect that an underlying reasoning behind this may be linked to the carbonic anhydrases being metalloproteins that coordinate zinc ions. Knockdown of the proteins would lead to an increase in intracellular zinc, which we show increases motility and a recent study reinforces the notion that cellular zinc concentration is linked to cell motility (Hogstrand et al. 2013).

This database will be made available online, where it will hopefully serve as a resource and be used in conjunction with the results from the MitoCheck database to molecular and developmental biologists investigating cell motility and cancer biology.

In order to test the hypothesis generated from the overlap of these large scale surveys, we developed a combination of spontaneous motility, spreading and immunofluorescence assays to assess the effect of siRNA knockdown of our selected candidates on these behaviours in cells expressing functional BRCA1 and ubiquitin ligase dead BRCA1. After an initial single cell motility screen, we further explored ezrin, a cytoskeletal-related protein, given its link to metastasis and its motility phenotype (Li et al. 2008; Elliott et al. 2005). Our results showed that cells expressing functional BRCA1 maintained strong cell-cell adhesions and exhibited comparable motility behaviour to MCF-7 cells, which endogenously express BRCA1. This was in contrast to cells expressing ubiquitin ligase dead BRCA1, which exhibited a hypermotile phenotype, reduced cell-cell contact, slower spreading and more erratic motility behaviour. We found that knocking down ezrin in the ubiquitin ligase dead cell line resulted in a reduction of

cell motility and a more similar motility phenotype to cells expressing functional BRCA1, consistent with a model in which ezrin and BRCA1 may engage in the same pathway to regulate cell motility.

These hypothesis testing results represent a work in progress and are not conclusive. Further expression and mutagenesis experiments will be required to confirm the hypothesis that BRCA1 ubiquitinates ezrin with the downstream effect of regulating motility. Moreover, some of the other candidates are also likely to be putative substrates, such as moesin due to the results from the proteomic experiments, showing strong enrichment with the WT BRCA1 expressing cell line.

There may be multiple candidate proteins that are involved, but affect motility under different circumstances *in vivo* such as hypoxia, different ECM density/composition, presence of different cell types and local inflammation. Moreover, if BRCA1 function is important in many different types of solid tumour, individual candidates may be more or less important in tumours with differing characteristics based on the nature of the ECM anchoring as an example.

Nevertheless, given the previous reports linking ezrin and BRCA1 (Coene et al. 2011) and its established role in metastasis, this pathway is very plausible and provides a conceivable mechanism to understand why BRCA1 mutated tumours are prone to invasive, metastatic behaviour. We have shown that BRCA1 has at least one new extranuclear function that is potentially important for the natural history of tumours in man.

6.1 Future work and clinical implications

From a biological perspective, this thesis is the starting point for further characterization of specific BRCA1 substrates. We predict ezrin is a substrate with a

lysine within the linker region, given the identification of a peptide detected in WT expressing cells. This next needs to be verified by production and expression of lysine to arginine point mutants of ezrin and observing their influence on motility behaviours in the presence of functional and mutant BRCA1. Specifically enriching ezrin followed by ubiquitin remnant profiling may confirm the identification of the specific lysine. In addition a number of our other candidates look promising, with moesin another candidate to explore further. It would also be wise to undertake the siRNA knockdown motility experiments on the BRCA1 null ovarian cancer cell line UWB1.289, transfected with WT and I26A BRCA1, to support the notion that the effects are BRCA1-dependent. After characterisation of the substrates *in vitro*, one could investigate the motility behaviours in murine models of tumour spread, to determine whether the effects are consistent and reproducible *in vivo*.

The findings from this thesis will be of use from a clinical perspective. Gaining a better understanding of BRCA1-mutated tumours will aid the development of tailored treatment plans for these cancers. Currently advice for treatment of BRCA1-mediated tumours is to adopt the same strategy as sporadic, non-hereditary tumours (Robson 2007; Bengala 2009). However, patients suffering from BRCA1-mutated tumours have poor prognoses and reduced 5-year survival rate even with early detection (Moller et al. 2007). Moreover, these tumours are less sensitive to the microtubule-targeting taxane-based treatments (Quinn et al. 2003). The use of estrogen-receptor targeting drugs such as tamoxifen are also less effective due to the majority of BRCA1-mutated tumours exhibiting a triple-negative phenotype (King et al. 2001). The more recent trials of PARP inhibitors looks promising as a treatment against BRCA1-mutated tumours, but genetic reversion can lead to resistance due to secondary mutations (Dhillon et al. 2011). Thus there is still an urgent requirement to improve and develop better therapeutics.

Targeting the metastatic cascade is of absolute importance when trying to develop novel therapeutics. The first step of this cascade is the acquisition of motile and invasive behaviour, which we and others have shown to be linked to BRCA1.

This work and further characterisation of the ubiquitin ligase substrates of BRCA1 will further improve our understanding of metastasis in BRCA1-mutated tumours. It is becoming apparent that post-translational modifications are a dynamic processes and perturbing the balance between their addition and removal can disturb cellular behaviour. This is of course also true for other PTMs such as phosphorylation, which is a balance of kinase and phosphatase activity. This gives a therapeutic opportunity apart from genetic engineering of all tumour cells to express the kinase – namely to inhibit the phosphatase.

Similarly, expressing or delivering functional BRCA1 specifically to tumour cells to restore the balance by increasing selective ubiquitination is a difficult task. However, if one of the factors contributing to the metastatic phenotype is the decreased ubiquitination of a substrate in the absence of functional BRCA1, leading to an altered stoichiometry of ubiquitinated versus non-ubiquitinated substrate, therapeutics could be developed to correct the equilibrium. An example may be administering a drug that specifically inhibits the action of the de-ubiquitinase responsible for the removal of ubiquitin from the protein of interest, e.g. ezrin. This would have the effect of reverting the balance of ubiquitinated ezrin versus non-ubiquitinated ezrin back to a relatively normal state. This work will hopefully contribute to the understanding of cancer biology and in the long-term to the development of novel therapeutics and treatment of these aggressive tumours.

Bibliography

- Alibés, A. et al., 2007. IDconverter and IDCligh: conversion and annotation of gene and protein IDs. *BMC bioinformatics*, 8(1), p.9.
- Ambühl, M.E. et al., 2012. High-resolution cell outline segmentation and tracking from phase-contrast microscopy images. *Journal of microscopy*, 245(2), pp.161–70.
- Antoniou, A. et al., 2003. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. *American journal of human genetics*, 72(5), pp.1117–30.
- Baer, R. & Ludwig, T., 2002. The BRCA1/BARD1 heterodimer, a tumor suppressor complex with ubiquitin E3 ligase activity. *Current opinion in genetics & development*, 12(1), pp.86–91.
- Beger, C. et al., 2001. Identification of Id4 as a regulator of BRCA1 expression by using a ribozyme-library-based inverse genomics approach. *Proceedings of the National Academy of Sciences of the United States of America*, 98(1), pp.130–5.
- Bell, A.W. et al., 2009. A HUPO test sample study reveals common problems in mass spectrometry-based proteomics. *Nature methods*, 6(6), pp.423–30.
- Bengala, C., 2009. Recent advances in managing triple-negative breast cancers. *F1000 medicine reports*, 1.
- Bianchi-Smiraglia, A., Paesante, S. & Bakin, A. V, 2013. Integrin $\beta 5$ contributes to the tumorigenic potential of breast cancer cells through the Src-FAK and MEK-ERK signaling pathways. *Oncogene*, 32(25), pp.3049–58.
- Brocardo, M. & Henderson, B.R., 2008. APC shuttling to the membrane, nucleus and beyond. *Trends in Cell Biology*, 18(12), pp.587–596.
- Brodie, K.M. & Henderson, B.R., 2012. Characterization of BRCA1 protein targeting, dynamics, and function at the centrosome: A role for the nuclear export signal, CRM1, and aurora a kinase. *Journal of Biological Chemistry*, 287(10), pp.7701–7716.
- Brooks, C.L. & Gu, W., 2011. p53 regulation by ubiquitin. *FEBS letters*, 585(18), pp.2803–9.
- Bruce, B. et al., 2007. Expression of the cytoskeleton linker protein ezrin in human cancers. *Clinical & experimental metastasis*, 24(2), pp.69–78.
- Brzovic, P.S. et al., 2003a. Binding and recognition in the assembly of an active BRCA1/BARD1 ubiquitin-ligase complex. *Proceedings of the National Academy of Sciences of the United States of America*, 100(10), pp.5646–51.

- Burgess, A. et al., 2010. Loss of human Greatwall results in G2 arrest and multiple mitotic defects due to deregulation of the cyclin B-Cdc2/PP2A balance. *Proceedings of the National Academy of Sciences of the United States of America*, 107(28), pp.12564–9.
- Cadwell, K. & Coscoy, L., 2005. Ubiquitination on nonlysine residues by a viral E3 ubiquitin ligase. *Science (New York, N.Y.)*, 309(5731), pp.127–30.
- Calvo, V. & Beato, M., 2011. BRCA1 counteracts progesterone action by ubiquitination leading to progesterone receptor degradation and epigenetic silencing of target promoters. *Cancer Research*, 71(9), pp.3422–3431.
- Charafe-Jauffret, E. et al., 2007. Moesin expression is a marker of basal breast carcinomas. *International journal of cancer. Journal international du cancer*, 121(8), pp.1779–85.
- Chen, C.F. et al., 1996. The nuclear localization sequences of the BRCA1 protein interact with the importin- α subunit of the nuclear transport signal receptor. *Journal of Biological Chemistry*, 271(51), pp.32863–32868.
- Chen, Y. et al., 1996. BRCA1 is a 220-kDa nuclear phosphoprotein that is expressed and phosphorylated in a cell cycle-dependent manner. *Cancer research*, 56(14), pp.3168–72.
- Chen, Z.J. & Sun, L.J., 2009. Nonproteolytic Functions of Ubiquitin in Cell Signaling. *Molecular Cell*, 33(3), pp.275–286.
- Chicooree, N. et al., 2013. Enhanced detection of ubiquitin isopeptides using reductive methylation. *Journal of the American Society for Mass Spectrometry*, 24(3), pp.421–30.
- Christensen, D.E., Brzovic, P.S. & Klevit, R.E., 2007. E2-BRCA1 RING interactions dictate synthesis of mono- or specific polyubiquitin chain linkages. *Nature structural & molecular biology*, 14(10), pp.941–8.
- Cianchi, F. et al., 2010. Selective inhibition of carbonic anhydrase IX decreases cell proliferation and induces ceramide-mediated apoptosis in human cancer cells. *Journal of Pharmacology and Experimental Therapeutics*, 334(3), pp.710–719.
- Coene, E.D. et al., 2011. A novel role for BRCA1 in regulating breast cancer cell spreading and motility. *Journal of Cell Biology*, 192(3), pp.497–512.
- Coene, E.D. et al., 2005. Phosphorylated BRCA1 is predominantly located in the nucleus and mitochondria. *Molecular biology of the cell*, 16(2), pp.997–1010.
- Cortez, D. et al., 1999. Requirement of ATM-dependent phosphorylation of brca1 in the DNA damage response to double-strand breaks. *Science (New York, N.Y.)*, 286(5442), pp.1162–6.
- Cram, E.J., Shang, H. & Schwarzbauer, J.E., 2006. A systematic RNA interference screen reveals a cell migration gene network in *C. elegans*. *Journal of cell science*, 119(Pt 23), pp.4811–8.

- Dacheux, E. et al., 2013. BRCA1-Dependent Translational Regulation in Breast Cancer Cells. P. J. Tofilon, ed. *PloS one*, 8(6), p.e67313.
- Debeir, O. et al., 2005. Tracking of migrating cells under phase-contrast video microscopy with combined mean-shift processes. *IEEE transactions on medical imaging*, 24(6), pp.697–711.
- Deshaies, R.J. & Joazeiro, C.A.P., 2009. RING domain E3 ubiquitin ligases. *Annual review of biochemistry*, 78, pp.399–434.
- Dhillon, K.K., Swisher, E.M. & Taniguchi, T., 2011. Secondary mutations of BRCA1/2 and drug resistance. *Cancer science*, 102(4), pp.663–9.
- Dizin, E. et al., 2006. BRCA1 interacts with poly(A)-binding protein: implication of BRCA1 in translation regulation. *The Journal of biological chemistry*, 281(34), pp.24236–46.
- Dmitrieva, N.I., Malide, D. & Burg, M.B., 2011. Mre11 is expressed in mammalian mitochondria where it binds to mitochondrial DNA. *American Journal of Physiology - Regulatory Integrative and Comparative Physiology*, 301(3), pp.R632–R640.
- Eakin, C.M. et al., 2007. Estrogen receptor alpha is a putative substrate for the BRCA1 ubiquitin ligase. *Proceedings of the National Academy of Sciences of the United States of America*, 104(14), pp.5794–9.
- Easton, D.F. et al., 1993. Genetic linkage analysis in familial breast and ovarian cancer: results from 214 families. The Breast Cancer Linkage Consortium. *American journal of human genetics*, 52(4), pp.678–701.
- Easton, D.F., Ford, D. & Bishop, D.T., 1995. Breast and ovarian cancer incidence in BRCA1-mutation carriers. Breast Cancer Linkage Consortium. *American journal of human genetics*, 56(1), pp.265–71.
- Elliott, B.E. et al., 2005. The membrane cytoskeletal crosslinker ezrin is required for metastasis of breast carcinoma cells. *Breast cancer research : BCR*, 7(3), pp.R365–73.
- Emanuele, M.J. et al., 2011. Global identification of modular cullin-RING ligase substrates. *Cell*, 147(2), pp.459–474.
- Esteller, M. et al., 2000. Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors. *Journal of the National Cancer Institute*, 92(7), pp.564–9.
- Even-Ram, S. et al., 2007. Myosin IIA regulates cell motility and actomyosin-microtubule crosstalk. *Nature cell biology*, 9(3), pp.299–309.
- Fabbro, M. et al., 2002. BARD1 induces BRCA1 intranuclear foci formation by increasing RING-dependent BRCA1 nuclear import and inhibiting BRCA1 nuclear export. *Journal of Biological Chemistry*, 277(24), pp.21315–21324.

- Fabbro, M. et al., 2004. BARD1 regulates BRCA1 apoptotic function by a mechanism involving nuclear retention. *Experimental Cell Research*, 298(2), pp.661–673.
- Fiaschi, T. et al., 2013. Carbonic anhydrase IX from cancer-associated fibroblasts drives epithelial-mesenchymal transition in prostate carcinoma cells. *Cell cycle (Georgetown, Tex.)*, 12(11), pp.1791–801.
- Fic, E. et al., 2010. Comparison of protein precipitation methods for various rat brain structures prior to proteomic analysis. *Electrophoresis*, 31(21), pp.3573–9.
- Fievet, B.T. et al., 2004. Phosphoinositide binding and phosphorylation act sequentially in the activation mechanism of ezrin. *The Journal of cell biology*, 164(5), pp.653–9.
- Figueroa, J.A., Yee, D. & McGuire, W.L., 1993. Prognostic indicators in early breast cancer. *The American journal of the medical sciences*, 305(3), pp.176–82.
- Foulkes, W.D. et al., 2003. Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer. *Journal of the National Cancer Institute*, 95(19), pp.1482–1485.
- Futreal, P.A. et al., 1994. BRCA1 mutations in primary breast and ovarian carcinomas. *Science (New York, N.Y.)*, 266(5182), pp.120–2.
- Girgis, H.S. et al., 2007. A comprehensive genetic characterization of bacterial motility. D. S. Guttman, ed. *PLoS genetics*, 3(9), pp.1644–60.
- Goshima, G. & Vale, R.D., 2005. Cell cycle-dependent dynamics and regulation of mitotic kinesins in *Drosophila* S2 cells. *Molecular biology of the cell*, 16(8), pp.3896–907.
- Grabbe, C., Husnjak, K. & Dikic, I., 2011. The spatial and temporal organization of ubiquitin networks. *Nature reviews. Molecular cell biology*, 12(5), pp.295–307.
- Griffin, N.M. et al., 2010. Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis. *Nature biotechnology*, 28(1), pp.83–9.
- Grillo, J. et al., Regulation of the angiogenesis inhibitor thrombospondin-1 by the breast cancer susceptibility gene-1 (BRCA1). *The breast journal*, 17(4), pp.434–5.
- Han, Y.H. et al., 2009. The effect of MG132, a proteasome inhibitor on HeLa cells in relation to cell growth, reactive oxygen species and GSH. *Oncology reports*, 22(1), pp.215–21.
- Hanahan, D. & Weinberg, R.A., 2011. Hallmarks of cancer: the next generation. *Cell*, 144(5), pp.646–674.
- Hand, A.J. et al., 2009. Automated tracking of migrating cells in phase-contrast video microscopy sequences using image registration. *Journal of microscopy*, 234(1), pp.62–79.
- Haralick, R.M. & Shapiro, L.G., 1993. *Computer and robot vision, Volume 2*, Addison-Wesley Pub. Co.

- Hashizume, R. et al., 2001. The RING heterodimer BRCA1-BARD1 is a ubiquitin ligase inactivated by a breast cancer-derived mutation. *The Journal of biological chemistry*, 276(18), pp.14537–40.
- Hitchcock, A.L. et al., 2003. A subset of membrane-associated proteins is ubiquitinated in response to mutations in the endoplasmic reticulum degradation machinery. *Proceedings of the National Academy of Sciences of the United States of America*, 100(22), pp.12735–40.
- Hjerpe, R. et al., 2009. Efficient protection and isolation of ubiquitylated proteins using tandem ubiquitin-binding entities. *EMBO reports*, 10(11), pp.1250–8.
- Hogstrand, C. et al., 2013. A mechanism for epithelial-mesenchymal transition and anoikis resistance in breast cancer triggered by zinc channel ZIP6 and STAT3 (signal transducer and activator of transcription 3). *The Biochemical journal*, 455(2), pp.229–37.
- Hsu, N.C. et al., 2013. Methylation of BRCA1 promoter region is associated with unfavorable prognosis in women with early-stage breast cancer. *PloS one*, 8(2), p.e56256.
- Huang, D.W., Sherman, B.T. & Lempicki, R.A., 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols*, 4(1), pp.44–57.
- Huen, M.S.Y., Sy, S.M.H. & Chen, J., 2010. BRCA1 and its toolbox for the maintenance of genome integrity. *Nature reviews. Molecular cell biology*, 11(2), pp.138–48.
- Iqbal, J. et al., 2012. The incidence of pancreatic cancer in BRCA1 and BRCA2 mutation carriers. *British journal of cancer*, 107(12), pp.2005–9.
- Jadeski, L. et al., 2008. IQGAP1 stimulates proliferation and enhances tumorigenesis of human breast epithelial cells. *The Journal of biological chemistry*, 283(2), pp.1008–17.
- Jedrychowski, M.P. et al., 2011. Evaluation of HCD- and CID-type fragmentation within their respective detection platforms for murine phosphoproteomics. *Molecular & cellular proteomics : MCP*, 10(12), p.M111.009910.
- Jeon, H.B. et al., 2007. A proteomics approach to identify the ubiquitinated proteins in mouse heart. *Biochemical and biophysical research communications*, 357(3), pp.731–6.
- Johnson, M., Sharma, M. & Henderson, B.R., 2009. IQGAP1 regulation and roles in cancer. *Cellular signalling*, 21(10), pp.1471–8.
- Jönsson, G. et al., 2012. The retinoblastoma gene undergoes rearrangements in BRCA1-deficient basal-like breast cancer. *Cancer research*, 72(16), pp.4028–36.

- Kaira, K. et al., 2011. Expression of excision repair cross-complementation group 1, breast cancer susceptibility 1, and β III-tubulin in thymic epithelial tumors. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer*, 6(3), pp.606–13.
- Kanehisa, M. et al., 2012. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic acids research*, 40(Database issue), pp.D109–14.
- Kanehisa, M. & Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 28(1), pp.27–30.
- Karpievitch, Y. et al., 2009. A statistical framework for protein quantitation in bottom-up MS-based proteomics. *Bioinformatics (Oxford, England)*, 25(16), pp.2028–34.
- Karpievitch, Y. V, Dabney, A.R. & Smith, R.D., 2012. Normalization and missing value imputation for label-free LC-MS analysis. *BMC bioinformatics*, 13 Suppl 1, p.S5.
- Khanna, C. et al., 2004. The membrane-cytoskeleton linker ezrin is necessary for osteosarcoma metastasis. *Nature medicine*, 10(2), pp.182–6.
- Kim, W. et al., 2011. Systematic and quantitative assessment of the ubiquitin-modified proteome. *Molecular cell*, 44(2), pp.325–40.
- King, M.C. et al., 2001. Tamoxifen and breast cancer incidence among women with inherited mutations in BRCA1 and BRCA2: National Surgical Adjuvant Breast and Bowel Project (NSABP-P1) Breast Cancer Prevention Trial. *JAMA : the journal of the American Medical Association*, 286(18), pp.2251–6.
- Kleiman, F.E. et al., 2005. BRCA1/BARD1 inhibition of mRNA 3' processing involves targeted degradation of RNA polymerase II. *Genes & development*, 19(10), pp.1227–37.
- Klemke, R.L., 1997. Regulation of Cell Motility by Mitogen-activated Protein Kinase. *The Journal of Cell Biology*, 137(2), pp.481–492.
- Köbel, M. et al., 2006. Ezrin promotes ovarian carcinoma cell invasion and its retained expression predicts poor prognosis in ovarian carcinoma. *International journal of gynecological pathology : official journal of the International Society of Gynecological Pathologists*, 25(2), pp.121–30.
- Komander, D., 2009. The emerging complexity of protein ubiquitination. *Biochemical Society transactions*, 37(Pt 5), pp.937–53.
- Kriebel, P.W. et al., 2008. Collective cell migration requires vesicular trafficking for chemoattractant delivery at the trailing edge. *The Journal of cell biology*, 183(5), pp.949–61.
- Laemmli, U.K., 1970. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature*, 227(5259), pp.680–685.

- Lai, Y. et al., 2011. Quantitative proteomics identifies the Myb-binding protein p160 as a novel target of the von Hippel-Lindau tumor suppressor. L. Zhang, ed. *PloS one*, 6(2), p.e16975.
- Lakhani, S.R. et al., 2000. The pathology of familial breast cancer: histological features of cancers in families not attributable to mutations in BRCA1 or BRCA2. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 6(3), pp.782–9.
- Lee, H.-C. & Wei, Y.-H., 2009. Mitochondrial DNA instability and metabolic shift in human cancers. *International journal of molecular sciences*, 10(2), pp.674–701.
- Lee, K.A. et al., 2011. Ubiquitin ligase substrate identification through quantitative proteomics at both the protein and peptide levels. *The Journal of biological chemistry*, 286(48), pp.41530–8.
- Lee, M.-N. et al., 2007. Epigenetic inactivation of the chromosomal stability control genes BRCA1, BRCA2, and XRCC5 in non-small cell lung cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 13(3), pp.832–8.
- Leongamornlert, D. et al., 2012. Germline BRCA1 mutations increase prostate cancer risk. *British Journal of Cancer*, 106(10), pp.1697–1701.
- Li, H. et al., 2010. Wild-type BRCA1, but not mutated BRCA1, regulates the expression of the nuclear form of beta-catenin. *Molecular cancer research : MCR*, 8(3), pp.407–20.
- Li, M.L. & Greenberg, R.A., 2012. Links between genome integrity and BRCA1 tumor suppression. *Trends in biochemical sciences*, 37(10), pp.418–24.
- Li, Q. et al., 2008. Ezrin silencing by small hairpin RNA reverses metastatic behaviors of human breast cancer cells. *Cancer letters*, 261(1), pp.55–63.
- Liang, C.-C., Park, A.Y. & Guan, J.-L., 2007. In vitro scratch assay: a convenient and inexpensive method for analysis of cell migration in vitro. *Nature protocols*, 2(2), pp.329–33.
- Liu, N.Q. et al., 2013. Quantitative Proteomic Analysis of Microdissected Breast Cancer Tissues: Comparison of Label-Free and SILAC-based Quantification with Shotgun, Directed, and Targeted MS Approaches. *Journal of proteome research*.
- Livasy, C.A. et al., 2006. Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*, 19(2), pp.264–71.
- Lopitz-Otsoa, F. et al., 2012. Integrative analysis of the ubiquitin proteome isolated using Tandem Ubiquitin Binding Entities (TUBEs). *Journal of proteomics*, 75(10), pp.2998–3014.

- Lotti, L.V. et al., 2002. Subcellular localization of the BRCA1 gene product in mitotic cells. *Genes, chromosomes & cancer*, 35(3), pp.193–203.
- Loyo, M. et al., 2011. A survey of methylated candidate tumor suppressor genes in nasopharyngeal carcinoma. *International journal of cancer. Journal international du cancer*, 128(6), pp.1393–403.
- Lucas, A.L. et al., 2013. High prevalence of BRCA1 and BRCA2 germline mutations with loss of heterozygosity in a series of resected pancreatic adenocarcinoma and other neoplastic lesions. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 19(13), pp.3396–403.
- Ma, Y. et al., 2010. BRCA1 regulates acetylation and ubiquitination of estrogen receptor-alpha. *Molecular endocrinology (Baltimore, Md.)*, 24(1), pp.76–90.
- Mak, H. et al., 2012. Ezrin phosphorylation on tyrosine 477 regulates invasion and metastasis of breast cancer cells. *BMC cancer*, 12, p.82.
- Mallery, D.L., Vandenberg, C.J. & Hiom, K., 2002. Activation of the E3 ligase function of the BRCA1/BARD1 complex by polyubiquitin chains. *The EMBO journal*, 21(24), pp.6755–62.
- Mani, A. & Gelmann, E.P., 2005. The ubiquitin-proteasome pathway and its role in cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 23(21), pp.4776–89.
- Maniccia, A.W. et al., 2009. Mitochondrial localization, ELK-1 transcriptional regulation and growth inhibitory functions of BRCA1, BRCA1a, and BRCA1b proteins. *Journal of Cellular Physiology*, 219(3), pp.634–641.
- Maxwell, C.A. et al., 2011. Interplay between BRCA1 and RHAMM regulates epithelial apicobasal polarization and may influence risk of breast cancer. *PLoS Biology*, 9(11).
- Meierhofer, D. et al., 2008. Quantitative analysis of global ubiquitination in HeLa cells by mass spectrometry. *Journal of proteome research*, 7(10), pp.4566–76.
- Melchor, L. et al., 2008. Distinct genomic aberration patterns are found in familial breast cancer associated with different immunohistochemical subtypes. *Oncogene*, 27(22), pp.3165–75.
- Merajver, S.D. et al., 1995. Germline BRCA1 mutations and loss of the wild-type allele in tumors from families with early onset breast and ovarian cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 1(5), pp.539–44.
- Michalski, A. et al., 2011. Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Molecular & cellular proteomics : MCP*, 10(9), p.M111.011015.

- Miki, Y. et al., 1994. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science*, 266(5182), pp.66–71.
- Mitulovic, G. & Mechtler, K., 2006. HPLC techniques for proteomics analysis--a short overview of latest developments. *Briefings in functional genomics & proteomics*, 5(4), pp.249–60.
- Moilanen, J. et al., 2003. Ezrin immunoreactivity in relation to survival in serous ovarian carcinoma patients. *Gynecologic oncology*, 90(2), pp.273–81.
- Moller, P. et al., 2007. Surveillance for familial breast cancer: Differences in outcome according to BRCA mutation status. *International journal of cancer. Journal international du cancer*, 121(5), pp.1017–20.
- Monteiro, A.N., 2000. BRCA1: exploring the links to transcription. *Trends in biochemical sciences*, 25(10), pp.469–74.
- Mootha, V.K. et al., 2003. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature genetics*, 34(3), pp.267–73.
- Moran, A. et al., 2012. Risk of cancer other than breast or ovarian in individuals with BRCA1 and BRCA2 mutations. *Familial Cancer*, 11(2), pp.235–242.
- Morris, J.R. et al., 2009. The SUMO modification pathway is involved in the BRCA1 response to genotoxic stress. *Nature*, 462(7275), pp.886–90.
- Neumann, B. et al., 2010. Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature*, 464(7289), pp.721–7.
- Nigg, E.A. & Raff, J.W., 2009. Centrioles, centrosomes, and cilia in health and disease. *Cell*, 139(4), pp.663–78.
- Olsen, J. V, Ong, S.-E. & Mann, M., 2004. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Molecular & cellular proteomics : MCP*, 3(6), pp.608–14.
- Orelli, B.J., Logsdon Jr, J.M. & Bishop, D.K., 2001. Nine novel conserved motifs in BRCA1 identified by the chicken orthologue. *Oncogene*, 20(32), pp.4433–8.
- Paull, T.T. et al., 2001. Direct DNA binding by Brca1. *Proceedings of the National Academy of Sciences of the United States of America*, 98(11), pp.6086–91.
- Peng, J. et al., 2003. A proteomics approach to understanding protein ubiquitination. *Nature biotechnology*, 21(8), pp.921–6.
- Proia, T.A. et al., 2011. Genetic predisposition directs breast cancer phenotype by dictating progenitor cell fate. *Cell Stem Cell*, 8(2), pp.149–163.

- Promkan, M. et al., 2009. BRCA1 modulates malignant cell behavior, the expression of survivin and chemosensitivity in human breast cancer cells. *International journal of cancer. Journal international du cancer*, 125(12), pp.2820–8.
- Quinn, J.E. et al., 2003. BRCA1 functions as a differential modulator of chemotherapy-induced apoptosis. *Cancer research*, 63(19), pp.6221–8.
- Raasi, S. et al., 2004. Binding of polyubiquitin chains to ubiquitin-associated (UBA) domains of HHR23A. *Journal of molecular biology*, 341(5), pp.1367–79.
- Reguart, N. et al., 2008. BRCA1: A new genomic marker for non-small-cell lung cancer. *Clinical Lung Cancer*, 9(6), pp.331–339.
- Reid, L.J. et al., 2008. E3 ligase activity of BRCA1 is not essential for mammalian cell viability or homology-directed repair of double-strand DNA breaks. *Proceedings of the National Academy of Sciences of the United States of America*, 105(52), pp.20876–81.
- Ren, L. et al., 2012. Dysregulation of ezrin phosphorylation prevents metastasis and alters cellular metabolism in osteosarcoma. *Cancer research*, 72(4), pp.1001–12.
- Robson, M.E., 2007. Treatment of Hereditary Breast Cancer. *Seminars in Oncology*, 34(5), pp.384–391.
- Rodriguez, J.A., Au, W.W.Y. & Henderson, B.R., 2004. Cytoplasmic mislocalization of BRCA1 caused by cancer-associated mutations in the BRCT domain. *Experimental cell research*, 293(1), pp.14–21.
- Rodríguez, J.A. & Henderson, B.R., 2000. Identification of a functional nuclear export sequence in BRCA1. *The Journal of biological chemistry*, 275(49), pp.38589–96.
- Rodriguez, M. et al., 2003. Phosphopeptide binding specificities of BRCA1 COOH-terminal (BRCT) domains. *The Journal of biological chemistry*, 278(52), pp.52914–8.
- Rodríguez-Pinilla, S.M. et al., 2006. Prognostic significance of basal-like phenotype and fascin expression in node-negative invasive breast carcinomas. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 12(5), pp.1533–9.
- Ross, S.H. et al., 2011. Ezrin is required for efficient Rap1-induced cell spreading. *Journal of cell science*, 124(Pt 11), pp.1808–18.
- Roy, R., Chun, J. & Powell, S.N., 2012. BRCA1 and BRCA2: different roles in a common pathway of genome protection. *Nat Rev Cancer*, 12(1), pp.68–78.
- Ruffner, H. et al., 1999. BRCA1 is phosphorylated at serine 1497 in vivo at a cyclin-dependent kinase 2 phosphorylation site. *Molecular and cellular biology*, 19(7), pp.4843–54.

- Ruffner, H. et al., 2001. Cancer-predisposing mutations within the RING domain of BRCA1: loss of ubiquitin protein ligase activity and protection from radiation hypersensitivity. *Proceedings of the National Academy of Sciences of the United States of America*, 98(9), pp.5134–9.
- Sadowski, M. & Sarcevic, B., 2010. Mechanisms of mono- and poly-ubiquitination: Ubiquitination specificity depends on compatibility between the E2 catalytic core and amino acid residues proximal to the lysine. *Cell division*, 5(1), p.19.
- Sahai, E., 2007. Illuminating the metastatic process. *Nature reviews. Cancer*, 7(10), pp.737–49.
- Salmena, L. & Pandolfi, P.P., 2007. Changing venues for tumour suppression: Balancing destruction and localization by monoubiquitylation. *Nature Reviews Cancer*, 7(6), pp.409–413.
- Sarrió, D. et al., 2006. Abnormal ezrin localization is associated with clinicopathological features in invasive breast carcinomas. *Breast cancer research and treatment*, 98(1), pp.71–9.
- Satija, Y.K., Bhardwaj, A. & Das, S., 2013. A portrayal of E3 ubiquitin ligases and deubiquitylases in cancer. *International journal of cancer. Journal international du cancer*, pp.1–10.
- Sato, K. et al., 2004. Nucleophosmin/B23 is a candidate substrate for the BRCA1-BARD1 ubiquitin ligase. *The Journal of biological chemistry*, 279(30), pp.30919–22.
- Schackmann, R.C.J. et al., 2013. Loss of p120-catenin induces metastatic progression of breast cancer by inducing anoikis resistance and augmenting growth factor receptor signaling. *Cancer research*, 73(15), pp.4937–49.
- Schlegel, B.P., Starita, L.M. & Parvin, J.D., 2003. Overexpression of a protein fragment of RNA helicase A causes inhibition of endogenous BRCA1 function and defects in ploidy and cytokinesis in mammary epithelial cells. *Oncogene*, 22(7), pp.983–91.
- Seery, L.T. et al., 1999. BRCA1 expression levels predict distant metastasis of sporadic breast cancers. *International journal of cancer. Journal international du cancer*, 84(3), pp.258–62.
- Shabbeer, S. et al., 2012. BRCA1 targets G2/M cell cycle proteins for ubiquitination and proteasomal degradation. *Oncogene*.
- Shakya, R. et al., 2011. BRCA1 tumor suppression depends on BRCT phosphoprotein binding, but not its E3 ligase activity. *Science*, 334(6055), pp.525–528.
- Shao, G. et al., 2009. MERIT40 controls BRCA1-Rap80 complex integrity and recruitment to DNA double-strand breaks. *Genes & development*, 23(6), pp.740–54.

- Shen, Z. et al., 2012. Kindlin-2: a novel adhesion protein related to tumor invasion, lymph node metastasis, and patient outcome in gastric cancer. *American journal of surgery*, 203(2), pp.222–9.
- Shi, X. & Wu, C., 2008. A suppressive role of mitogen inducible gene-2 in mesenchymal cancer cell invasion. *Molecular cancer research : MCR*, 6(5), pp.715–24.
- Shimizu, Y., Okuda-Shimizu, Y. & Hendershot, L.M., 2010. Ubiquitylation of an ERAD substrate occurs on multiple types of amino acids. *Molecular cell*, 40(6), pp.917–26.
- Sidani, M. et al., 2007. Cofilin determines the migration behavior and turning frequency of metastatic cancer cells. *The Journal of cell biology*, 179(4), pp.777–91.
- Silver, D.P. & Livingston, D.M., 2012. Mechanisms of BRCA1 tumor suppression. *Cancer discovery*, 2(8), pp.679–84.
- Simpson, K.J. et al., 2008. Identification of genes that regulate epithelial cell migration using an siRNA screening approach. *Nature cell biology*, 10(9), pp.1027–38.
- Smolen, G.A. et al., 2010. A genome-wide RNAi screen identifies multiple RSK-dependent regulators of cell migration. *Genes & development*, 24(23), pp.2654–65.
- Song, M. et al., 2011. Quantitative Proteomic Identification of the BRCA1 Ubiquitination Substrates. *Journal of Proteome Research*, 10(11), pp.5191–5198.
- Sorlie, T. et al., 2003. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proceedings of the National Academy of Sciences of the United States of America*, 100(14), pp.8418–23.
- Sørli, T. et al., 2001. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences of the United States of America*, 98(19), pp.10869–74.
- Sporn, J.C., Hothorn, T. & Jung, B., 2011. BARD1 expression predicts outcome in colon cancer. *Clinical Cancer Research*, 17(16), pp.5451–5462.
- Starita, L.M. et al., 2005. BRCA1/BARD1 ubiquitinate phosphorylated RNA polymerase II. *The Journal of biological chemistry*, 280(26), pp.24498–505.
- Starita, L.M. et al., 2004. BRCA1-dependent ubiquitination of gamma-tubulin regulates centrosome number. *Molecular and cellular biology*, 24(19), pp.8457–66.
- Starita, L.M. & Parvin, J.D., 2003. The multiple nuclear functions of BRCA1: transcription, ubiquitination and DNA repair. *Current opinion in cell biology*, 15(3), pp.345–50.
- Stoppa-Lyonnet, D. et al., 2000. Familial invasive breast cancers: worse outcome related to BRCA1 mutations. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 18(24), pp.4053–9.

- Struwing, J.P. et al., 1997. The risk of cancer associated with specific mutations of BRCA1 and BRCA2 among Ashkenazi Jews. *The New England journal of medicine*, 336(20), pp.1401–8.
- Subramanian, A. et al., 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, 102(43), pp.15545–50.
- Suchy, J. et al., 2010. BRCA1 mutations and colorectal cancer in Poland. *Familial cancer*, 9(4), pp.541–4.
- Tabb, D.L. et al., 2010. Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. *Journal of proteome research*, 9(2), pp.761–76.
- Taron, M. et al., 2004. BRCA1 mRNA expression levels as an indicator of chemoresistance in lung cancer. *Human molecular genetics*, 13(20), pp.2443–9.
- Taylor, K.M. et al., 2012. Protein kinase CK2 triggers cytosolic zinc signaling pathways by phosphorylation of zinc channel ZIP7. *Science signaling*, 5(210), p.ra11.
- Thakur, S. et al., 1997. Localization of BRCA1 and a splice variant identifies the nuclear localization signal. *Molecular and Cellular Biology*, 17(1), pp.444–452.
- The Cancer Geonme Atlas Network, 2012. Comprehensive molecular portraits of human breast tumours. *Nature*, 490(7418), pp.61–70.
- Thompson, D. & Easton, D.F., 2002. Cancer Incidence in BRCA1 mutation carriers. *Journal of the National Cancer Institute*, 94(18), pp.1358–65.
- Thompson, M.E. et al., 1995. Decreased expression of BRCA1 accelerates growth and is often present during sporadic breast cancer progression. *Nature genetics*, 9(4), pp.444–50.
- Tian, F. et al., 2013. BRCA1 promotes the ubiquitination of PCNA and recruitment of translesion polymerases in response to replication blockade. *Proceedings of the National Academy of Sciences of the United States of America*, 110(33), pp.13558–63.
- Tomlinson, G.E. et al., 1998. Characterization of a breast cancer cell line derived from a germ-line BRCA1 mutation carrier. *Cancer research*, 58(15), pp.3237–42.
- Trinh, H. V et al., 2013. iTRAQ-Based and Label-Free Proteomics Approaches for Studies of Human Adenovirus Infections. *International journal of proteomics*, 2013, p.581862.
- Trudgian, D.C. et al., 2011. Comparative evaluation of label-free SING normalized spectral index quantitation in the central proteomics facilities pipeline. *Proteomics*, 11(14), pp.2790–7.

- Trudgian, D.C. et al., 2010. CPFP: a central proteomics facilities pipeline. *Bioinformatics (Oxford, England)*, 26(8), pp.1131–2.
- Turner, N.C. et al., 2007. BRCA1 dysfunction in sporadic basal-like breast cancer. *Oncogene*, 26(14), pp.2126–32.
- Udeshi, N.D. et al., 2012. Methods for quantification of in vivo changes in protein ubiquitination following proteasome and deubiquitinase inhibition. *Molecular & cellular proteomics : MCP*, 11(5), pp.148–59.
- Vasilescu, J. et al., 2005. Proteomic analysis of ubiquitinated proteins from human MCF-7 breast cancer cells by immunoaffinity purification and mass spectrometry. *Journal of proteome research*, 4(6), pp.2192–200.
- Waddell, N. et al., 2010. Subtypes of familial breast tumours revealed by expression and copy number profiling. *Breast cancer research and treatment*, 123(3), pp.661–77.
- Wagner, S.A. et al., 2011. A proteome-wide, quantitative survey of in vivo ubiquitylation sites reveals widespread regulatory roles. *Molecular & cellular proteomics : MCP*, 10(10), p.M111.013284.
- Wallace, D.C., 2012. Mitochondria and cancer. *Nature reviews. Cancer*, 12(10), pp.685–98.
- Wang, W., Eddy, R. & Condeelis, J., 2007. The cofilin pathway in breast cancer invasion and metastasis. *Nature reviews. Cancer*, 7(6), pp.429–40.
- Wang, X. et al., 2006. Analysis of cell migration using whole-genome expression profiling of migratory cells in the Drosophila ovary. *Developmental cell*, 10(4), pp.483–95.
- Ware, M., Wells, A. & Lauffenburger, D., 1998. Epidermal growth factor alters fibroblast migration speed and directional persistence reciprocally and in a matrix-dependent manner. *J. Cell Sci.*, 111(16), pp.2423–2432.
- Wei, M. et al., 2005. BRCA1 promoter methylation in sporadic breast cancer is associated with reduced BRCA1 copy number and chromosome 17 aneusomy. *Cancer research*, 65(23), pp.10692–9.
- Williams, R.S. et al., 2003. Detection of protein folding defects caused by BRCA1-BRCT truncation and missense mutations. *The Journal of biological chemistry*, 278(52), pp.53007–16.
- Wilson, C.A. et al., 1999. Localization of human BRCA1 and its loss in high-grade, non-inherited breast carcinomas. *Nature Genetics*, 21(2), pp.236–240.
- Wu, W. et al., 2007. BRCA1 ubiquitinates RPB8 in response to DNA damage. *Cancer research*, 67(3), pp.951–8.
- Wu, W. et al., 2008. The ubiquitin E3 ligase activity of BRCA1 and its biological functions. *Cell division*, 3, p.1.

- Wu, Z.-Q. et al., 2012. Canonical Wnt signaling regulates Slug activity and links epithelial-mesenchymal transition with epigenetic Breast Cancer 1, Early Onset (BRCA1) repression. *Proceedings of the National Academy of Sciences of the United States of America*, 109(41), pp.16654–9.
- Wu-Baer, F. et al., 2003. The BRCA1/BARD1 heterodimer assembles polyubiquitin chains through an unconventional linkage involving lysine residue K6 of ubiquitin. *The Journal of biological chemistry*, 278(37), pp.34743–6.
- Xia, Y. et al., 2003. Enhancement of BRCA1 E3 ubiquitin ligase activity through direct interaction with the BARD1 protein. *The Journal of biological chemistry*, 278(7), pp.5255–63.
- Xu, B., Kim St & Kastan, M.B., 2001. Involvement of Brca1 in S-phase and G(2)-phase checkpoints after ionizing irradiation. *Molecular and cellular biology*, 21(10), pp.3445–50.
- Xu, G., Paige, J.S. & Jaffrey, S.R., 2010. Global analysis of lysine ubiquitination by ubiquitin remnant immunoaffinity profiling. *Nature biotechnology*, 28(8), pp.868–73.
- Yamane, K., Katayama, E. & Tsuruo, T., 2000. The BRCT regions of tumor suppressor BRCA1 and of XRCC1 show DNA end binding activity with a multimerizing feature. *Biochemical and biophysical research communications*, 279(2), pp.678–84.
- Yasmeen, A. et al., 2008. BRCA1 mutations contribute to cell motility and invasion by affecting its main regulators. *Cell cycle (Georgetown, Tex.)*, 7(23), pp.3781–3.
- Yu, X. et al., 2006. BRCA1 ubiquitinates its phosphorylation-dependent binding partner CtIP. *Genes & development*, 20(13), pp.1721–6.
- Yu, Y. et al., 2013. Kindlin 2 promotes breast cancer invasion via epigenetic silencing of the microRNA200 gene family. *International journal of cancer. Journal international du cancer*, 133(6), pp.1368–79.
- Zaarour, R.F. et al., 2012. Ezrin ubiquitylation by the E3 ubiquitin ligase, WWP1, and consequent regulation of hepatocyte growth factor receptor activity. R. E. Means, ed. *PloS one*, 7(5), p.e37490.
- Zhang, J. et al., 2012. PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Molecular & cellular proteomics : MCP*, 11(4), p.M111.010587.
- Zhang, Z.-Z. et al., 2013. Loss of BRCA1 expression leads to worse survival in patients with gastric carcinoma. *World journal of gastroenterology : WJG*, 19(12), pp.1968–74.
- Zhao, T. et al., 2013. Kindlin-2 promotes genome instability in breast cancer cells. *Cancer letters*, 330(2), pp.208–16.

Zhu, Q. et al., 2011. BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. *Nature*, 477(7363), pp.179–84.