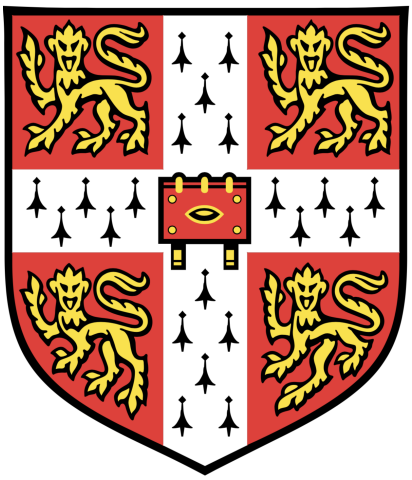
The role of NEDD9 in TGFβ mediated tumour initiating cell dynamics in the Claudin-low breast cancer subtype



Wendy Fay Greenwood

Cancer Research UK - Cambridge Institute
University of Cambridge

This thesis is submitted for the degree of
Doctor of Philosophy

Girton college

March 2021

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text.

This thesis is the result of work carried out at the Cancer Research UK Cambridge Institute between April 2012 to March 2021. It is not substantially the same as any that I have submitted or is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

This thesis does not exceed the prescribed word limit of 60,000 words for the Clinical Medicine and Veterinary Medicine Degree Committee, excluding tables, appendices and bibliography.

Wendy Fay Greenwood

March 2021

Abstract

Title: The role of NEDD9 in TGFβ mediated tumour initiating cell dynamics in the Claudin-low breast cancer subtype

Author: Wendy Fay Greenwood

Breast cancer is an extremely heterogeneous disease comprising at least ten different subtypes, each exhibiting different characteristics in progression, prognosis and response to treatment.

Transforming growth factor β (TGFβ) a member of the TGF superfamily of cytokines differentially regulates breast tumour-initiating cells (BTICs). In the claudin-low breast cancer subtype, TGFβ increases the tumour initiating capacity through the ability of the scaffolding protein NEDD9 to unite the TGFβ/SMAD and Rho-Actin-SRF pathways. Previously, oncogenicity has been ascribed to the level of NEDD9 protein expression. However, my analysis indicates this mediation by NEDD9 is irrespective of NEDD9 protein expression, which is ubiquitously overexpressed in the majority of cancer types including breast cancer.

In this thesis, I demonstrate how in the Claudin-low breast cancer subtype NEDD9 protein expression and post-translational modification are influenced by TGFβ pathway activation. I identify significant NEDD9 interacting proteins and their downstream effectors which contribute towards oncogenic TGFβ signalling pathways. A key TGFβ specific NEDD9 interactor identified in this subtype is the metabolic isoenzyme PKM2. Through a variety of techniques, I explore the known roles of PKM2 in the regulation of oncogenic metabolic reprogramming and demonstrate how these processes are influenced by its association with NEDD9. Finally, I investigate potential translational applications and biomarkers of TGFβ mediated, NEDD9/PKM2 dependent downstream signalling pathways, using large clinical breast cancer datasets and patient-derived tumour xenograft (PDTX) models. These data suggest a novel mechanism by which oncogenic TGFβ signalling regulates cellular proliferation and self-renewal via β-Catenin/c-Myc regulation of altered metabolism in the Claudin-low breast cancer subtype, a process which is dependent upon the scaffolding protein NEDD9 and the metabolic enzyme PKM2.

Together, these data suggest that a combined biomarker of TGFβ signalling and c-Myc expression may be useful in identifying a subset of Claudin-low breast cancer patients who would be sensitive to inhibition of Wnt/β-Catenin signalling. Additionally, due to the dependence of these tumours on c-Myc driven glutamine dependent metabolic processes, metabolic magnetic resonance imaging may be useful for monitoring response in these patients.

Acknowledgements

I would like to acknowledge the following people who have all been instrumental in enabling me to complete this work:

My supervisors Professor Carlos Caldas and Dr Alejandra Bruna for having faith in my abilities, and permitting me to fulfil my dream of completing a PhD. Also, Dr Alejandra Bruna for being a mentor and great friend through the challenges I have faced, both professionally and personally, and for teaching me everything I know about breast cancer biology.

My good friend and neighbour Dr Stephen Doyle who stepped up at short notice to review my manuscript.

Dr Oscar Rueda and Raquel Manzano Garcia for their excellent bioinformatics support.

The CRUK-CI tumour modelling team especially Lisa Young, Yi Cheng and Steven Kupczak, along with the entire BRU team for their dedication to their work and the animals they care for.

All the staff past and present at the Cambridge Breast Cancer Research Unit for their support and collaboration.

The CRUK-CI core facilities for all their help and expertise.

Cancer Research UK and the University of Cambridge for funding my research

And finally, all of my fantastic friends in the Caldas lab past and present, for making the last 12 years a blast!

Dedication

I dedicate this thesis to my loving parents and their ceaseless support, my wonderful, patient husband and my three beautiful sons Isaac, Seth and Alexander. Together your love and support has empowered me to achieve my dreams.

This work is also dedicated to the hundreds of breast cancer patients who gave so generously, that future generations might not suffer as they did.

Table of Contents

List of figures

List of tables

List of abbreviations

BCaPE - Breast cancer PDTX encyclopedia

BTIC - Breast tumour initiating cell

CAS - Crk associated substrate

CN - Copy number

CNA - Copy number aberration

CSC - cancer stem cell

DMSO - Dimethyl sulfoxide

ECAR - Extracellular acidification rate

ECM - extracellular matrix

ELISA - Enzyme-linked immunosorbent assay

EMT - Epithelial to mesenchymal transition

ES - Enrichment score

FAK - focal adhesion kinase

FDR - False discovery rate

GCMS - Gas-chromatography mass-spectrometry

GEMM - Genetically engineered mouse models

GPCR - G protein-coupled receptor

HEF1 - Enhancer of Filamentation 1

HER2 - human epidermal growth factor receptor 2

IHC - Immunohistochemistry

LCMS - Liquid chromatography-mass spectrometry

M1 - First-generation mammosphere assay

M2 - Second-generation mammosphere assay

MS - Mass spectrometry

MS-IC - Mammosphere initiating cell

NEDD9 - Neural Precursor Cell Expressed, Developmentally Down-Regulated 9

NHEJ - Non-homologous end-joining

NSCLC - Non-small cell lung carcinoma

OCR - Oxygen consumption rate

PARP - Poly(ADP-ribose) polymerase

PDTC - Patient-derived tumour cells

PDTX - Patient-derived tumour xenografts

PET - positron emission tomography

PIP - Protein Identification Probability

PPP - Pentose phosphate pathway

PR - Progesterone receptor

RIME - Rapid immunoprecipitation mass spectrometry of endogenous proteins

ROS - reactive oxygen species

RPPA - Reverse-phase protein array

SILAC - Stable isotope labelling by amino acids in cell culture

SB - TGFβ inhibitor SB431542

SD - Substrate domain

TCA - Tricarboxylic acid

TGFβ - Transforming growth factor-beta

TIC - Tumour initiating cell

TMA - Tissue microarray

TMT - Tandem mass tags

TMT-RIME - Tandem mass tag-rapid immunoprecipitation mass spectrometry of endogenous proteins

TPM - Transcripts per million

1. Introduction

1.1 Breast cancer incidence

Breast cancer is the most commonly diagnosed cancer in the world. In the UK, it accounts for approximately 15% of all new cancer cases and is the leading cause of cancer mortality in women (Torre et al. 2016). Between 2015 and 2017, there were in the order of 55,200 new breast cancer cases diagnosed each year in the UK (www.cancerresearchuk.org). Breast cancer is curable in 70-80% of early-stage disease, however, once metastasis occurs, advanced breast cancer is considered incurable (Harbeck et al. 2019). In the clinical setting, breast cancer is stratified by measures of tumour growth (grade) size and local/regional spread (stage), and by the expression of cell surface and nuclear markers, and known genetic markers of predisposition. These classifications allow for informed treatment decisions to be made by oncologists at a semi-personalised level, however, it is clear that the intrinsic molecular heterogeneity of breast cancer means this "one size fits all" approach is not sufficient in all cases. At present, treatment for breast cancer is broadly divided into two categories, loco-regional (surgery and radiotherapy) and systemic approaches. Systemic treatments tend to be based on the expression of nuclear hormone receptors for oestrogen receptor (ER), progesterone receptor (PR) and the cell surface marker, human epidermal growth factor receptor 2 (HER2, encoded by the ERBB2 gene). Patients with ER-positive and/or HER2 overexpressing disease are generally stratified for surgery to remove the tumour followed by long-term endocrine therapy, whereas patients with hormone receptor-negative disease generally receive chemotherapeutic agents, followed by surgery. More recently, poly(ADP-ribose) polymerase (PARP) inhibitors have become common for use in patients with BRCA mutations. This "broad-brush" approach to cancer treatment is generally successful, however, in approximately 20% of treated patients, recurrence will occur and these patients will inevitably succumb to advanced metastatic disease. This continuing mortality despite advances in screening, early detection and targeted therapy highlights the fact that there are many complex mechanisms of disease progression and resistance to therapy that remain yet understood.

1.2 Tumour heterogeneity

Breast cancer is an extremely heterogeneous disease at both inter- and intra-tumoural levels. Major differences in the biological and histological properties of breast cancer are due to variations at the

genetic, epigenetic and transcriptomic level, and these differences result in the variations in prognosis and treatment response making targeted treatment challenging, and drug resistance and relapse common (Polyak 2011). In 2000, Perou and Sørlie described a new classification of breast cancer based on gene expression in 42 patients (Perou et al. 2000). Their classification distinguished five intrinsic breast cancer subtypes; Normal breast-like, Luminal A and Luminal B (oestrogen receptor (ER) positive), Basal-like (hormone receptor and HER2 negative) and HER2 overexpressing (but ER receptor negative). In general, the ER-negative subtypes (basal and HER2 overexpressing) exhibit poorer patient outcomes with regards to overall survival when compared to the ER-positive groups. Following on from this seminal publication, a major shift in the focus for breast cancer treatment and management occurred, from being predominantly based on tumour burden to becoming tumour-biology based (Sørlie et al. 2001), and to this day, classification based on the expression of these markers has remained the principal method used to stratify breast cancer patients for treatment. However, what these classifications fail to take into account are the complex levels of intra-tumour heterogeneity present in breast cancer, and it is this heterogeneity along with the process of tumour evolution that underlies drug resistance and relapse. In 2012, Curtis and colleagues published an in-depth analysis of 2,000 breast tumours using copy number, gene expression and somatic mutation analysis with long term clinical follow to further characterise breast cancer into ten molecular subtypes (Curtis et al. 2012; Pereira et al. 2016). This work allows us to better understand the patient prognosis and therapeutic response and to more effectively utilise targeted therapy. More recently, studies at the single-cell level have further highlighted the significance of inter-tumour phenotypic heterogeneity and the importance of altered transcriptional programmes regulated by non-genomic processes in breast cancer. These data demonstrate how treatment response and the emergence of drug resistance are influenced by such processes (Shah et al. 2012, 2009; Callari et al. 2021)). Many groups have demonstrated that several factors contribute to tumour heterogeneity, including genetic mutations both germ-line and somatic, epigenetic alterations, tumour cell interaction with the microenvironment plus the existence or not of a cellular hierarchy in cancer.

1.2.1 Cancer cell hierarchy

The frank heterogeneity of breast cancer has led researchers to hypothesise that different breast tumour subtypes originate from distinct cellular compartments within the mammary gland (Izumchenko et al. 2009; Perou et al. 2000; Hennessy et al. 2009). Increasing evidence exists that there is a subpopulation of cells within cancer that retain stem cell-like features and are responsible for tumour initiation, resistance to therapy, recurrence and metastasis (Al-Hajj et al. 2003; Shan et al. 2021). Cells belonging to this subpopulation, commonly known as tumour initiating cells (TICs) or cancer stem cells (CSC) are defined by their ability to self-renew in both *in vitro* and *in vivo* assays, and by their expression of stem cell-like

gene signatures (Wicha 2014). CSCs were first identified as a $CD34^{+}/CD38^{-}$ subpopulation of cells in human acute myeloid leukaemia, which when transplanted into immunocompromised mice were capable of recapitulating the disease as a whole (Bonnet and Dick 1997), and CSCSs have now been identified in a plethora of cancer types (Sun et al. 2019).

Two mechanisms of hierarchy have been proposed to account for the existence of CSCs; the clonal evolution, or stochastic theory and the cancer stem cell (CSC) theory. The clonal evolution theory or stochastic model was first proposed by Peter Nowell in 1976 and hypothesises that a tumour arises from a single mutated cell and gathers mutations as it progresses (Nowell 1976). By a process of Darwinian evolution, subpopulations arise in response to pressure within the tumour microenvironment allowing clones with an evolutionary advantage to flourish (Swanton 2012). The cancer stem cell theory hypothesises that a small population of cells within the tumour, but not necessarily derived from normal tissue stem cells, is responsible for sustaining tumour growth and this promotion of growth is not regulated by genetic mutations. The theory proposes that these stem cells can both self-renew to generate daughter stem cells and differentiate into non-tumorigenic progeny (Shackleton et al. 2009).

Recent findings have begun to cast doubt on the existence of a "universal" CSC as great diversity in both inter-tumoural and intra-tumoural CSC phenotype has been observed. These findings confirm that the CSC theory is a valid concept but the reality is more likely a combination of the two theories. Many groups now favour the term tumour initiating cell (TIC) as a functional description of a subset of tumour cells with "stem-like" characteristics which are capable of establishing tumour growth. Several experimental protocols exist for the identification and quantification of CSCs in cell lines or primary human tissue; these are discussed in chapter 1.4.4.

1.3 Claudin-low breast cancer

Seven years on from the initial classification of breast cancer by Perou and Sørlie, a further subtype was identified through an integrated analysis of human and murine mammary tumours (Herschkowitz et al. 2007). This new subtype, named Claudin-low, was characterised by reduced expression of luminal genes and genes related to tight junctions and cell-adhesion, including *Claudin 3,4 and 7, E-Cadherin* and *Occludin*. In this analysis, the vast majority of these tumours were ER-negative and all were classified as infiltrating ductal carcinoma, grade II or III. More recently high expression of epithelial to mesenchymal transition (EMT) genes and stem cell-like gene expression patterns have been added to the characteristics

that define these tumours (Prat et al. 2010), along with pronounced stromal and immune cell infiltration (Dias et al. 2017). In 2020, Fougner and colleagues re-defined the accepted perception of the Claudin-low subtype through comprehensive genomic, transcriptomic and clinical analyses (Fougner et al. 2020). Claudin-low tumours are not defined by any specific set of genomic aberrations, and globally they exhibit a greater extent of copy number aberration (CNA) and variation in mutational burden. Therefore, rather than being a distinct subtype additional to those first described by Perou and Sørlie, Fougner et al conclude that Claudin-low tumours should be considered as complex entities with similarities across all intrinsic subtypes, and suggest that the term Claudin-low should be used to describe the breast cancer phenotype specific to these tumours.

1.4 Breast cancer models in research

1.4.1 Immortalised cancer cell lines

To fully capture the heterogeneity of breast cancer in research, we must have viable models to reflect this nature. Historically, immortalised breast cancer cell lines have been used for oncology research, drug discovery and preclinical studies. They are generally stable and clonal, preserving the primary somatic driver mutation, and tend to exhibit responses to pharmacological agents that are similar to the tumour of origin (H. S. Kim, Sung, and Paik 2015). Since the establishment of HeLa, the first immortalised human cancer cell line in the 1950s (Masters 2002), there have been thousands of cell lines generated across scores of cancer types. These have remained the mainstay of oncology research and drug development and are still widely used in research where alternatives are not practically or financially available and are also particularly useful in research into specific gene mutations or signalling pathways that can be easily perturbed in cell lines. More recently models involving xenotransplantation of established cell lines into immunocompromised mice have been utilised as *in vivo* tumour models in cancer research. Whilst this method has proven useful in studying metastatic processes and mechanisms of drug resistance (Kersten et al. 2017), they still fail to reflect the heterogeneity of human cancer.

1.4.2 Genetically engineered mouse models (GEMM)

The term (genetically engineered mouse models) GEMM covers a range of mouse models but generally refers to a mouse with an altered genome achieved by genetic engineering techniques to study a specific disease process or pathway. Conventional GEMM contain modifications in a targeted pathway or pathways specific to the disease they are being used to study, whereas OncoMice can bear either germline

mutations in tumour-suppressor genes or activated oncogenes under the control of tissue-specific promoters. These mice are utilised to accurately mimic sporadic human cancer, as tumours arising in these models closely mimic the molecular and histological features and the trajectory of the cancers they are imitating. More recently the use of conditional knockout models and CRISPR/Cas9 technology has allowed for targeted genome editing in somatic cells (Jonkers and Berns 2002; Platt et al. 2014), allowing for the *in vivo* functional analysis of genetic mutations in disease processes such as tumorigenesis and metastasis. However, particularly of importance is the use of CRISPR/Cas9 technology to study complex combinations of genetic mutations, such as gain-of-function mutations in proto-oncogenes, and loss-of-function mutations in tumour suppressor genes and their effects on the processes of tumorigenesis.

1.4.3 Patient-derived tumour xenografts (PDTX)

Patient-derived tumour xenografts (PDTX) are *in vivo* models derived from the implantation of fresh tumour tissue biopsies into immunocompromised mice. Unlike cell line transplantation methods, PDTX models retain the intra-tumour heterogeneity, as well as the genetic, molecular and histological features seen in the originating tumour even following serial passage (Cassidy, Caldas, and Bruna 2015). (Bruna et al. 2016). Because of this ability, PDTX models possess a clear advantage over traditional models when used in oncology research, drug discovery, and preclinical studies. Historically, where cell line panels have been used in preclinical studies (H. S. Kim, Sung, and Paik 2015), the inherent inability of cell lines to capture the full heterogeneity of human tumours means only a handful of drugs in development ever reach the clinic. Bruna et al described a methodology for the use of patient-derived tumour cells (generated from PDTX) in high-throughput drug testing, demonstrating a robust platform for preclinical drug studies (Bruna et al. 2016). Of course, PDTX models also present limitations, for example, the engraftment rate for tumours is extremely variable with ER-positive breast cancer, and these models often require prolonged periods and re-engraftment to become established. Another limitation is the necessity to use immunocompromised mice for engraftment, therefore bypassing interactions between the innate immune system and cancer cells that would otherwise be present. However, even taking into account these limitations, PDTX models represent a promising and powerful resource for preclinical drug development studies, and also allows for complex *in vivo* investigation of the processes of drug resistance, recurrence and clonal evolution.

1.4.4 Cancer stem cell cultures

Many *in vitro* and *in vivo* methods are routinely used for the identification and quantification of CSCs from cell lines, human normal or human tumour tissue. The mammosphere assay used in breast cancer

research was adapted from the neurosphere assay developed by Reynolds et al (Reynolds and Weiss 1992) and has become a quick, cost-effective alternative to the gold standard method of the *in vivo* limiting dilution assay (Dontu et al. 2003). Briefly, the mammosphere assay incorporates two separate methods; a first-generation assay (M1) where cells are dissociated to a single cell suspension and plated at clonal density on a non-adherent substrate, and a second-generation assay (M2) as a readout of CSC numbers from the first generation assay. The readout assay usually comprises a second-generation mammosphere assay and a colony-forming cell assay, measuring respectively the quantity of CSC in the original sample, and the ability of these CSCs to proliferate and differentiate when seeded upon an adherent substrate. As mentioned above, the gold standard method for assessing CSC numbers within a cell population is the *in vivo* limiting dilution assay. This method involves the transplantation of single-cell suspensions from M1 assays into the mammary fat pad of immunocompromised mice in limiting dilution. The L-Calc™ limiting dilution software (StemCell technologies) is then used to calculate the number of CSC in the originating sample. The above methods represent extremely useful tools in estimating CSC populations within samples of interest, especially when comparing paired sets of data, for example in response to treatment.

1.5 Growth factor signalling and cancer

Under normal conditions, tissue homeostasis is tightly controlled, primarily by fate-determining signals received by cells via polypeptide growth factors released from the surrounding tissues. Although the initial transformation from normal cells to malignant cells is due to the accumulation of genetic mutations, growth factors are the major regulators of tumorigenesis, controlling processes such as cellular adhesion and migration, angiogenesis, and tumour cell extravasation and invasion of distant sites. A key regulator in the control of epithelial cell homeostasis is Transforming growth factor-beta.

1.5.1 TGFβ and the SMAD Signalling Pathway

Transforming growth factor-beta (TGFβ) is a member of the TGFβ superfamily of cytokines important in organ development and tissue homeostasis and the TGFβ signalling pathway was one of the first pathways controlling multicellular life to emerge with the appearance of the first animal species (Huminiecki et al. 2009). TGFβ is a pleiotropic cytokine and is instrumental in the regulation of numerous cellular processes including cell proliferation, differentiation, apoptosis, motility, invasion, angiogenesis, immune response and extracellular matrix (ECM) production. Through its ability to activate specific transcription factors TGFβ is known to regulate hundreds of TGFβ target genes, consistent with its role in

a variety of processes (Mullen et al. 2011). Because of the crucial role of TGFβ in controlling cellular programmes regulating proliferation, differentiation and tissue regeneration and its importance in the evolution of early life it comes as no surprise that this pathway is often hijacked in diseases resulting from malfunctions in these processes, cancer for example (Massagué 2008).

TGFβ signalling has historically been understood to be propagated via the canonical TGFβ/SMAD pathway. However, increasing data on TGFβ activated non-canonical (non-SMAD) pathways are emerging. These non-canonical pathways include multiple components of the phosphatidylinositol-3-kinase (PI3K)/AKT, MAPK and Rho-like GTPase signalling pathways (Y. E. Zhang 2009).

The canonical TGFβ signalling pathway is activated by the binding of the TGFβ ligand to a heterodimeric complex of type I and type II membrane-bound serine/threonine-protein kinase receptors, TGFβRI and TGFβRII respectively. Upon ligand binding, TGFβRII phosphorylates the glycine- and serine-rich TGFβRI GS domain creating a binding site for the phosphorylation of TGFβ-specific transcription factors called SMADs. The designation SMAD is a portmanteau of the homologous proteins; Drosophila homolog mothers against decapentaplegic (MAD) and *Caenorhabditis elegans* protein SMA (Massagué, Seoane, and Wotton 2005). SMAD proteins are classified by their structure and function into three groups: 1) The receptor-regulated SMADs (R-SMAD), which contain a C-terminal Ser-Ser-X-Ser motif and are directly phosphorylated by type I receptor kinases; 2) The common mediator SMADs (co-SMAD), which complex with R-SMADs and translocate to the nucleus to mediate the regulation of TGFβ-specific target genes; and 3) the inhibitory SMADs (I-SMADs), which interfere with receptor activation of R-SMADs or complex formation with co-SMADs (Miyazono 2000).

Structurally, SMAD proteins comprise two globular domains, MH1 and MH2, joined by a variable linker region. The amino-terminal MH1 domain acts as a DNA binding region, and the C-terminal MH2 domain contains a series of hydrophobic regions which mediate SMAD interactions with both cytoplasmic and nuclear pore proteins, and with DNA-binding co-factors and signalling regulators (Massagué, Seoane, and Wotton 2005). The SMAD variable linker domain contains multiple phosphorylation sites for protein kinases including mitogen-activated protein kinases (MAPK) (Massagué, Seoane, and Wotton 2005; Kretzschmar, Doody, and Massagu 1997) and cyclin-dependent kinases (CDK), both of which are closely involved in the regulation of TGFβ/SMAD signalling (Matsuura et al. 2004). These phosphorylation domains allow SMADs to act as integration hubs for multiple regulatory processes, for example, cell cycle progression and growth factor signalling.

In the instance of canonical TGFβ signalling, the TGFβ type I receptor phosphorylates the R-SMADs SMAD2 and SMAD3, leading to their translocation to the nucleus and complex formation with the co-SMAD, SMAD4. Once in the nucleus, this complex recruits DNA-binding transcription factors resulting in transcriptional regulation and expression of TGFβ target genes. Signalling through TGFβ/Smads is a tightly regulated process and the SMAD protein complex translocates to the nucleus only whilst TGFβ receptor activation is taking place. This localisation is maintained through repeated phosphorylation/dephosphorylation and nuclear-cytoplasmic shuttling cycles. Following transcriptional activation, SMADs undergo rapid phosphorylation at the linker region by CDK8 and the cyclinT1/CDK9 complex. This phosphorylation primes the linker region for further phosphorylation by glycogen synthase kinase 3 (GSK3) and targets SMADs for polyubiquitylation and degradation in the proteasome (Massagué 2012).

TGFβ is crucially involved in the maintenance of tissue homeostasis. At the cellular level, TGFβ effects are well recognised for their pleiotropic nature. In development, embryonic stem cells and lineage-committed progenitors exhibit elevated levels of TGFβ signalling, where it functions to regulate pluripotency and differentiation. In mature epithelial, haematopoietic and neural cells, TGFβ negatively regulates the cell cycle through several mechanisms, including the expression of the Cyclin-dependent kinase inhibitors CDKN1A, CDKN1C, CDKN2B and repression of the proto-oncogene MYC (Siegel and Massagué 2003). Because of the crucial role of TGFβ signalling in organ development and in maintaining tissue homeostasis, deregulation of the pathway is commonly seen in diseases, including cancer.

1.5.2 TGFβ and cancer

TGFβ signalling is strongly implicated in many aspects of cancer progression. The TGFβ dogma supports that TGFβ functions mainly as a potent tumour suppressor in the early stages of cancer development through the potent cytostatic effects observed in the normal epithelium. Evidence of inactivation or loss of TGFβ pathway components in various cancer types strongly supports this hypothesis. For example, inactivating mutations in TGFβRII are frequently observed in cancers associated with defects in the DNA mismatch repair system, such as gastric, colorectal, biliary and lung adenocarcinoma (Levy and Hill 2006). Loss of SMAD2 and SMAD4 is commonly observed in pancreatic and colon cancers due to mutation, or loss of chromosome 18q (Massagué and Padua 2009; Massagué, Seoane, and Wotton 2005), and CDKN2B loss is frequent in melanoma (Massagué, Seoane, and Wotton 2005). In sharp contrast, however, other cancers progress with an intact TGFβ/SMAD pathway, where evasion of tumour-suppressive TGFβ activity occurs downstream of SMAD signalling. In this context which is most commonly seen in breast cancer, melanoma and glioma, the TGFβ/SMAD signalling pathway remains

intact but tumour growth is no longer suppressed by TGFβ signalling. Consistently, increased TGFβ activity is a marker of malignancy in non-small cell lung carcinoma (NSCLC), colorectal, prostate, gastric carcinoma, glioblastoma and breast cancer (Massagué and Padua 2009; Bruna et al. 2007, 2012; Massagué, Seoane, and Wotton 2005). In the later stages of disease, TGFβ plays key roles in the regulation of metastasis via the process of epithelial to mesenchymal transition (EMT). Here, TGFβ orchestrates signalling pathways that allow cancer cells to lose polarity, become motile and invade distant sites. The key event in the initiation of this process is the switch from expression of E-Cadherin to the expression of N-Cadherin, this switch is mediated by downstream effectors of TGFβ signalling which are transcriptional repressors of E-cadherin, namely SNAIL1, SLUG and TWIST (Michael K. Wendt, Allington, and Schiemann 2009).

In 2007, Bruna and colleagues presented evidence supporting opposing roles for TGFβ signalling in glioma, where it can both suppress and promote tumour formation (Bruna et al. 2007). More recently, under the leadership of Alejandra Bruna, my colleagues and I published further data providing evidence that in breast cancer, the paradoxical role of TGFβ activity is not necessarily linked to the stage of breast cancer progression (Bruna et al. 2012). Here, we demonstrate that oncogenic TGFβ signalling occurs in the molecularly defined Claudin-low subset of breast cancer, via the convergence of TGFβ/SMAD signalling and the Rho-MRTF-SRF pathway. The integration of these pathways drives a TGFβ-specific stem-cell-like gene expression program which is associated with worse outcomes in a large breast cancer clinical cohort, irrespective of ER status. In breast tumour initiating cell (BTIC) enriched populations of cell lines belonging to the Claudin-low subtype, TGFβ binding to cell surface receptors induces phosphorylation of the SMAD TGFβ-specific transcription factors, followed by SMAD partnering with co-SMADs leading to nuclear translocation. SMAD phosphorylation also leads to increased activation of Rho GTPase, which in turn activates the polymerisation of globular (G) actin into Filamentous (F) actin and results in the release of Myocardin-related transcription factor (MRTF). MRTF liberation induces the nuclear transcription factor serum response factor (SRF) to modulate expression of the gene signature enriched in stem-cell processes. Key in uniting the TGFβ/SMAD and Rho-MRTF-SRF pathways is the scaffolding protein NEDD9 which itself is subject to increased gene expression in response to TGFβ via a positive feedback loop (figure **1.1a**) (Bruna et al. 2012; Tufegdzic Vidakovic et al. 2015).

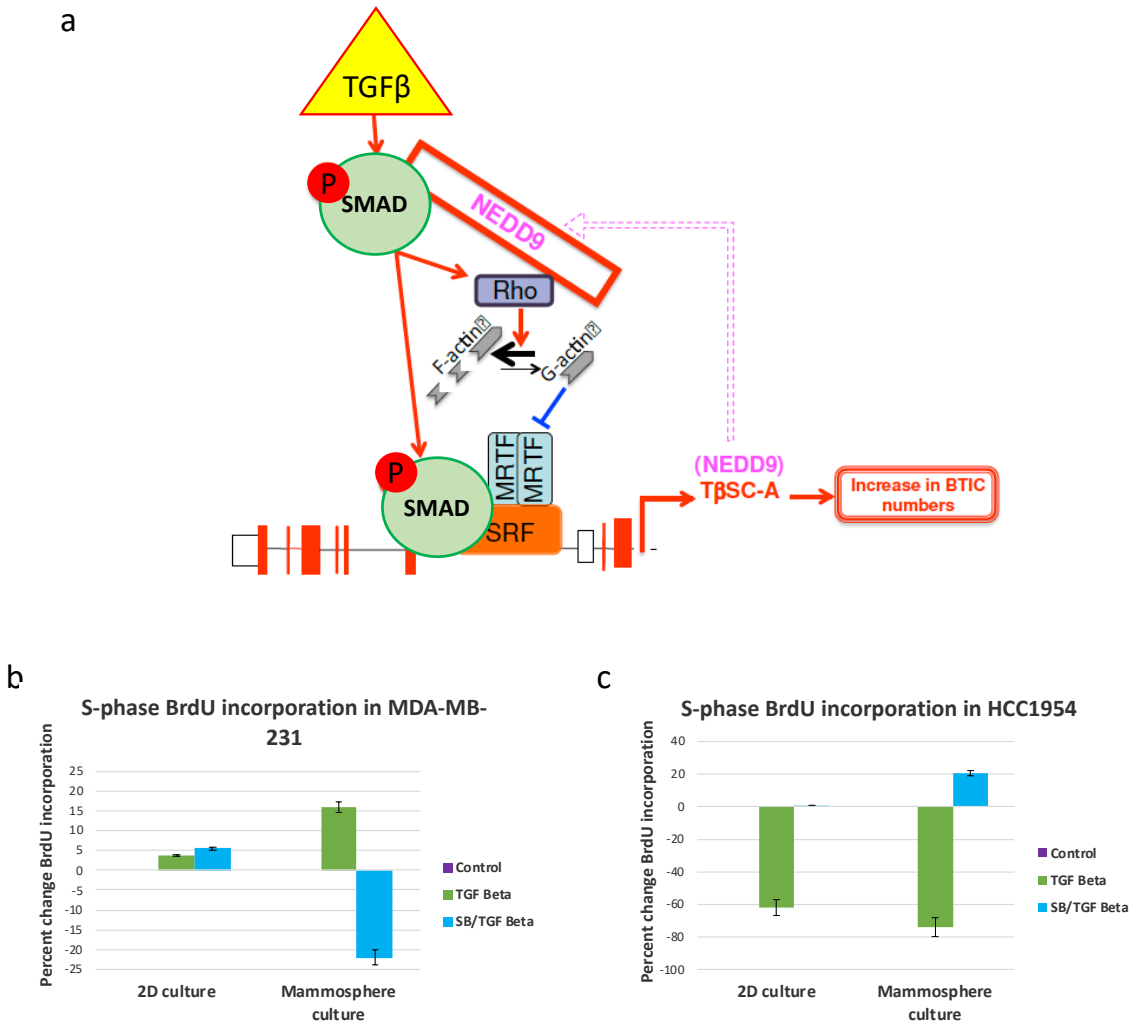


Figure 1.1. The mechanism of oncogenic TGFβ signalling in Claudin-low cell lines
(a) A NEDD9-Smad-SRF-dependent pathway regulates a TGFβ-specific transcriptional program which is associated with human stem cell features, and necessary for TGFβ-induced BTIC activity. Upon TGFβ binding to cell surface receptors phosphorylation of TGFβ-specific transcription factor SMADs occurs, followed by partnering with co-SMADs and nuclear translocation. SMAD phosphorylation also leads to increased Rho activity, this activates the polymerisation of globular (G) actin into Filamentous (F) actin and results in the release of Myocardin-related transcription factor (MRTF). MRTF liberation in turn induces the nuclear transcription factor serum response factor (SRF) to modulate expression of the gene signature enriched in stem-cell processes. Key in uniting these pathways is the scaffolding protein NEDD9 which itself is subject to increased gene expression in response to TGFβ via a positive feedback loop. Analysis of cellular proliferation in S-phase as measured by Bromodeoxyuridine (BrdU) incorporation in **(b)** MDA-MB-231 and **(c)** HCC1954 cultured both in attached and suspension mammosphere culture.

We propose that as a result of this gene expression program, following TGFβ treatment an increase in BTIC occurs as evidenced by Bromodeoxyuridine (BrdU) incorporation assays (figures **1.1b** and **1.1c**.). Here in the Claudin-low cell line MDA-MB-231, TGFβ treatment induced an increase in cellular proliferation as measured by BrdU incorporation, only when the cells were cultured as suspension mammospheres to enrich in BTICs (figure **1.1b**). Neither TGFβ induced cellular proliferation, nor proliferation differences between attached and suspension cultures were observed in the corresponding experiment in the non-Claudin-low cell line HCC1954 (figure **1.1c**). To date, other studies have failed to demonstrate a clear association between TGFβ-specific gene signatures and clinical outcomes, however, this is possibly because these studies did not isolate TGFβ specific gene signatures from the opposing tumour promoting and tumour suppressive branches.

The effects of TGFβ signalling in breast cancer should not be taken in the context of the cell-autonomous compartment alone; a significant number of TGFβ effects in cancer are mediated by the interaction between epithelial cells and the tumour microenvironment (stromal cells). Tumour stroma is a complex tissue comprising numerous cell types together with extracellular matrix proteins, all of which express TGFβ receptors and respond differently to TGFβ signalling. In contrast to the paradoxical role of TGFβ signalling in tumour epithelial cells, TGFβ signalling in tumour stromal cells mainly contributes towards promoting a favourable environment for tumour growth (Bruna et al. 2007). TGFβ signalling is also fundamental in immune regulation of the tumour microenvironment by inducing tumour immune-evasion, allowing tumour cells to evade host immune surveillance. TGFβ signalling suppresses the host immune system by many mechanisms including the inhibition of M1 macrophages and expansion of tumour-promoting M2 macrophages, the suppression of cytotoxic T cells and NK cells and an increase in the population of T-helper 2 cells which possess humoral but no cytotoxic activity (Siegel and Massagué 2003). All of the above observations highlight the complex nature of TGFβ regulation and its role in cancer, and the importance of fully understanding the pleiotropy of cellular response to TGFβ signalling.

1.5.3 TGFβ and Cancer therapy

The deregulation of TGFβ signalling in tumour initiation progression and metastasis has positioned TGFβ signalling as an increasingly interesting target for oncology drug development. However, because of TGFβ pleiotropic effects, signalling in normal epithelial cells and its paradoxical role in the growth regulation of different cancer types, TGFβ inhibition as a cancer therapy is not a straightforward prospect. The development of anti-TGFβ therapy to date has focussed on three main areas: 1) inhibition of TGFβ synthesis by antisense oligonucleotides either delivered intravenously or delivered as an allogeneic cancer cell vaccine via genetically engineered immune cells; 2) targeting TGFβ signalling at the ligand-receptor

TGFβ ligand inhibitors	Target	Trial ID	Status
Fresolimumab (GC1008) Sanofi/Aventis®	TGFβ1, -β2, -β3	NCT00356460 NCT00923169 NCT01472731 NCT01112293 NCT01401062	Phase I studies completed in RCC, melanoma and glioma. Phase II study completed in mesothelioma and breast cancer.
Trabedersen (AP12009) Antisens Pharma®	TGFβ2	NCT00844064 NCT00431561 NCT00761280	Phase I study in melanoma, pancreatic, and CRC completed. Phase II study in glioma completed. Phase III study in astrocytoma and secondary glioblastoma terminated.
Lucanix (Belagenpumatucel-L) NovaRx Corporation®	TGFβ2	NCT01058785 NCT00676507	Phase II and phase III study in NSCLC completed.
FANG™ Vaccine (rhGMCSF/shRNAfurin) Gradalis®	TGFβ1, -β2	NCT01061840 NCT01309230 NCT01505166 NCT01453361	Phase I study in NSCLC and liver cancer, ongoing. Phase II study in melanoma, ovarian and CRC ongoing.
Disitertide (P144) Digna Biotech®	TGFβ1		Progress outside oncology Pre-clinical development in glioma

TGFβ receptor inhibitors	Target	Trial ID	Status
Galunisertib (LY157299) Eli Lilly®	TGFβR1	NCT01246986 NCT01373164 NCT01220271 NCT02178358 NCT01582269	Phase I study in metastatic cancer completed. Phase II study in HCC, glioma and pancreatic cancer ongoing.
LY3200882 Eli Lilly®	TGFβR1	NCT02937272	Phase I study ongoing
TEW-7197 MedPacto®	TGFβRI	NCT02160106	Phase I study in progress in solid tumours
PF-03446962 Pfizer®	TGFβRI	NCT00557856 NCT01337050 NCT01911273 NCT01486368 NCT01620970 NCT02116894	Phase I study in CRC completed Phase II study in mesothelioma and urothelial cancer completed Phase II study in HCC terminated Phase I study in combination with regorafenib in CRC completed
IMC-TR1 (LY3022859) Eli Lilly®	TGFβRII	NCT01646203	Phase I study in solid tumours completed

RCC: renal cell carcinoma; CRC: colorectal carcinoma; NSCLC: non-small cell lung carcinoma; HCC: hepatocellular carcinoma

Table 1.1 Drugs targeting TGFβ signalling and their current status in clinical drug trials

level using monoclonal antibodies or peptides; and 3) TGFβ receptor kinase inhibitors, which function to prevent signal transduction. A list of current drugs in development can be seen in Table 1.1 adapted from Neuzillet et al (Neuzillet et al. 2015). Although the prospect of TGFβ inhibition seems attractive, considering the role of TGFβ tumour progression and the results of early drug trials appear promising, it is crucial to consider the opposing effect of TGFβ signalling in different tumour types to ensure patients are correctly stratified for treatment. Perhaps more importantly current research should focus on identifying molecular markers of TGFβ tumour promoting activity, which is crucial for appropriate patient stratification and treatment monitoring.

1.6 NEDD9

As mentioned above, Bruna and colleagues recently identified the integrin scaffolding protein Neural Precursor Cell Expressed, Developmentally Down-Regulated 9 (NEDD9) is a key component in regulating stem cell dynamics in response to TGFβ/SMAD signalling in the Claudin-low breast cancer subset. It is hypothesised that in this breast cancer subtype, NEDD9 acts as a signalling hub to coordinate signalling pathways in response to TGFβ pathway activation.

1.6.1 Discovery and relevance

NEDD9, also known as Enhancer of Filamentation 1 (HEF1) and CASL, is a non-catalytic member of the CAS (Crk associated substrate) family of scaffolding proteins. First identified in 1992, NEDD9 was found to be expressed in the early embryonic mouse brain but not in the adult, and thought to play crucial roles in embryonic development and differentiation of the central nervous system (Kumar, Tomooka, and Noda 1992). Following this discovery, the complete gene sequence and an elementary protein function analysis were presented by two independent research groups in 1996. Law et al used a yeast system to screen a human cDNA library of oncogenes and antioncogenes for proteins that enhanced pseudohyphal budding (Law et al. 1996). They identified HEF1, similar to the recently described docking protein p130 Cas (Sakai et al. 1994) as the cDNA producing the most striking phenotype, and hypothesised that HEF1 may be an important mediator of extracellular signalling and cytoskeletal regulation. Mingenishi et al demonstrated that their previously discovered pp105 protein was a novel p130 Cas family member important for integrin signalling in T lymphocytes, and designated their protein Cas-L (Crk-associated substrate-related protein, Lymphocyte type) (Minegishi et al. 1996).

In vertebrates, four CAS family members have now been identified: NEDD9/HEF1/Cas-L (Law et al. 1996), BCAR1/p130Cas (Sakai et al. 1994), EFS/Sin (Ishino et al. 1995) and CASS4/HEP (Singh et al. 2008) although NEDD9 and BCAR1 remain the most extensively investigated members. NEDD9 is implicated in the regulation of diverse cellular effects both physiologically and pathophysiologically, via integrin-dependent cellular mechanisms of adhesion and migration, invasion, apoptosis, cell cycle, and survival through the coordination of focal adhesion kinase (FAK) and Src signalling cascades (Natarajan et al. 2006; O'Neill and Golemis 2001; Tikhmyanova and Golemis 2011).

The overexpression of CAS proteins has been strongly linked to poor prognosis and metastasis in addition to conferring resistance to first-line chemotherapeutics in multiple tumour types, including breast cancer (Dorssers et al. 2004; Brinkman et al. 2000). Additionally, *in vitro* NEDD9 overexpression has been demonstrated to induce invasive behaviour in breast cancer cell lines (Fashena et al. 2002). Recent meta-analyses have further highlighted the association between NEDD9 overexpression and poor overall survival in patients with multiple solid tumour types, indicating that NEDD9 expression may function as an effective predictive cancer biomarker (Gu et al. 2019; Shagisultanova et al. 2015).

1.6.2 NEDD9 expression and regulation

The human NEDD9 gene is located on the antisense strand of the short arm of chromosome 6. The NEDD9 promoter has two transcriptional start sites and there are four known protein-coding transcript variants annotated in the image from the NIH genome browser (figure **1.2a**) (www.ncbi.nlm.nih.gov): Transcript variants one (NM_006403) and three (NM_001142393) both encode full-length proteins but with different N-termini (MKYK and MWTR, respectively), translation of variant NM_001142393 is initiated upstream of NM_006403, both transcripts contain seven exons. Transcript variant four (NM_001271033) lacks an alternate in-frame exon when compared to transcript variant one and therefore encodes a shorter isoform of six exons. Transcript variant two lacks several exons and has an alternate 3' terminal exon compared to variant one.

NEDD9 has consistently been demonstrated by western blot to migrate as a doublet which is substantially larger than its predicted size of 93kDa. At least four NEDD9 protein forms have been described in the literature, designated $p115^{HEF1}$,$p105^{HEF1}$, $p65^{HEF1}$ and $p55^{HEF1}$, with the larger 115kDa protein identified as a heavily phosphorylated form of the full-length 105kDa protein (Law et al. 1998). The smaller $p65^{HEF1}$ and $p55^{HEF1}$ protein forms are truncated versions of the full-length protein with $p55^{HEF1}$ generated by caspase-cleavage of the full-length protein. The p55 species contains NEDD9 amino acid sequences mapping to the amino-terminal of the protein (SH3 domain and SH2-binding site sequences) and the p65

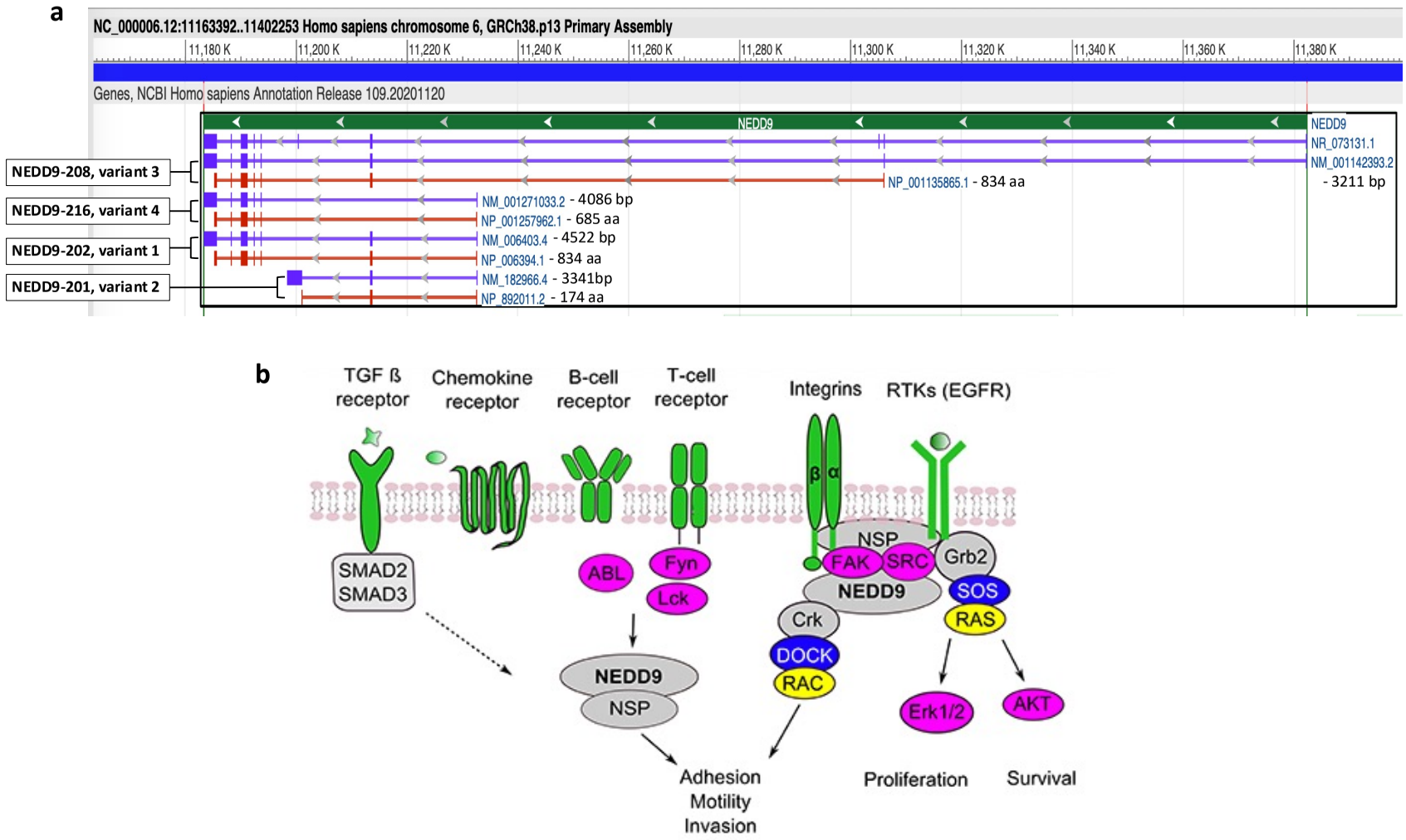


Figure 1.2. NEDD9 expression and regulation of cellular pathways/effects
(**a**) NEDD9 isoform expression from the NIH genome browser (www.ncbi.nlm.nih.gov) GRCh38.p13 reference genome. (**b**) An overview of the pathways and cellular processes regulated by NEDD9 (Gaponova Anna, CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons)

species contains carboxy-terminal NEDD9 sequences, however, $p65^{HEF1}$ is generally poorly detected and its regulation is not fully understood (Law et al. 1998). NEDD9 is a non-catalytic scaffolding protein that acts as a docking station for multiple transcriptional pathways and, therefore, regulates a host of cellular effects. An overview of the pathways and cellular effects mediated by NEDD9 can be seen in figure **1.2b**. The protein comprises four structural domains; a highly conserved N-terminal SH3 domain, an unstructured substrate domain (SD), a serine-rich region (SR), and a focal adhesion targeting domain (FAT). Both the SR and FAT domains are predicted to fold into four-helix bundles (Briknarová et al. 2005; Mace et al. 2011). The SH3 domain mediates the binding of NEDD9 to interacting proteins containing multiple proline residues. The SD contains multiple YxxP motifs, which are phosphorylated by members of the Src family of kinases; once phosphorylated, these motifs act as binding sites for proteins containing SH2 regulatory module domains thus coordinating multiple signalling networks. The FAT domain is responsible for the localisation of focal adhesion kinase (or protein tyrosine kinase 2 PTK2) to focal adhesion structures to mediate cell migration processes.

In embryonic brain development, regulation of the NEDD9 promoter is known to occur by all-trans retinoic acid (ATRA) (Aquino et al. 2009), however, in cancer, several different gene regulation mechanisms have been proposed in numerous tumour types. These mechanisms include; Wnt signalling in colorectal cancer (Y. Li et al. 2011), Forkhead transcription factors in hepatocellular carcinoma (L. Xia et al. 2013), hypoxia in colorectal cancer (S.-H. Kim et al. 2010), the estrogen receptor alpha corepressor SAFB1 in breast cancer (Hammerich-Hille et al. 2010), and TGFβ in liver, prostate and breast cancer (Zheng and McKeown-Longo 2002). NEDD9 has also been shown to promote a stem/progenitor cell phenotype in many cell lineages, and significant binding of the pro-stem cell factor SOX2 is observed in the NEDD9 promoter (Aquino et al. 2009; Vogel et al. 2010; Boyer et al. 2005).

At the protein level, the NEDD9 mechanism of action is highly regulated by extensive phosphorylation in response to cell attachment, growth factors and hormones. This phosphorylation influences the scaffolding activity of NEDD9 along with cellular localisation and proteasomal degradation. Tyrosine phosphorylation of CAS proteins in response to cell attachment occurs through the integrin signalling pathway involving FAK and SRC (RAFTK and SRC in NEDD9), and also occurs in the presence of growth factors and hormones including VEGF and IGFR. Alternatively, TGFβ can induce serine/threonine phosphorylation, independent of focal adhesion and an intact actin cytoskeleton (Zheng and McKeown-Longo 2002). The phosphorylation state of NEDD9 is cell cycle synchronised, the more phosphorylated form p115 kDa (serine/threonine) accumulates in the G2/M phase and is more susceptible to proteolysis (Zheng and McKeown-Longo 2002). The cellular location of NEDD9 shifts throughout the

cell cycle in response to changes in phosphorylation state and at interphase the majority of NEDD9 is localised to focal adhesions. However, upon entry into mitosis, in response to phosphorylation NEDD9 migrates along the mitotic spindle eventually localising at the midbody at cell division (Pugacheva and Golemis 2005). Post-transcriptional regulation of the NEDD9 transcript occurs by miR-145, which binds the untranslated 3' region. MiR-145 is a known repressor of pluripotency in embryonic stem cells and an oncogene in several cancer types (Lu et al. 2014; Guo et al. 2013). Speranza et al successfully demonstrated the downregulation of miR-145 in glioblastoma resulting in increased NEDD9 mediated invasion of glioblastoma cells in *in vitro* assays (Speranza et al. 2012).

TGFβ is known to phosphorylate NEDD9 and, in some conditions, it has been shown to trigger NEDD9 proteasomal degradation by ubiquitination (X. Liu et al. 2000; Zheng and McKeown-Longo 2002). Wendt and colleagues demonstrated that P130Cas/BCAR1, another member of the Cas protein family, was able to block TGFβ induced Smad 2/3 activation and subsequently promote TGFβ induced activation of p38 MAPK when BCAR1 was overexpressed in mammary tumour cells. As a confirmation of this effect, the authors demonstrated that inhibition of BCAR1 restored TGFβ/Smad induced tumour inhibition and blocked TGFβ/MAPK promotion of tumour progression (M. K. Wendt, Smith, and Schiemann 2009).

When taken into consideration alongside the evidence that NEDD9$^{-/-}$ MMTV mouse tumour models demonstrate increased tumour latency and reduced tumour number (Izumchenko et al. 2009), and our findings that NEDD9 is a key regulator of TGFβ induced breast cancer stem cell self-renewal (Bruna et al. 2012), BCAR1 and NEDD9 are emerging as integral determinants in both the tumour suppressive and promoting effects of TGFβ in cancer signalling. However, what still remains unclear is the molecular mechanism by which NEDD9 mediates the signalling pathways responsible for these effects in breast cancer stem cells from different breast cancer subtypes.

1.7 Hypothesis and aims

There exists extensive data on the correlation between NEDD9 expression and tumour aggressiveness and metastasis in multiple cancer types, and the role of TGFβ signalling in tumour progression. Recent data from Bruna and colleagues identified NEDD9 as a key modulator of TGFβ oncogenic effects in the Claudin-low subset of breast cancer cell lines, and the gene expression signature from this analysis correlates with poor prognosis in a breast cancer clinical cohort (Bruna et al. 2012). However, what remains unclear is the mechanism by which NEDD9 mediates oncogenic TGFβ signalling in the Claudin-low subset of breast cancer when compared with other breast cancer subtypes. Due to my

involvement in the project which led to the publication of the above findings (Bruna et al. 2012), this thesis represents a logical continuation of that research under the supervision of Dr Alejandra Bruna.

The main aim of this thesis is to investigate the hypothesis that breast cancer prognosis is not only dependent on tumour subtype and that it is more complex than merely expression levels of NEDD9, and TGFβ signalling. This thesis addresses the above hypothesis by exploring the following aims:

In the first chapter, I aim to gain an understanding of NEDD9 gene mutations and expression across multiple cancer types before focusing on breast cancer. I will then characterise the NEDD9 gene and protein expression in a panel of breast cancer cell lines in response to the activation of TGFβ signalling. I will also use RNA sequencing methods to investigate gene expression under the perturbation of NEDD9 protein expression to compare TGFβ responsive signalling networks which are dependent upon NEDD9. Finally, I will use RNA sequencing data to investigate which of the known NEDD9 isoforms are predominantly expressed in response to TGFβ treatment, and compare these data with those from breast cancer clinical cohorts.

In the second chapter, I will use a variety of techniques to identify NEDD9 interacting proteins to try and predict signalling networks coordinated by NEDD9 contributing to oncogenic TGFβ signalling. I will start with simple co-immunoprecipitation methods and then advance to complex quantitative analyses to confirm my findings.

The third chapter will focus on investigating the mechanism by which PKM2 (a key NEDD9 interactor in oncogenic TGFβ signalling) influences tumour initiating cell dynamics. I will explore the known roles of PKM2 in metabolism and transcriptional activation and will investigate whether the results from the RNA sequencing and phospho-proteome analyses indicate activation of metabolic programs.

Finally, I will investigate the translational application for my findings in patient treatment. I will investigate the expression of the oncogenic TGFβ/NEDD9 signature in a large clinical cohort and our comprehensive patient-derived tumour xenograft (PDTX) biobank, and determine how this relates to response to treatment with therapeutic agents.

If successful in addressing the aims of my study I hope to further the understanding of the complex regulation of breast cancer stem cell dynamics in the Claudin-low subtype. I also hope that this work may

help to identify biomarkers of oncogenic TGFβ signalling and point towards therapeutic agents which would be advantageous in the treatment in a breast cancer subtype-specific manner.

2. NEDD9 gene and protein expression in breast cancer

2.1 Preface

As discussed, my previous work, which was performed under the supervision of Dr Alejandra Bruna before the start of my PhD, identified NEDD9 as a key component in the regulation of oncogenic TGFβ signalling in the Claudin-low subset of breast cancer (Bruna et al. 2012). In this chapter, I aimed to build upon that knowledge by first characterising NEDD9 expression in a panel of breast cancer cell lines, followed by a more specific analysis of NEDD9 expression in response to TGFβ signalling in the Claudin-low subtype.

2.2 NEDD9 gene expression in cancer

To first gain a current understanding of global NEDD9 RNA expression levels in a range of cancer types, I queried the TCGA PanCancer Atlas at cBioPortal (www.cbioportal.org) (Cerami et al. 2012; Gao et al. 2013), a curated database of 32 clinical cancer studies. Figure **2.1a** details RSEM normalised (B. Li and Dewey 2011) relative expression levels of NEDD9 and detected mutations across a wide range of cancer types. In studies included in the TCGA PanCancer atlas, NEDD9 expression was increased as indicated by the fold change values on the y axis. Additionally, NEDD9 gene mutations were frequent, as indicated by coloured circles. To investigate this observation in more detail I then queried NEDD9 expression and mutational profile in the METABRIC breast cancer cohort (Curtis et al. 2012; Pereira et al. 2016). Figure **2.1b** details the NEDD9 mutational profile of 2509 breast cancers from the METABRIC clinical cohort, including details of patient ER and HER2 status, breast cancer intrinsic subtype and Integrative cluster subtype. The zoomed insert panel (figure **2.1b**) illustrates in more detail the annotation of the 6% of patients with NEDD9 mutations, all in the form of copy number amplifications and/or high mRNA expression. There was no correlation between NEDD9 copy number amplification and ER status, of the 45 patients with NEDD9 copy number amplification 23 are ER-positive and 22 are ER-negative. However, a definite correlation exists with regards to HER2 expression, with 43 out of 45 patients having HER2 negative disease. There was no overrepresentation of any breast cancer subtype in the NEDD9 copy number amplification group, with both the PAM50 and Integrative cluster subtypes evenly distributed, and Kaplan–Meier estimates (figure **2.1c**) on patients with NEDD9 copy number

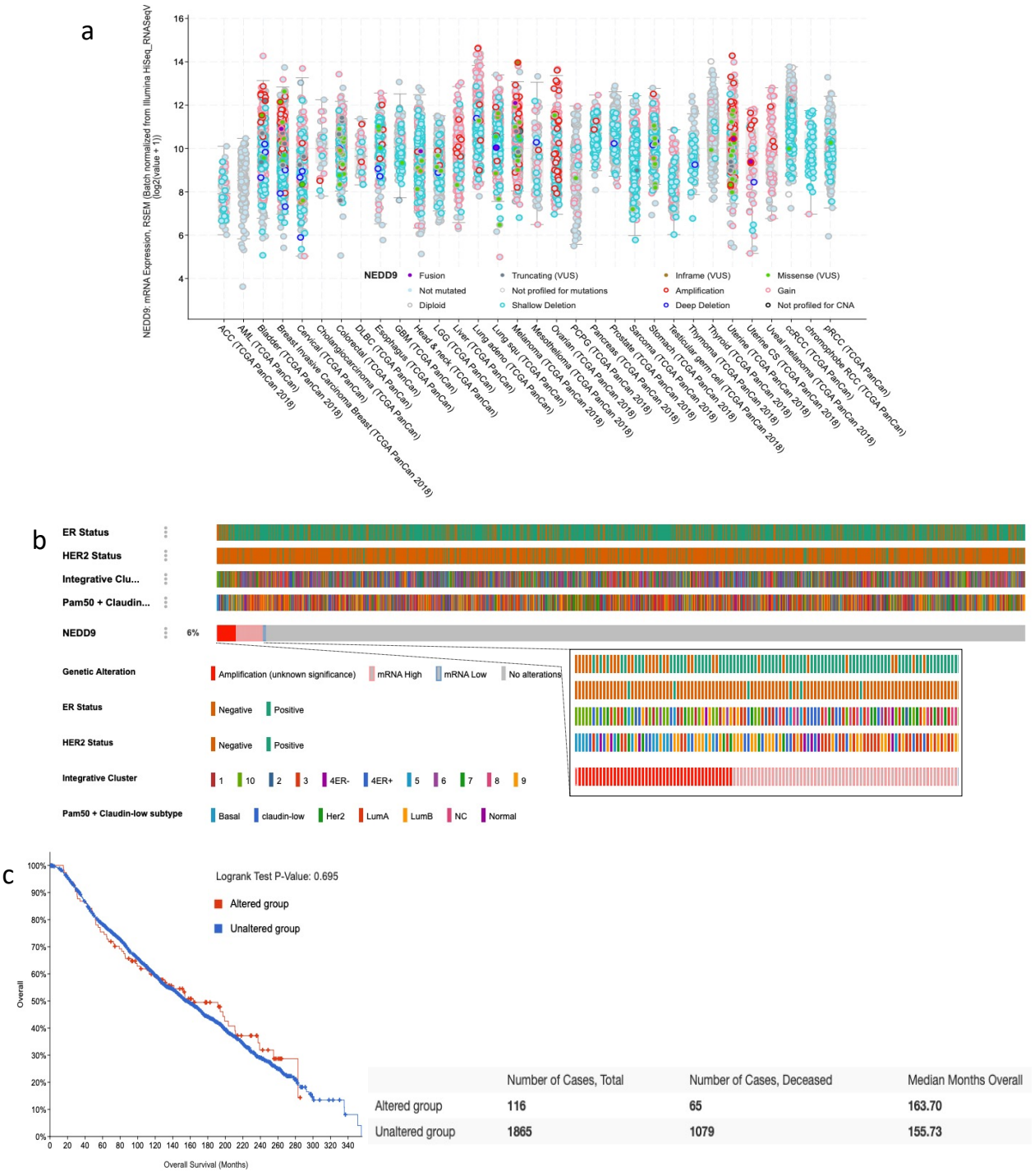


Figure 2.1. NEDD9 expression in cancer

(**a**) NEDD9 RSEM normalised RNA-Seq expression (Log2 transformed) from the TCGA PanCancer Atlas (cbioportal.com) across multiple cancer types, including mutations and types where profiled. **(b)** NEDD9 mutational profile of 2509 breast cancers from the METABRIC clinical cohort, including intrinsic subtype and Integrative cluster annotation. Zoomed insert illustrates annotation details for the 6% of patients with a NEDD9 amplification and/or high mRNA expression. . **(c)** Kaplan–Meier overall survival estimates for patients profiled for NEDD9 alterations (red) versus the unaltered group (blue).

amplifications showed no effect on overall survival, with a median survival of 163.7 months in the altered group compared with 155.7 months in the unaltered group. These data confirm that NEDD9 expression and mutational status do not have a major effect on patient outcome and strengthen our hypothesis that TGFβ oncogenic effects mediated by NEDD9 are more complex than being regulated by NEDD9 expression levels alone. This observation prompted further investigation into the diversity of NEDD9 expression levels in a panel of breast cancer cell lines representative of the heterogeneity of breast cancer.

2.3 NEDD9 protein diversity, and response to TGFβ in a panel of cell lines

NEDD9 is an integrin signalling adaptor protein involved in the regulation of cell motility and adhesion (Zhong et al. 2012; Kondo et al. 2012). Many groups have observed variation in NEDD9 protein bands identified by western blot, both within different cell lines, and between cell lines grown in suspension versus adherent culture. NEDD9 has consistently been demonstrated to migrate as a doublet by western blot substantially higher than its predicted size of 93kDa. Two full-length protein forms designated $p105^{HEF1}$ and $p115^{HEF1}$ have been described, with the larger 115kDa protein identified as a heavily phosphorylated version of the 105kDa protein (Law et al. 1998). These bands are thought to represent the multiple post-translational modifications of the NEDD9 protein known to arise in response to several different signalling pathways (Zhao et al. 2019; Zheng and McKeown-Longo 2002, 2006). Our previous work identified NEDD9 as a hub integrating signalling pathways in the modulation of breast tumour initiating cell expansion in response to TGFβ (Bruna et al. 2012). We, therefore, hypothesised that NEDD9 post-translational modification may be important in coordinating signalling pathways in response to TGFβ treatment, and the downstream regulation of gene expression programs of tumour initiating cell self-renewal.

To investigate the heterogeneity of NEDD9 protein expression across breast cancer subtypes, I first selected a panel of breast cancer cell lines representing the heterogeneity of breast cancer observed in the patient population, as illustrated below.

CellLine	Claudin-low score	LogFold	Classification	TP53 status	Her2 status	ER status	PR status	CD49f/EpCAM
HCC1954	-0.41227	-4.22942	Basal A	MUT	+	-	-	0.59705
SKBR7	0.19866	-1.80834	Normal like	WT	-	-	-	0.97883
MCF7	-0.42141	-1.03579	Luminal	MUT	-	+	+	0.49400
MDA-MB-231	0.48312	0.96505	Claudinlow	MUT	-	-	-	1.09372
BT549	0.65611	0.63117	Claudinlow	WT	-	-	-	1.03261
SUM159	0.59517	2.65371	Claudinlow	MUT	-	-	-	1.09572

Table 2.1. Classification and characteristics of the breast cancer cell line panel used to investigate NEDD9 protein diversity by western blot

The panel included cell lines from across a range of PAM50 subtypes, TP53 mutational status, and cell surface marker expression as detailed above. The Claudin-low score is calculated as the difference between the Spearman correlation value of expression data for each cell line with the Claudin-low centroid, and the Spearman correlation with the "others" centroid as defined by Prat and colleagues (Prat et al. 2010). Cell lines with a higher Claudin-low score correlate more closely with the Claudin-low centroid and are therefore classified as Claudin-low. The LogFold is the log2 fold change in the number of second-generation mammospheres in TGFβ-treated compared to untreated cultures: positive and negative values indicate where TGFβ increases or decreases BTIC self-renewal activity, respectively. The CD49f/EpCAM ratio is a measure of tumour aggressiveness, a higher ratio correlates with shorter disease-free survival (Ye et al. 2015).

For the western blot analysis, cells were grown in suspension mammosphere culture for seven days to enrich for breast tumour initiating cells (BTIC), and left untreated, or treated with TGFβ 100 nM. Figure **2.2a** demonstrates the variation in NEDD9 protein bands identified by western blot, in the Claudin-low subtype (labelled in red) compared to all other subtypes (labelled in blue), bands corresponding to $p115^{HEF1}$ and $p105^{HEF1}$ are indicated by arrows. In cell lines belonging to the Claudin-low category (MDA-MB-231, BT549 and SUM159), TGFβ treatment increased the intensity of the $p115^{HEF1}$ band when compared to the no-treatment control, and decreased the intensity of the highest molecular weight band above $p115^{HEF1}$. The significance of this uppermost band will be discussed later. This pattern was not as pronounced in cell lines belonging to other breast cancer subtypes. This pattern of NEDD9 protein expression was consistently observed over numerous experiments. Increased levels of phospho-Smad2 in the TGFβ treatment conditions indicates activation of the TGFβ pathway, and consistent levels of Smad2/3 demonstrate that the TGFβ pathway is intact in these cell lines. β-Actin was used as a loading control. Based on these observations, two cell lines were chosen which were representative of the ability of TGFβ to induce BTIC self-renewal, or not. These cell lines, respectively MDA-MB-231 (ATCC® HTB-26™) and HCC1954 (ATCC® CRL-2338™) were used for all further experiments.

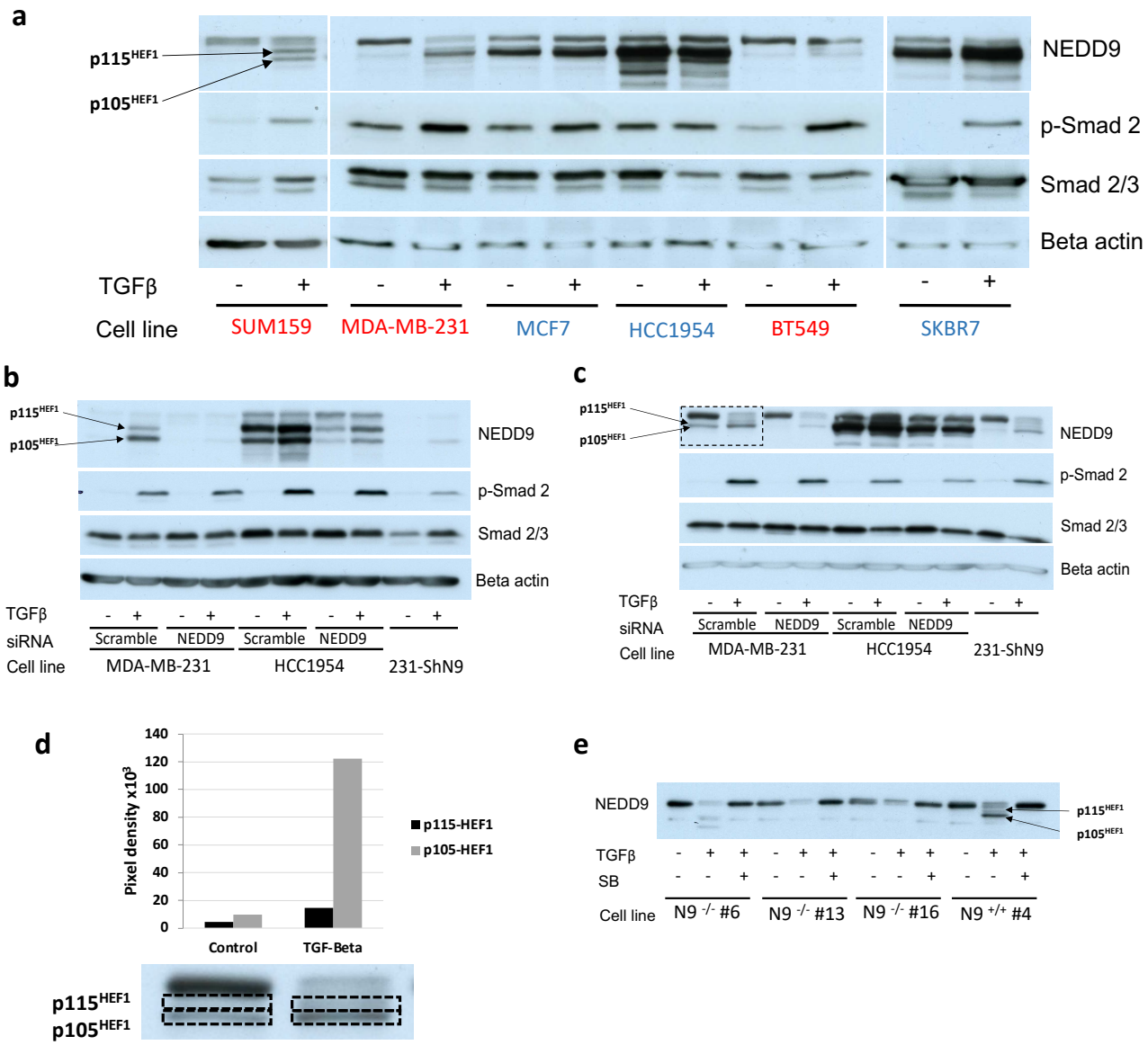


Figure 2.2. NEDD9 expression and post-translational modification in response to TGFβ treatment
(**a**) Western blot of whole cell protein lysates from mammosphere cultures of cell lines representative of the Claudin-low subtype (red) versus all others (blue) following seven day treatment +/- 100 nM TGFβ. Membranes were probed for levels of NEDD9, phospho-Smad2 and Smad2/3 to confirm TGFβ pathway activation, and β-Actin as a loading control. Western blot from whole cell protein lysates of cell lines MDA-MB-231 (type A) and HCC1954 (type B) in attached culture (**b**) and suspension mammosphere culture (**c**). Cultures were treated for seven days +/- 100 nM TGFβ. NEDD9 expression was perturbed by use of a commercial SiRNA directed towards NEDD9 or a control scramble SiRNA, or by a stable knockdown of NEDD9 in MDA-MB-231 expressing a short-hairpin generated against the NEDD9 coding sequence (231-ShN9). Membranes were probed as above. (**d**) NEDD9 band quantification of phospho-forms p105^{HEF} and p115^{HEF} from the MDA-MB-231 scramble SiRNA condition in figure **2.2c**. Units represent pixel density analysis using Image J. (**e**) Western blot of whole cell protein lysates from mammosphere cultures of MDA-MB-231 NEDD9 Crispr knock-out cell line clones (#6, #13 and #16) and electroporation control (#4). Cultures were treated for seven days +/- 100 nM TGFβ or 100 nM TGFβ plus 10 μM TGFβ inhibitor SB431542 10 μM (SB). Membrane was probed for NEDD9 protein expression. The position of the NEDD9 phospho-forms p105^{HEF} and p115^{HEF} is located as indicated by arrows. Densitometry analysis was performed on a single experiment, but western blot figures are representative of at least three biological replicates.

Figures **2.2b** and **2.2c** represent experiments performed to identify NEDD9 specific bands identified by western blot in both adherent (figure **2.2b**) and suspension mammosphere culture (figure **2.2c**). It is evident that in enriching for tumour initiating cells by culture in suspension, the protein expression of NEDD9 is altered. Here, following TGFβ treatment an increase in the expression of $p105^{HEF1}$ was observed, this represents increased NEDD9 mRNA expression and translation of new protein, and post-translational modification of $p105^{HEF1}$ to generate $p115^{HEF1}$, processes which have both been ascribed in response to TGFβ signalling. However, when the expression of NEDD9 was perturbed partially either by siRNA or shRNA the levels of both $p115^{HEF1}$ and $p105^{HEF1}$ were decreased, although less so in the shNEDD9 cell line. To quantify the observed changes in expression of $p105^{HEF1}$ and subsequent post-translational modification of $p105^{HEF1}$ to $p115^{HEF1}$ I performed pixel density analysis using Image J (Schneider, Rasband, and Eliceiri 2012) on the protein bands from the MDA-MB-231 Scramble siRNA condition, in the presence and absence of TGFβ treatment (dashed box). Figure **2.2d** quantifies the previously observed increase in the expression of both phospho-forms in response to TGFβ treatment. To further validate NEDD9 protein expression, western blots were performed on whole-cell lysates of MDA-MB-231 suspension cultures from clones where NEDD9 was completely knocked out by CRISPR-Cas9 genome editing, the knock-out was confirmed by sequencing (figure **2.2e**). These cultures were treated as previously with, or without 100 nM TGFβ, but with the addition of 10 μM TGFβ inhibitor SB431542 (SB) to confirm the changes observed in NEDD9 expression are due to TGFβ signalling. Here, the $p115^{HEF1}$ and $p105^{HEF1}$ bands which were seen in the electroporation control ($N9^{+/+}$#4) were not present in any of the three CRISPR knock-out clones ($N9^{-/-}$#6, $N9^{-/-}$#13, or $N9^{-/-}$#16). Interestingly, however, as observed in the previous suspension culture analyses (figures **2.2a and 2.2c**) the top band, larger than $p115^{HEF1}$ was still present, even in the NEDD9 knock-out clones. This suggests that this is a non-specific band, but is detecting a protein still modulated by TGFβ signalling.

As mentioned in the introduction, NEDD9 phosphorylation by TGFβ is thought to target the protein for proteolysis. To investigate this, seven-day suspension cultures of MDA-MB-231 and HCC1954 with, or without 100 nM TGFβ treatment, were treated for four hours with the proteasome inhibitor MG132 before collection of whole-cell lysates. Figure **2.3a** demonstrates that a modest increase in the level of $p115^{HEF1}$ was observed in the TGFβ plus MG132 treatment condition when compared to the TGFβ condition alone (dashed box), and these changes were quantified, as previously using Image J (figure **2.3b**). No corresponding changes in NEDD9 protein expression were observed in the non-Claudin-low cell line HCC1954. Together these results indicate that NEDD9 protein expression and post-translational modification in response to TGFβ signalling is different in the Claudin-low subset of breast tumour cell

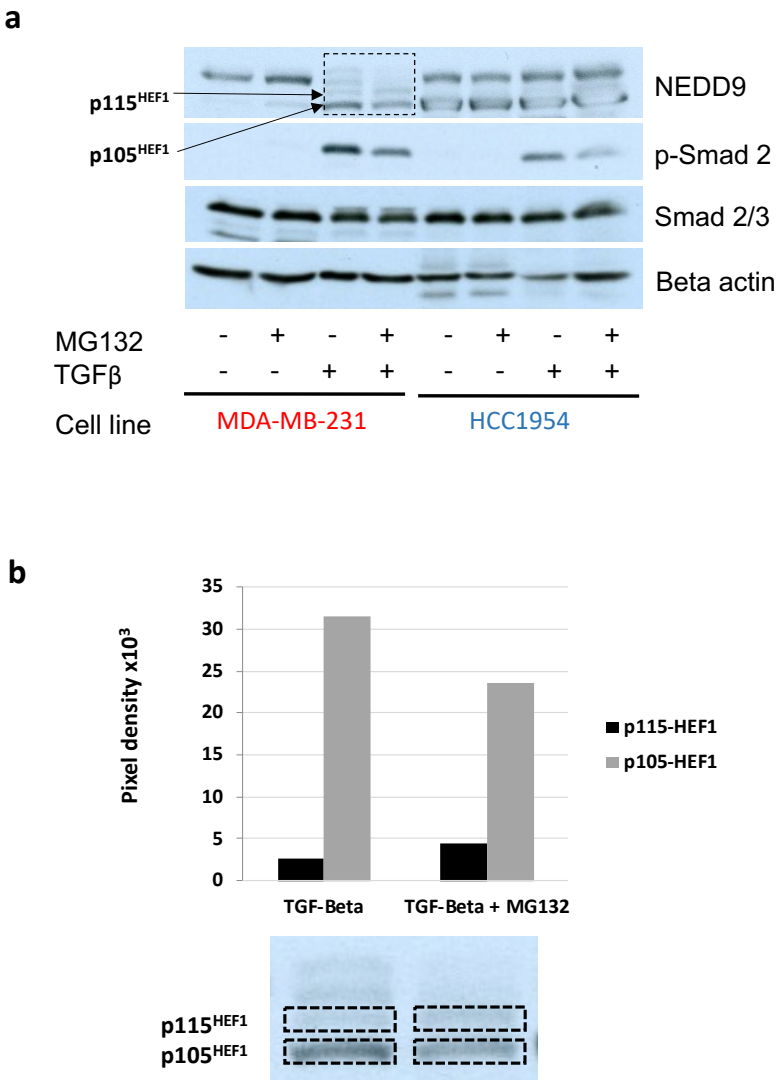


Figure 2.3. NEDD9 expression and post-translational modification in response to proteasomal inhibition

(**a**) Western blot from whole cell protein lysates of cell lines MDA-MB-231 (type A) and HCC1954 (type B) following seven day treatment +/- 100 nM TGFβ, and +/- 4 hours 10 μM MG132 treatment at day 7. (**b**) NEDD9 band quantification of phospho-forms $p105^{HEF}$ and $p115^{HEF}$ from the MDA-MB-231 + 100 nM TGFβ condition in figure **2.3a**. Units represent pixel density analysis using Image J. Densitometry analysis was performed on a single experiment, but western blot figures are representative of at least three biological replicates.

lines when compared to other breast cancer cell line subtypes, even in the presence of an intact TGFβ signalling pathway. These results also indicate that in the Claudin-low cell line MDA-MB-231, TGFβ treatment targets NEDD9 for proteasomal degradation. These results are consistent with the findings of our lab and others (Bruna et al. 2012; Zheng and McKeown-Longo 2002)

2.4 Validation of NEDD9 protein bands

To further investigate our initial observation that the NEDD9 uppermost band seen by western blot may be a non-specific artefact, I went on to perform an in-depth proteomics analysis. NEDD9 was immunoprecipitated (IP) from whole-cell extracts of MDA-MB-231 mammosphere cultures which had been treated for seven days with, or without 100 nM TGFβ. To eliminate background antibody interference, the NEDD9 antibody (Abcam) used for the immunoprecipitation was covalently linked to Dynabeads® magnetic beads (Life technologies). Following the IP, proteins of interest were eluted from the magnetic beads reducing background interference by antibody heavy and light chains in the mass spectrometry analysis.

The resulting immunoprecipitates were run on an 8% polyacrylamide gel and stained with Coomassie blue alongside a prestained protein ladder. Bands were excised from the gel at locations corresponding with numbers 1-6, in both the control and TGFβ treated condition (figure **2.4b**). The gel slices were then processed and analysed by liquid chromatography-mass spectrometry (LCMS) in the CRUK-CI proteomics core facility following the workflow outlined (figure **2.4a**).

Figure **2.4b** (left panel) is a confirmation western blot using 5% of the IP product; it shows that the IP method resulted in an enrichment of protein at around 105 kDa to 115 kDa when compared to the input (right panel). To confirm the presence of NEDD9 in the remaining IP product, I excised bands as indicated. Figure **2.4c** represents the Protein Identification Probability (PIP) value for each band being identified as NEDD9 as generated by Scaffold QPlus (Proteome software); this is a calculated measure of the percentage probability of protein identification being correct and is calculated based on several parameters, including the number of peptides identified and the protein coverage. In the IP NEDD9 annotated figure (figure **2.4c**) a pattern of three bands analogous to those seen in western blot was visible in both the control and TGFβ treatment condition (highlighted by the black box). These bands correspond to the positions marked 2, 3 and 4 in **2.4b**, and fall into the size range expected for $p105^{HEF1}$, $p115^{HEF1}$ and the accompanying higher mass band also seen by western blot. When calculated from the LCMS data, in the TGFβ condition bands 4 and 3 (corresponding to $p105^{HEF1}$ and $p115^{HEF1}$ respectively) demonstrated a NEDD9 PIP value of 100%. Whereas the upper band had a NEDD9 PIP of 0%, confirming the results from the knock-down experiments, that this is a non-specific band detected by the antibody.

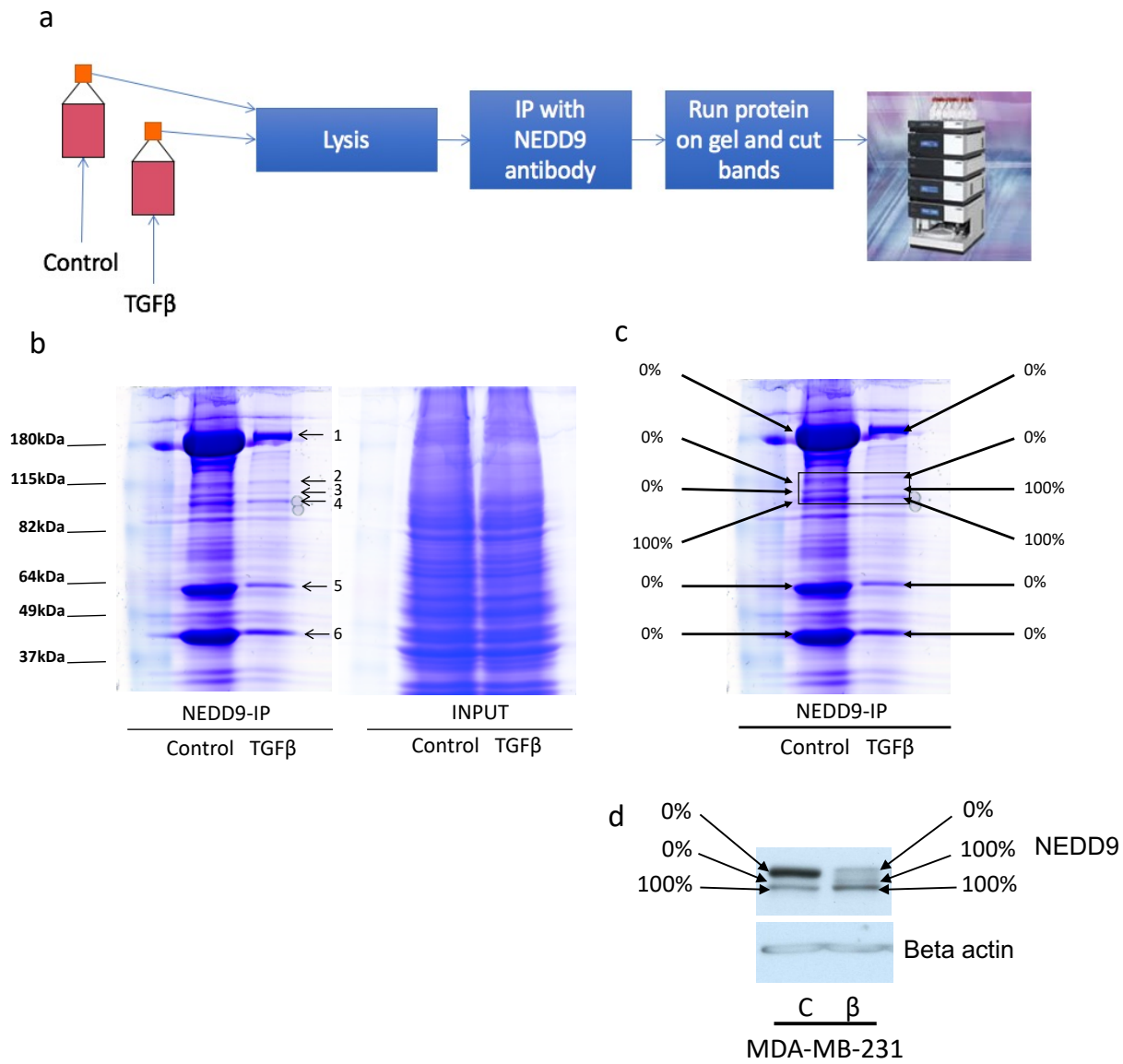


Figure 2.4. NEDD9 protein band identification using IP-LCMS
(**a**) Schematic demonstrating the workflow for IP-LCMS identification of NEDD9 protein bands. Large scale seven-day MDA-MB-231 suspension cultures +/- 100nM TGFβ were prepared. Following cell lysis NEDD9 protein was pulled down using an antibody against NEDD9 and the resulting protein run on a polyacrylamide gel and bands cut for LCMS analysis. (**b**) Coomassie stained polyacrylamide gel of the immunoprecipitate from the NEDD9 pull-down showing the locations (numbered arrows 1-6) of bands excised for LCMS analysis. The right hand panel is a Coomassie stained polyacrylamide gel of protein extracted from the whole cell lysate (input) as a control. (**c**) An illustration of the NEDD9 Protein identification probability (PIP) for each of the corresponding numbered bands collected in the control and TGFβ treatment conditions (**b**). (**d**) Representation of how the PIP values correspond to bands seen by western blot in a seven day mammosphere culture of MDA-MB-231 +/- 100nM TGFβ. The western blot figure is representative of at least three biological replicates.

When we turned to the PIP values in the control condition, only band 4 ($p105^{HEF1}$) was identified as NEDD9 with 100% confidence. Figure **2.4d** represents the location of the analysed bands, and PIP percentage, in a representative western blot for MDA-MB-231 with or without TGFβ treatment.
The uppermost band (2) was a non-specific antibody artefact targeted by neither siRNA, short-hairpin nor by CRISPR-Cas9 genome editing, but interestingly it was differentially regulated by TGFβ. This indicates that it is probably a NEDD9 partner protein, and further analysis of the LCMS data confirmed that Breast cancer anti-oestrogen resistance protein 1 (BCAR1), a known NEDD9 interacting protein, was identified with 100% probability in bands 2 and 3 in both conditions. What is not clear, however, is why this band decreases in intensity in response to an increase in NEDD9 expression and vice versa.

The conclusion from this experiment is consistent with our observations by western blot, that of the three bands identified in the region of the NEDD9 full-length protein, only the lowest band, corresponding to $p105^{HEF1}$ was identified as NEDD9 with a 100% probability in both conditions. The band corresponding to $p115^{HEF1}$ was only identified as NEDD9 in the TGFβ condition, indicating that this is the phosphorylated form as identified by Law et al (Law et al. 1998) and that this phosphorylation is induced by TGFβ signalling.

2.5 Characterisation of NEDD9 post-translational modifications

Next, I proceeded to investigate NEDD9 phosphorylation to predict substrate and kinase binding motifs. The presence of specific motifs may indicate the role of TGFβ activated, NEDD9 dependent pathways. To this aim, I immunoprecipitated NEDD9 from whole-cell extracts of MDA-MB-231 mammosphere cultures treated for seven days with vehicle control (DMSO), 100 nM TGFβ, or 100 nM TGFβ plus 10 μM TGFβ inhibitor SB431542. As previously, the IP was performed using Dynabead-linked NEDD9 antibody, the immunoprecipitated protein was then eluted from the beads and phosphopeptides enriched using titanium dioxide [Methods 3]. The digestion, enrichment and LCMS analysis were performed by the CRUK-CI core proteomics team following the outlined workflow (figure **2.5a**). Figures **2.5b-d** represent the identified phospho-peptides in each of the treatment conditions; vehicle control (figure **2.5b**), TGFβ (figure **2.5c**) and TGFβ plus TGFβ inhibitor SB431542 (figure **2.5d**). As previously observed by western blot, NEDD9 protein abundance was higher in the TGFβ condition and the LCMS analysis confirmed this, with a protein coverage of 43.53% when compared to control (24.22%) and TGFβ inhibitor plus TGFβ (17.39%). The colour code for individual amino acid modifications relates to the percentage probability that the detection was correct.

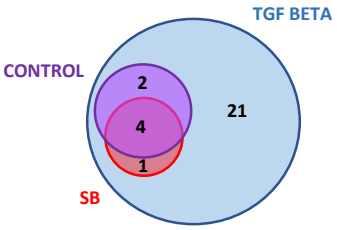
Figure 2.5. Characterisation of NEDD9 post translational modifications

(**a**) Schematic demonstrating the workflow for Ti02 enrichment LCMS identification of NEDD9 post-translational modifications. Large scale seven-day MDA-MB-231 suspension cultures +/- 100nM TGFβ and +/- 10μM TGFβ inhibitor SB431542 were prepared. Following cell lysis NEDD9 protein immunoprecipitated, followed by Ti02 phospho-peptide enrichment and LCMS analysis of phospho-peptides. (**b-d**) Primary structure maps of the NEDD9 protein showing LCMS coverage (green highlight) and phosphorylated amino acids identified. Maps correspond to the following treatment conditions; vehicle control (**b**), TGFβ (**c**) and TGFβ plus TGFβ inhibitor SB431542 (**d**). Colour codes relate to the percentage probability of the post-translational modification identification being correct.

Although a higher coverage was achieved in the TGFβ condition, the individual phosphorylation sites which were identified in the control and TGFβ inhibitor conditions, were also all identified in the TGFβ condition as illustrated in the Venn diagram (figure **2.6a**). However, not all individual modified amino acids identified in the TGFβ condition were covered in the control and TGFβ conditions so it was impossible to comment on whether they would have been modified if they had been included. Nevertheless, there remained a large number of amino acids which had coverage in all three conditions but were modified in the TGFβ condition alone.

Using this data I then returned to the literature on NEDD9 phosphorylations previously identified, and any published links with cancer by querying the PhosphoSitePlus® online analysis tool (www.phosphosite.org, as of April 2020). Figure **2.6b** describes the NEDD9 phosphorylated residues detected in each treatment condition and the protein domain in which the modifications are located. The number of references represents the number of papers referencing the particular modification in both low throughput and high throughput analyses, and the cancer types in which these modifications are significant in relation to disease. From the list of TGFβ specific phosphorylation sites, I selected the modifications where coverage was present in all three treatment conditions (highlighted with bold borders) and performed phospho-motif prediction using the PhosphoSitePlus® online analysis tool (figure **2.6c**) This tool predicts both phosphorylation-based substrate motifs (recognised by serine/threonine/tyrosine kinases or phosphatases) and phosphorylation-based binding motifs (motifs that once phosphorylated become targets for binding of phosphorylated residues). In the NEDD9 substrate domain, phosphorylation at Tyrosine 317 was identified with high confidence; this residue is within a motif associated with a substrate-binding domain for SHP1 (or PTPN6), a member of the protein tyrosine phosphatase family (PTP). SHP1 has known roles in numerous cellular mechanisms including oncogenic transformation. The same residue (Y317) is also identified with high probability in a motif for binding of Src Kinase substrates, in the TGFβ condition alone. The interaction of NEDD9 with Src, and the PI3K and MAPK pathways have been previously demonstrated (Izumchenko et al. 2009; Zheng and McKeown-Longo 2002). The NEDD9 substrate domain (SD) is known to contain multiple YxxP motifs which are phosphorylated by members of the Src family of kinases which, once phosphorylated, act as binding sites for proteins containing SH2 regulatory module domains thus coordinating multiple signalling networks. These data suggest an interaction between TGFβ signalling and activation of Src signalling independent of integrins, as is the usual method of activation. Izumchenko and colleagues demonstrated that NEDD9 is not only a target of FAK and Src kinases but is also involved in a feedback loop supporting their continuous activation (Izumchenko et al. 2009).

a



b

TGFβ	Control	SB	References	Cancer	Domain
Y92			337	Many inc. Breast	Substrate domain
Y166			578	Many inc. Breast	Substrate domain
Y168			6	Lung, Pancreatic, Leukaemia	Substrate domain
Y177			78	Many inc. Breast	Substrate domain
S182			22	Liver, Lung, Gastric, Esophageal, Ovarian	Substrate domain
T184			8	Bile, Lung, Ovarian	Substrate domain
T185			7	Gastric, Lung	Substrate domain
Y189			121	Many inc. Breast	Substrate domain
S202			NONE		Substrate domain
Y214	Y214		229	Many inc. Breast	Substrate domain
Y241	Y241	Y241	238	Many inc. Breast	Substrate domain
S296	S296	S296	5	Colorectal, Esophageal, Ovarian, Melanoma	Substrate domain
S298	S298	S298	1	Breast	Substrate domain
S309	S309	S309	1	Esophageal	Substrate domain
S312		S312	6	Gastric, Liver, Lung	Substrate domain
Y317			708	Many inc. Breast	Substrate domain
T332			NONE		Substrate domain
S333			1	Liver	Substrate domain
Y345	Y345		388	Many inc. Breast	Substrate domain
S369			4	Colorectal	Substrate domain
S371			NONE		Substrate domain
S510			NONE		Serine rich
S512			NONE		Serine rich
S572			NONE		Serine rich
Y629			66	Many inc. Breast	DUF3513
Y631			48	Many inc. Breast	DUF3513
S667			NONE		DUF3513
S677			NONE		DUF3513

PTM site probability	Site not covered
25%-45%	
45%-75%	
75%-99%	
99%-100%	

c

Amino acid	Location	Sequence	Modification	Motif prediction
Tyrosine	315 - 317	DAY	[E/D]X**pY**	SHP1 phosphatase substrate motif
Tyrosine	317 - 318	YD	**pY**[A/G/S/T/E/D]	Src kinase substrate motif
Threonine	331 - 334	ETSE	XX[pS/**pT**]E	G protein-coupled receptor kinase 1 substrate motif
Threonine	331 - 335	ETSEK	[E/D][pS/**pT**]XXX	b-Adrenergic Receptor kinase substrate motif
Threonine	331 - 336	ETSEKA	X[pS/**pT**]XXX[A/P/S/T]	G protein-coupled receptor kinase 1 substrate motif
Serine	333 - 335	SEK	[**pS**/pT]X[R/K]	PKA kinase substrate motif
Serine	333 - 335	SEK	[**pS**/pT]X[R/K]	PKC kinase substrate motif
Serine	572 - 576	SIMNS	**pS**XXX[pS/pT]	MAPKAPK2 kinase substrate motif
Serine	572 - 576	SIMNS	**pS**XXXpS*	GSK3 kinase substrate motif
Serine	666 - 668	LSQ	X**pS**Q	DNA dependent Protein kinase substrate motif
Serine	667 - 668	SQ	**pS**Q	ATM kinase substrate motif

Figure 2.6. NEDD9 phospho-motif prediction

(**a**) Venn diagram illustrating the number of phosphorylated residues common and unique between treatment conditions.(**b**) NEDD9 phosphorylation modifications identified by Ti02 enrichment LCMS in all three treatment conditions and NEDD9 protein location. Number of references (PhosphoSitePlus ® as of April 2020) for each modification and cancer types to which this modification is referenced. Colour codes (as in figure 5) relate to the percentage probability of the post-translational modification identification being correct. Grey boxes indicate where protein coverage was not achieved in the control or TGFβ inhibitor conditions corresponding to the TGFβ induced phosphorylation detected. Modifications highlighted by bold borders, where coverage for all three treatment conditions are present was analysed for phopho-motif prediction using the PhophoSitePlus® online analysis tool (**c**).

The other phosphorylated residues which appear with high confidence to be activated by TGFβ alone and with complete data from the other two conditions were in the Serine rich region and the DUF (domain of unknown function) of the NEDD9 protein at S572 and S667, respectively. Phosphorylation at S572 predicts a MAPKAPK2 substrate-binding motif, MAPKAPK2 is involved in the MAPK signalling network and is a known downstream target of TGFβ signalling. Also predicted at S572 is a GSK3 kinase substrate motif. GSK3 (Glycogen Synthase Kinase3) is a serine/threonine kinase and has been implicated in the phosphorylation of over 100 different proteins in a variety of different cellular processes, including cellular proliferation, migration, glucose regulation, and apoptosis (McCubrey et al. 2014). Predicted at the S667 phosphorylation site motif were two kinases that regulate response to DNA double-strand breaks: ATM protein kinase Member of PIKK (PI3K-like protein kinase family) which is also involved in metabolic processes and cell signalling and DNA Dependent protein kinase (DNA-PK) which controls DNA double-strand repair by Non-homologous end joining (NHEJ). DNA-PK is known to cooperate with ATR and ATM in the phosphorylation of proteins involved in the DNA damage checkpoint process (Blackford and Jackson 2017). As DNA damage is fundamental in the development and progression of cancer, it is interesting that these data suggest such mechanisms may be increased by TGFβ signalling, dependent upon NEDD9 phosphorylation. Also interestingly, neither of these site phosphorylations have been previously reported to play any significant roles in the pathogenesis of cancer. Other sites of modification identified with less confidence in this analysis were phosphorylations at T332 and S333. These phosphorylations identified possible binding motifs for G protein-coupled receptor (GPCR) kinase substrates and Protein Kinase A/C (PKA, PKC) substrates. GPCR mediated transactivation of TGFβR1 has been described as leading to increased cellular phospho-Smad2 levels via RhoA/ROCK signalling (Chaplin et al. 2017).

Interestingly, due to the nature of the experimental protocol, in this analysis, several interacting proteins were co-immunoprecipitated alongside NEDD9, and so were also identified by LCMS with high confidence. These proteins included; the Cas protein family members BCAR1 and BCAR3 which also function as docking proteins in the regulation of tyrosine kinase based signalling pathways. This finding corroborates the results in section 2.3, where BCAR1 was co-immunoprecipitated in the NEDD9 protein band validation experiment. Also identified with high confidence as a NEDD9 interacting protein was CITED1, which is a SMAD co-activator that functionally enhances SMAD-mediated transcription by strengthening the SMAD/p300/CBP transcriptional complex; and the pyruvate kinase PKM2, an enzyme involved in the regulation of glucose metabolism.

Together these results confirm that in the MDA-MB-231 cell line NEDD9 undergoes phosphorylation in response to TGFβ signalling at specific amino acid residues. Analysis of these residues indicates that NEDD9 phosphorylation at Y317 is dependent on Src kinase activity in response to TGFβ signalling, and not via the canonical integrin signalling route, consistent with the findings of others (Zheng and McKeown-Longo 2002; Izumchenko et al. 2009). Also indicated, is that signalling pathways involved in the control of DNA damage repair and metabolic processes are regulated by phosphorylation of NEDD9 in response to TGFβ treatment.

2.6 Whole proteome analysis in response to TGFβ signalling

The previous experiments detail the range of methods I used to characterise changes in the expression, and post-translational modifications of NEDD9 in response to TGFβ signalling, and how I used these data to indicate signalling pathways and processes which may be regulated by these changes. With the aim of both validating these findings and better understanding the interplay between TGFβ signalling, NEDD9 expression and pathway activation I moved on to the analysis of the global proteome by mass spectrometry. In using this method I was able to perturb NEDD9 expression to identify pathways that are activated in response to TGFβ, and dependent upon NEDD9. I performed both whole proteome and phospho-proteome analysis of whole-cell extracts from MDA-MB-231 mammosphere cultures expressing either a short-hairpin empty vector or short-hairpin targeting NEDD9 mRNA [methods 9] in either the no-treatment condition or 100 nM TGFβ 24 hour treatment condition. Ideally, this analysis would have been performed on 7-day mammosphere cultures, however, because of the need for equal protein quantity in each sample and the fact that TGFβ induces cell proliferation, this was not possible. Eleven biological replicates were collected in total to allow for quantitative analysis in one 11-plex Tandem mass tag-rapid immunoprecipitation mass spectrometry of endogenous proteins (TMT-RIME) experiment. TMT are isobaric chemical tags that allow for multiplexing in quantitative proteomics experiments. Their use allows for the simultaneous analysis of multiple samples from different treatment groups and gives results that are quantitative and directly comparable. We collected two replicates from the empty vector, no treatment control condition, and three from each of the other three conditions. Biological replicates were prepared from successive passages for both cell lines.

As a control for TGFβ induction of NEDD9 expression, the average levels of NEDD9 protein detected in each condition normalised to TMT reporter levels were calculated (figure **2.7a**). NEDD9 protein expression was induced by TGFβ treatment, and these levels were significantly reduced in the NEDD9

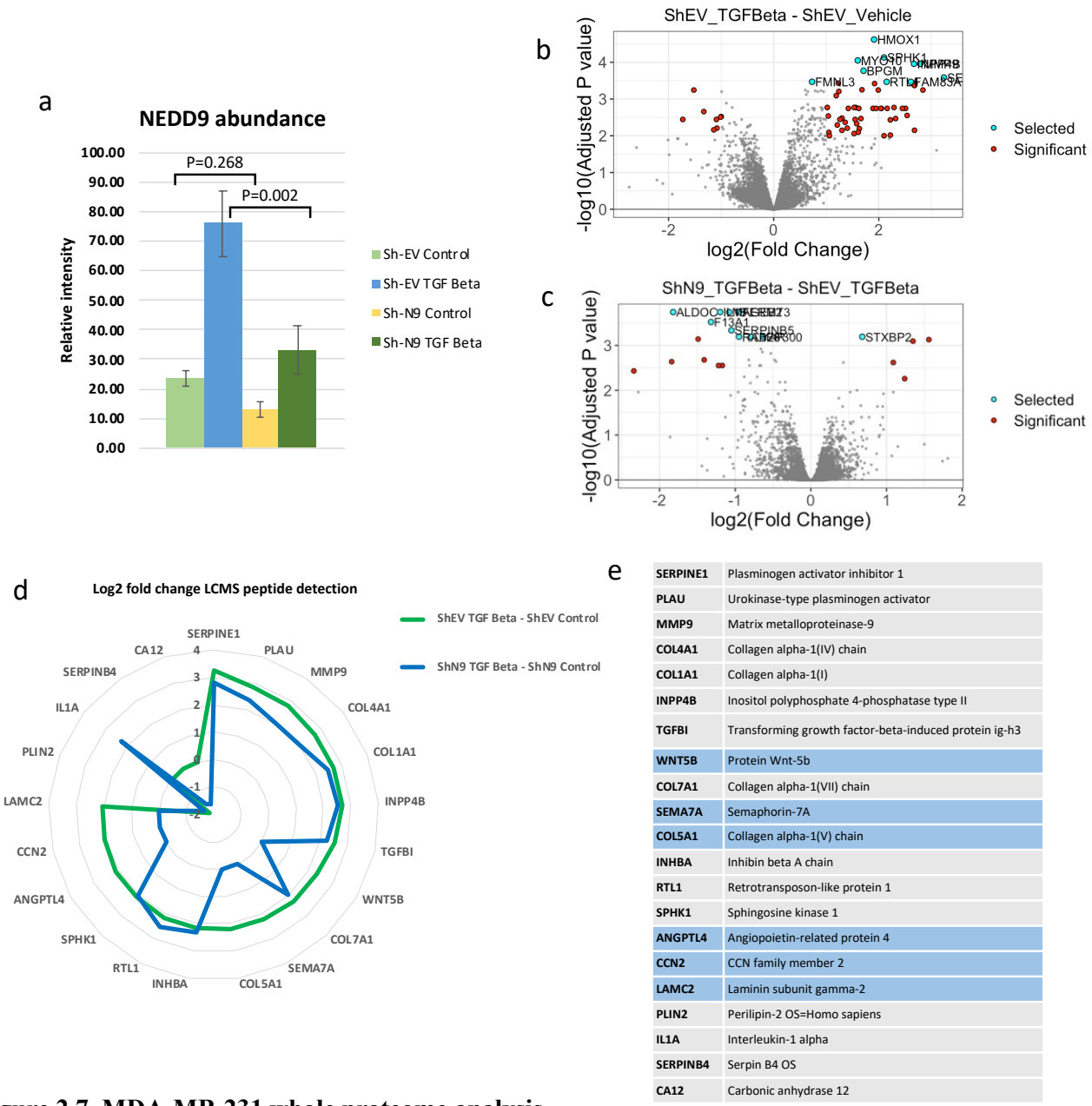
short-hairpin cell line in the presence of TGFβ. However, basal NEDD9 expression levels were not significantly reduced in the presence of the short-hairpin. Table 2.2 below summarises the numbers of total peptides and proteins detected and the number of differentially expressed proteins in each contrast tested.

Total	Number
Peptides	92976
Proteins	8502

Contrasts	Number
ShEV_TGFBeta - ShEV_Vehicle	344
ShN9_TGFBeta - ShN9_Vehicle	477
ShN9_TGFBeta - ShEV_TGFBeta	87

Table 2.2 Total number of peptides and proteins detected in the full proteome LCMS analysis, and the total number of differentially expressed proteins detected in each contrast listed.

The three contrasts listed were selected because when combined, they would identify proteins and pathways which were both regulated by TGFβ and dependent upon NEDD9. Contrast one (ShEV TGF Beta - ShEV Vehicle) identified proteins regulated by TGFβ, and the second contrast (ShN9 TGF Beta - ShN9 Vehicle) identified proteins regulated by TGFβ in the absence of NEDD9. The final contrast (ShN9 TGF Beta - ShEV TGF Beta), compared the proteins regulated in contrasts 1 and 2 to hopefully identify proteins that are both regulated by TGFβ and dependent on NEDD9. Figure **2.7b** represents the fold change in peptides detected in the empty vector short-hairpin TGFβ condition versus the empty vector short-hairpin no-treatment condition (contrast 1), and figure **2.7c** represents the same analysis in the empty vector short-hairpin TGFβ versus NEDD9 short-hairpin TGFβ, where NEDD9 expression was stably perturbed (contrast 3). The proteins represented by both the red and blue coloured dots have significant ($p=<0.05$ absolute logFC $>=1$) change in expression. Only the proteins with higher fold change values (blue dots) were annotated for visualisation purposes. TGFβ treatment caused an increase in protein expression, this can be seen in both table **2.2** (contrast 1) and figure **2.7b**, and this increase in expression also occurs where NEDD9 was perturbed in contrast 2 (table **2.2**). However, when we compare changes in expression in response to TGFβ in the NEDD9 normal versus NEDD9 knockdown condition (contrast 3) (table **2.2** and figure **2.7c)**, it is clear that there was a reduction in the number of proteins that are differentially expressed.



e

SERPINE1	Plasminogen activator inhibitor 1
PLAU	Urokinase-type plasminogen activator
MMP9	Matrix metalloproteinase-9
COL4A1	Collagen alpha-1(IV) chain
COL1A1	Collagen alpha-1(I)
INPP4B	Inositol polyphosphate 4-phosphatase type II
TGFBI	Transforming growth factor-beta-induced protein ig-h3
WNT5B	Protein Wnt-5b
COL7A1	Collagen alpha-1(VII) chain
SEMA7A	Semaphorin-7A
COL5A1	Collagen alpha-1(V) chain
INHBA	Inhibin beta A chain
RTL1	Retrotransposon-like protein 1
SPHK1	Sphingosine kinase 1
ANGPTL4	Angiopoietin-related protein 4
CCN2	CCN family member 2
LAMC2	Laminin subunit gamma-2
PLIN2	Perilipin-2 OS=Homo sapiens
IL1A	Interleukin-1 alpha
SERPINB4	Serpin B4 OS
CA12	Carbonic anhydrase 12

Figure 2.7. MDA-MB-231 whole proteome analysis

(a) Mean NEDD9 peptide abundance in each condition, units represent relative intensity of signal compared to a TMT reporter. Error bars represent ±s.d. of three independent biological replicates (two in EV control condition). Volcano plots representing the the log2 fold change values of peptides detected which are (**b**) TGFβ specific in the empty vector control vs. empty vector TGFβ condition, and (**c**) both TGFβ/NEDD9 specific in the short-hairpin NEDD9 plus TGFβ vs. empty vector TGFβ condition. Values represent Log2 fold change and coloured dots (red and blue) represent peptides which have a significant (p=<0.05) change. (**d**) Peptide detection fold change in the two comparisons; Sh-EV control vs. Sh-EV TGFβ, Sh-N9 control vs. Sh-N9 TGFβ, to identify peptides that are both dependent upon TGFβ signaling and NEDD9. Log2 fold change values of >1.5 and <-1.5 were included where the adjusted p value was <0.05. **(e)** Table of proteins significantly up- or down-regulated by TGFβ treatment, proteins highlighted in blue are those which are NEDD9 dependent and TGFβ specific.

Fold changes in the highly significant (Log2FC = >1.5 or <-1.5, and adjusted p=<0.05) group of peptides are illustrated in figure **2.7d**. Here, this radar plot illustrates the proteins which were differentially expressed in response to TGFβ treatment (contrast 1) in green and compares them to proteins that were differentially expressed in response to TGFβ treatment in the NEDD9 knock-down condition (contrast 3) in blue. The full list of differentially expressed proteins in response to TGFβ treatment is listed in figure **2.7e,** and proteins that demonstrated a change in expression between contrast 1 and contrast 3 are highlighted in blue. We hypothesise that the regulation of these proteins is TGFβ specific but dependent on NEDD9.

Of this NEDD9 dependent and TGFβ specific list of proteins, many are involved in cell adhesion and integrin binding. For example, COL5A1, SEM7A, CCN2 and LAMC2 were all down-regulated where NEDD9 was perturbed, this indicates that NEDD9 is important in mediating cell adhesion and integrin-binding processes via these proteins/pathways. WNT5B, also in the NEDD9 dependent and TGFβ specific group, is a key regulatory factor in Wnt signalling (along with WNT5A) which activates the Wnt pathway via both canonical and non-canonical mechanisms. WNT5B was recently shown to be a key regulatory factor in the stem-like phenotype of basal breast cancer (Jiang et al. 2019). ANGPTL4 was also differentially regulated in this group, this is a gene known to play roles in many biological processes, but is best characterised for its role in regulating lipid metabolism. Under hypoxic conditions, ANGPTL4 plays a key role in regulating the processes of cancer progression and metastasis and is directly induced by TGFβ via SMAD signalling in breast cancer metastasis to the lung (Padua et al. 2008). Here the authors demonstrate a role for secreted ANGPTL4 in the metastatic processes of intravasation from the primary tumour and extravasation at metastatic sites via disruption of cell-cell junctions. Interestingly, recently ANGPTL4 has also been identified as an inhibitor of Wnt signalling (Kirsch et al. 2017), a signalling pathway that results in the expression of genes regulating cell proliferation, migration and differentiation. Another component of the Wnt signalling pathway is GSK3, a substrate motif for which was identified as being activated in NEDD9 by TGFβ treatment in section 2.4. Our data point towards a combination of NEDD9 dependent WNT and ANGPTL4 signalling in response to TGFβ treatment occurring in our cell culture model, which could influence downstream cellular responses such as survival and differentiation through the integrin pathway.

This analysis identified a group of proteins, the expression of which were differentially regulated in response to TGFβ. When NEDD9 expression was perturbed, a number of these proteins were no longer regulated by TGFβ signalling, indicating that their regulation is dependent upon the presence of NEDD9. These proteins broadly fall into the category of proteins regulating the processes of cell adhesion and

integrin signalling, however, two proteins WNT5B and ANGPTL4 have previously been demonstrated to play roles in breast cancer oncogenesis. As mentioned in the introduction, Wnt signalling has been implicated in the regulation of NEDD9 in gastric and colorectal cancer (Y. Li et al. 2011; C. Zhang et al. 2019). The 2019 publication of Jiang and colleagues describe a role for Wnt signalling in the regulation of the stem-like phenotype in basal breast cancer (Jiang et al. 2019), but they did not investigate any association with NEDD9 expression. When considered together with published data, this analysis indicates a role for NEDD9 in TGFβ regulated Wnt signalling in the regulation of stem-cell processes in the Claudin-low subtype of breast cancer cell lines.

2.7 Differential gene expression analysis in response to TGFβ signalling

Next, with the aim of better understanding the association between TGFβ signalling, NEDD9 expression and signalling pathway activation, in addition to validating the signalling pathways identified in both the phospho-proteome and global proteome analyses, I performed total mRNA sequencing analysis. This analysis was performed to identify signalling pathways that are both activated in the presence of TGFβ, and dependent upon NEDD9. MDA-MB-231 mammosphere cultures were prepared, with NEDD9 expression perturbed by siRNA as previously described, in the vehicle control, 100 nM TGFβ, or 100 nM TGFβ plus 10 μM TGFβ inhibitor SB431542 24 hour treatment conditions. Total RNA was collected at 48 hours post-experiment initiation [methods 13], a time point selected to allow for maximal siRNA activity and sufficient TGFβ treatment for transcriptional activation and detection of downstream effects. Library preparation, RNA sequencing and bioinformatic analysis were performed as outlined [methods 6]. For the data processing, differential expression was analysed in the scramble siRNA control versus scramble siRNA TGFβ condition to identify TGFβ specific responses; the scramble siRNA TGFβ versus NEDD9 siRNA TGFβ to identify TGFβ specific, NEDD9 dependent responses; and a third no treatment day 0 versus day 2 contrast to control for gene expression induced by suspension mammosphere culture. It is known that the process of maintaining the MDA-MB-231 cell line in suspension mammosphere cultures induces autocrine TGFβ signalling (Bruna et al. 2012), however, the levels of TGFβ signalling are significantly lower in the no treatment control condition than in the TGFβ treatment condition. The number of upregulated and downregulated genes in each contrast is listed below in table 2.3:

Contrast	Up regulated	Down regulated
D2 control - D0 control	8196	6423
si-Scramble Beta - si-Srcamble Control	6213	5500
si-Scramble Beta - si-NEDD9 Beta	5233	4339

Table 2.3. The number of differentially expressed genes identified by mRNA-seq in each listed contrast.

First, to validate TGFβ signalling activity in the assay I used the molecular signatures database (MSigDB) to examine the data from the scramble siRNA TGFβ versus NEDD9 siRNA TGFβ comparison for gene set enrichment scores (ES) using the oncogenic signature gene set (www.gsea-msigdb.org). This dataset is a curated collection of gene sets representing signatures of cellular pathways which are frequently dysregulated in cancer. As anticipated, this analysis showed a correlation in the scramble siRNA TGFβ condition with genesets upregulated in TGFβ signalling, and samples from the NEDD9 siRNA group negatively correlated with this signature (figure **2.8a**). To further validate this observation, I looked at the normalised expression levels in transcripts per million (TPM) of the SMAD2 and SMAD3 components of the TGFβ pathway. SMAD2 expression remained relatively stable across all conditions (figure **2.8b**). Interestingly, however, TGFβ treatment significantly reduced levels of SMAD3 expression (paired two-sample t-test p=<0.05), and this was further decreased where NEDD9 expression was perturbed by siRNA (figure **2.8c**). SMAD3 has previously been shown to interact with members of the p130Cas family in response to TGFβ treatment, and in 2000 Liu et al (X. Liu et al. 2000) demonstrated how this interaction triggers proteasomal degradation of NEDD9 and subsequent increases in NEDD9 mRNA expression in a positive feedback loop. This was followed by the observations of Kim et al in 2008, and Wendt et al in 2009 that this interaction was responsible for the downregulation of canonical TGFβ signalling pathways in cancer and the increased activity of non-canonical pathways such as MAPK signalling pathways (W. Kim et al. 2008; M. K. Wendt, Smith, and Schiemann 2009). However, the effect I observed was in the opposite direction, where NEDD9 perturbation resulted in a decrease in SMAD3 expression. This result is, nonetheless, still interesting and merits further investigation. Analysis of NEDD9 normalised TPM expression levels (figure **2.8d**) showed TGFβ treatment significantly increased NEDD9 expression (paired two-sample t-test p=<0.05, dashed bracket), and in the presence of the NEDD9 siRNA this increase was significantly reduced, in both the control and TGFβ conditions (paired two-sample t-test p=<0.05, bold brackets) but not in the TGFβ plus TGFβ inhibitor condition. Next, I queried the data for signatures enriched in processes we had observed in previous investigations. In section 2.5, I identified that BCAR1 was immunoprecipitated alongside NEDD9, and was also differentially regulated by TGFβ treatment.

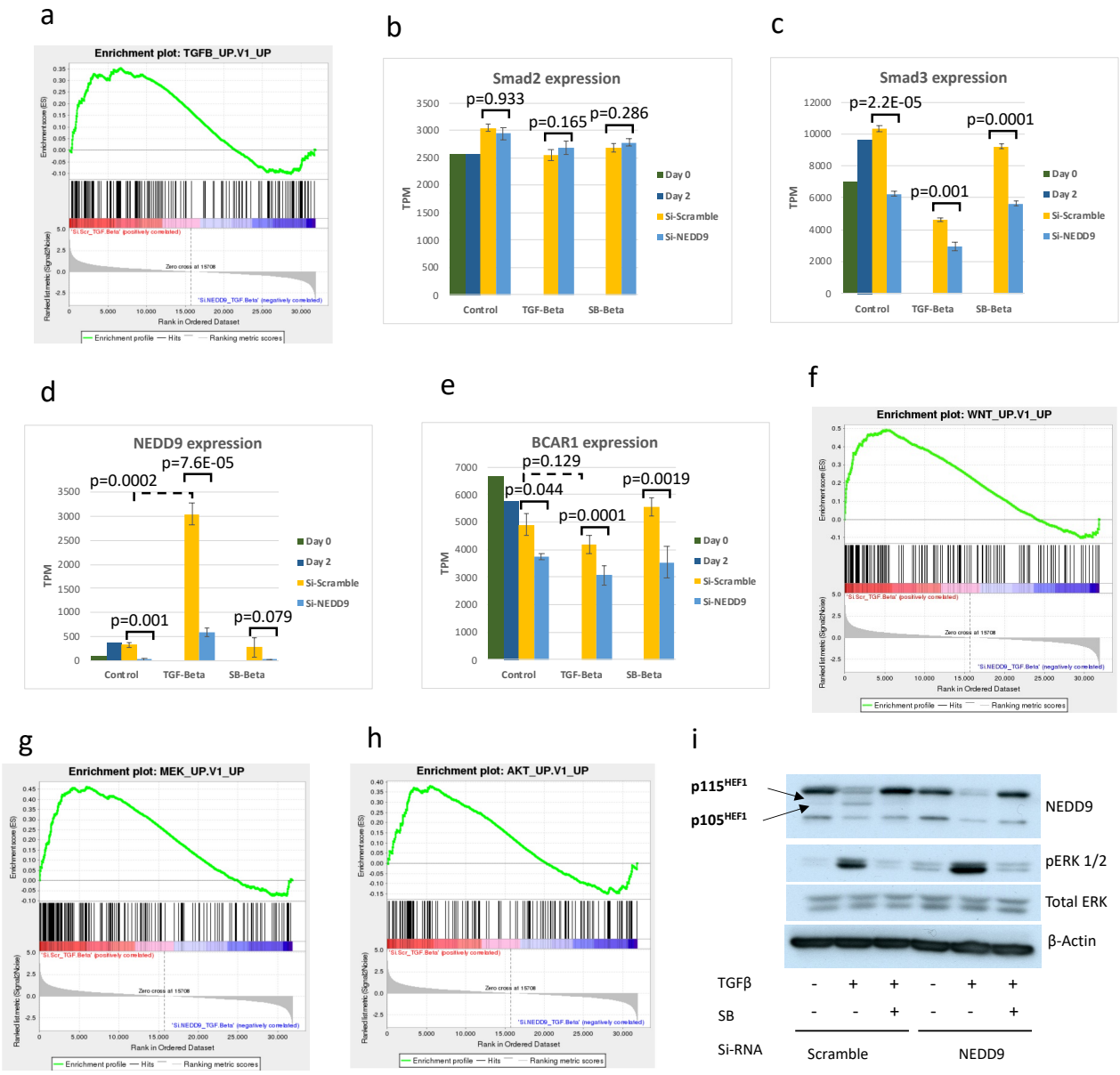
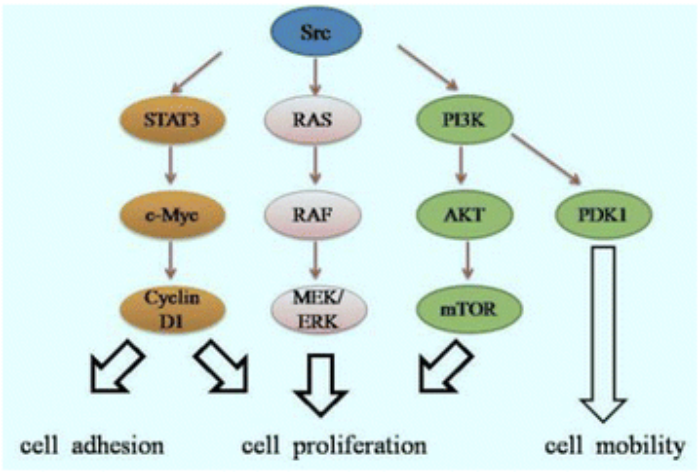


Figure 2.8. RNA-seq analysis in the scramble siRNA TGFβ versus NEDD9 siRNA TGFβ contrast
Enrichment plots for correlation with oncology gene sets from the molecular signatures database (MSigDB, www.gsea-msigdb.org) for TGFβ (**a**), WNT (**f**), MEK (**g**) and AKT (**h**) signalling signatures. RNA expression for SMAD2 (**b**) SMAD3 (**c**) NEDD9 (**d**) and BCAR1 (**e**) in transcripts per million (TPM) from day 2 mammosphere cultures in vehicle control, 100 nM TGFβ, or 100 nM TGFβ plus 10μM TGFβ inhibitor SB431542 (SB-Beta) 24 hour treatment conditions, in the si-scramble versus si-NEDD9 group. Significance calculated from paired two sample t-test, p=<0.05. Error bars represent ±s.d. of four independent biological replicates. (**i**) Western blot in the same treatment conditions for pERK1/2, ERK and NEDD9 and expression with Beta Actin as a loading control. The western blot figure is representative of at least three biological replicates.

The analysis of BCAR1 expression in TPM (figure **2.8e**) followed the same pattern of reduction in expression in response to TGFβ treatment, but this reduction was not significant (paired two-sample t-test p=<0.05, dashed bracket). There was, however, a significant reduction in BCAR1 expression in the NEDD9 siRNA group in all treatment conditions (paired two-sample t-test p=<0.05, solid brackets), and also the same pattern of change in expression was observed in response to TGFβ treatment. This indicates that the association between NEDD9 and BCAR1 is subject to regulation by a yet undefined mechanism, and not purely stoichiometric. In the whole proteome analysis (section 2.6), WNT5B was identified as a protein that was both TGFβ specific and NEDD9 dependent. When we looked at the enrichment plot for gene sets upregulated in Wnt signalling (figure **2.8f**), the scramble siRNA TGFβ condition demonstrated a positive correlation, and samples from the NEDD9 siRNA group negatively correlated with this signature.

As previously discussed in section 2.5, the IP NEDD9 TiO2 enrichment identified a Src Kinase substrate-binding motif at NEDD9 Y317. Other groups have demonstrated an association between NEDD9, Src signalling, and the PI3K and MAPK pathways, including a NEDD9 dependent positive feedback mechanism for the continuous activation of Src (Izumchenko et al. 2009; Zheng and McKeown-Longo 2002). Src kinases regulate a wide range of cellular effects via their influence on several signalling pathways as outlined below:



h

Cell signalling pathways induced by Src kinases (Jiao et al. 2018)

Investigation of oncology gene sets downstream of Src indicated that the scramble siRNA TGFβ sample group was enriched in both the MEK (figure **2.8g**) and AKT (figure **2.8h**) gene sets suggesting that signalling via both the RAS/RAF/MEK/ERK and PI3K/AKT/mTOR pathway may also be occurring in response TGFβ. Again, samples from the NEDD9 siRNA sample group were negatively correlated with

both of these gene signatures. However, at the protein level (figure **2.8i**), whilst we saw an increase in ERK signalling, represented by an increase in phospho-ERK in response to TGFβ treatment, this response did not appear to be dependent upon NEDD9.

Next, to gain an overview of the cellular processes involved in the experiment, I performed GO analysis (genentology.org 2019-01-01. 10.5281/zenodo.2529950) (Gene Ontology Consortium 2021; Ashburner et al. 2000) on the comparisons: 1) scramble siRNA control versus scramble siRNA plus TGFβ and 2) scramble siRNA TGFβ versus NEDD9 siRNA plus TGFβ. These comparisons aimed to identify processes that were primarily enriched by TGFβ treatment, and secondarily dependent upon NEDD9. To perform this analysis, I selected genes that were up-regulated (>2 Log2 fold change, p adjusted<0.05) in condition 1 (scramble siRNA control versus scramble siRNA plus TGFβ) (n=1034), and are therefore TGFβ dependent. I then compared this analysis to the GO analysis for up-regulated genes (>2 Log2 fold change, p adjusted<0.05) in condition2 (scramble siRNA plus TGFβ versus NEDD9 siRNA plus TGFβ) (n=204). What was immediately apparent was the number of significantly (>2 Log2 fold change, p adjusted<0.05) upregulated genes was greater than fivefold higher in condition 1 when compared with condition 2. Figure **2.9** shows the Go analysis of significantly up-regulated genes in condition 1, the GO terms listed demonstrate fold enrichment for over-represented GO annotations (FDR p<0.05) in each of the three sub-ontologies; biological process, molecular function and cellular compartment. The analysis of biological processes show the majority of the upregulated genes with higher fold change values were involved in biological processes related to DNA replication, this is probably a reflection of the increased cellular self-renewal we observe in this condition. Also significantly enriched (FC 7.91) is SMAD protein phosphorylation which would be expected as a response to TGFβ treatment. Also evident in the biological process sub-ontology was an increase in ERK signalling, as discussed earlier. The molecular function analysis reveals an enrichment of processes involved in signalling receptor activity and growth factor binding, consistent with TGFβ signalling activity. The increase in signatures involving extracellular matrix and collagen in the cellular compartment sub-ontology is likely a reflection of the increase in NEDD9 expression in this condition and the role of NEDD9 as a scaffolding protein. The inclusion of annotations relating to the extracellular space in this ontology is an indication of the role of NEDD9 in integrin signalling since integrins are the primary transmembrane linkers between the actin cytoskeleton and the extracellular matrix.

When I compared the list of GO annotations from both conditions, not one of the terms in condition 1 was present in condition 2 analysis (not shown). This indicates that NEDD9 is necessary for the regulation of the processes described in condition 1.

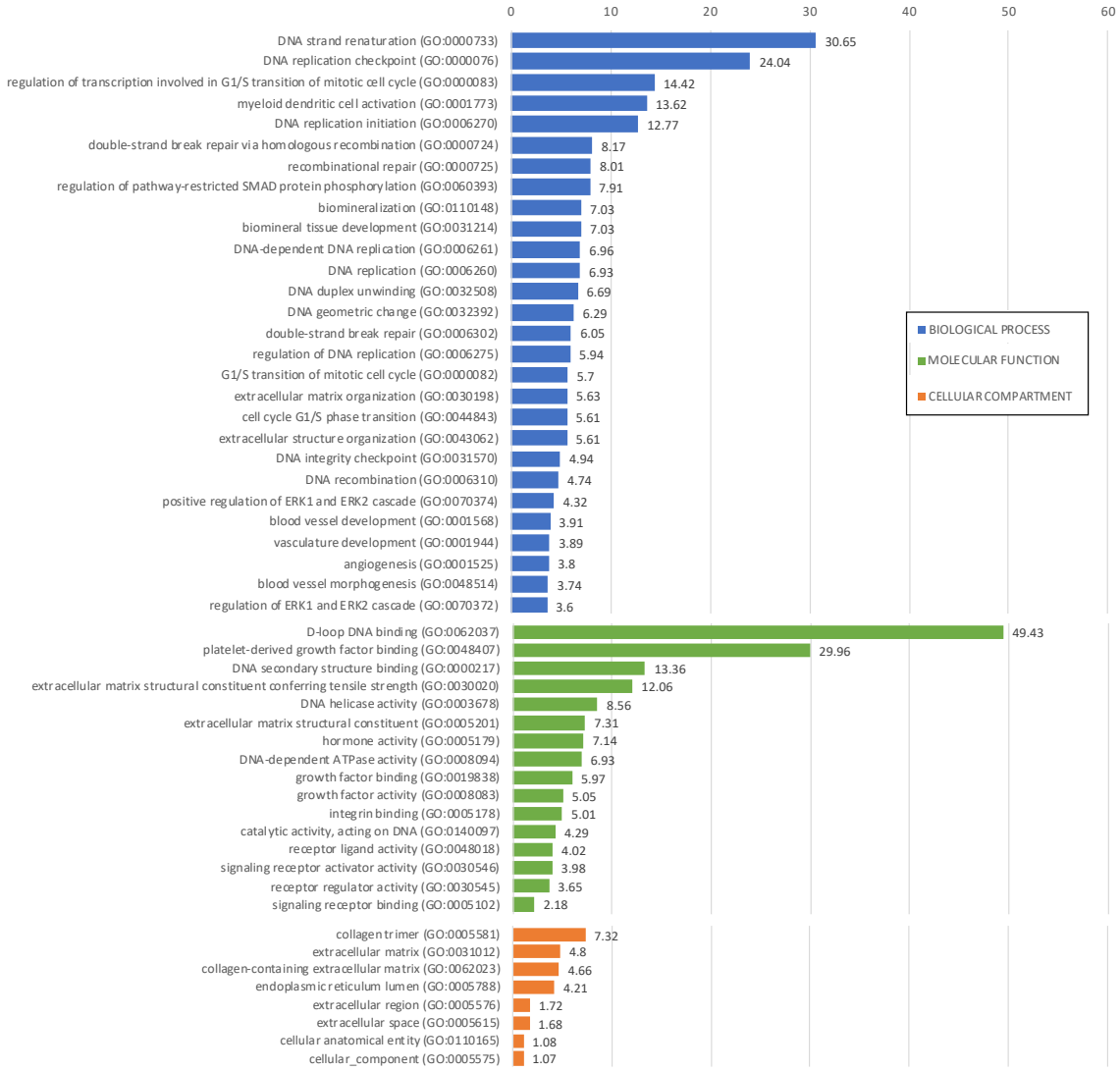


Figure 2.9. GO analysis of the scramble siRNA Control versus the scramble siRNA TGFβ conditions

Fold enrichment values for overrepresented GO annotations using (genentology.org 2019-01-01. 10.5281/zenodo.2529950) FDR P < 0.05. GO enrichment analysis for upregulated genes showing >2 log FC ($p<0.05$) increase in expression in the si-Scramble Control versus si-Scramble Beta treatment condition. The same gene expression list was used for analysis of each of the sub-ontologies (biological process, molecular function and cellular compartment).

In this analysis, I both strengthened the findings of the previous analyses and further investigated the cellular processes which are both activated by TGFβ treatment and dependent upon NEDD9. Through analysis of gene expression, I validated the effects on TGFβ signalling, and the expression of SMADs, NEDD9 and BCAR1, a NEDD9 partner protein, which is induced by TGFβ treatment. I was also able to reinforce earlier observations that Wnt signalling and Src activation are likely mechanisms by which TGFβ induces breast tumour initiating cell self-renewal in a NEDD9 dependent manner. Evident from the analysis of gene ontology annotations is that the processes of DNA damage repair, and DNA replication, and extracellular matrix remodelling are increased in the presence of TGFβ treatment and are dependent upon NEDD9.

2.8 Characterisation of NEDD9 isoform expression.

There are 16 NEDD9 transcript variants identified to date, however, as previously discussed, only four of these variants are protein-coding. These four variants are expressed from two different transcriptional start sites (figure **1.2**), with only variants one and three expressing a full-length protein product. The expression of these different isoforms has not been investigated with regards to cancer incidence and survival. Due to the differences in isoform expression and the probable changes in the resulting protein sequence, I was curious to know if these differences had any effect on NEDD9 function. NEDD9 variant three has a transcription start site which is upstream of that for variant one and also includes an alternative in-frame exon, whereas isoform four encodes a slightly shorter protein product with one alternative exon to both variants one and three. Could differences in isoform expression, which would be mirrored in the amino acid sequence of the resulting protein confer different binding abilities for kinases or other NEDD9 interacting proteins? And does TGFβ treatment lead to the expression of alternative isoforms compared to the no-treatment control?

To address this question I first looked at the expression of NEDD9 isoforms in a large clinical cohort, the Cambridge University hospitals Personalised Breast Cancer Program study (PBCP). The PBCP program is based at Addenbrooke's hospital, Cambridge, and aims to sequence the RNA and DNA of 2000 breast cancer patients. Analysis of the RNA-seq data for NEDD9 isoform expression was carried out by Raquel Manzano Garcia, another PhD student in the Caldas lab. To estimate Transcript Per Million (TPM) values, Raquel employed Salmon (Patro et al. 2017) version 0.14.1 on read-based mode with default parameters. The plots in figures **2**.10**a** and **2**.**10b** represent RNA-Seq analysis of NEDD9 isoform expression in transcripts per million (TPM) of normal samples (figure **2**.**10a**) and breast tumour samples

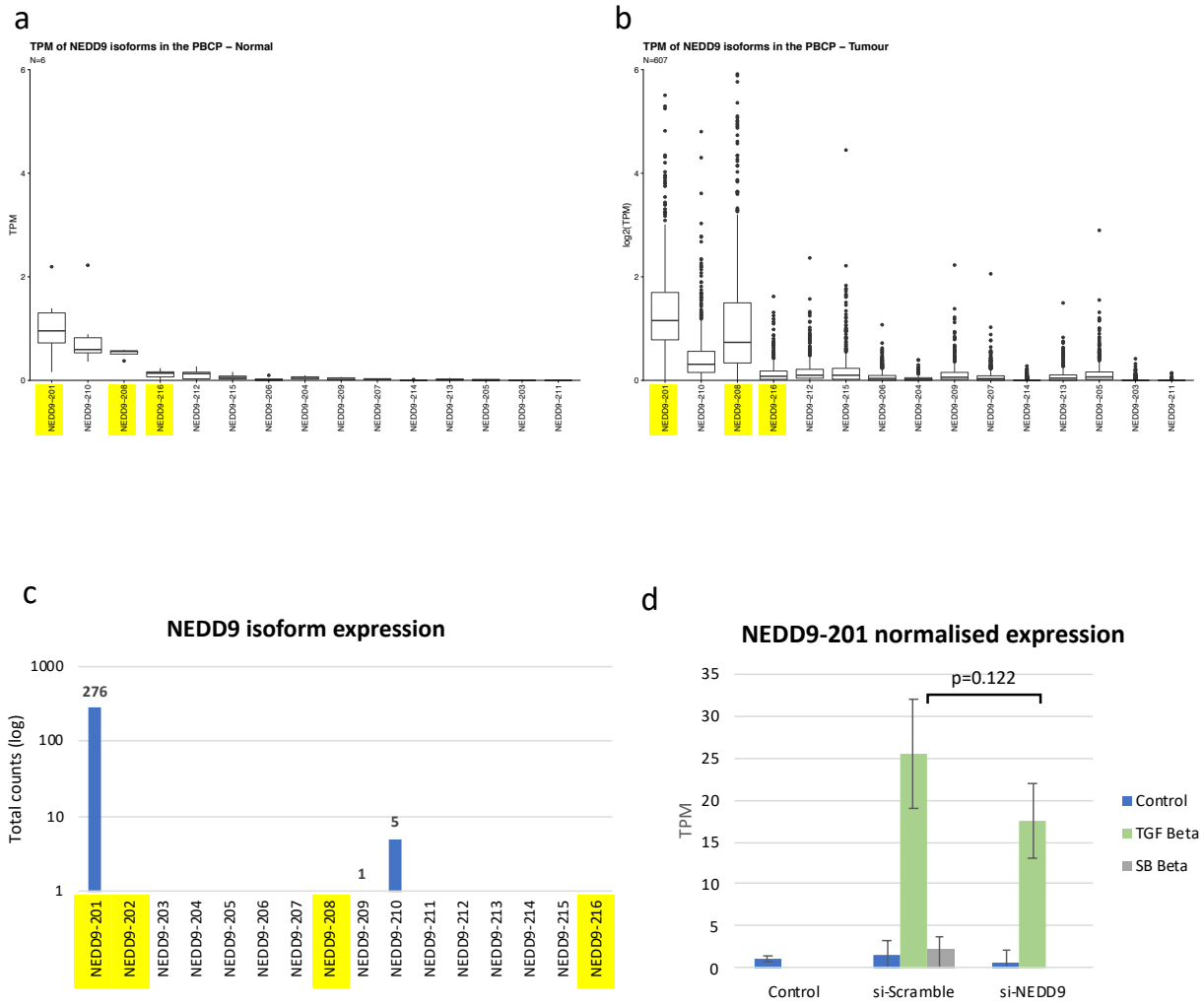


Figure 2.10. NEDD9 isoform expression

(**a**) RNA-Seq NEDD9 isoform expression analysis of normal samples from PBCP data (n=6). Expression is normalized to transcripts per million (TPM). **(b)** RNA-Seq NEDD9 isoform expression analysis of patient breast tumour samples from PBCP data (n=607). Expression is normalized to transcripts per million (TPM). Isoforms highlighted in yellow in figures **a** and **b** are the known protein coding variants. **(c)** NEDD9 isoform expression of all known isoforms from cell line RNA-Seq analysis. Data represents total counts for each isoform detected. Isoforms highlighted in yellow are the known protein coding variants. **(d)** Normalised TPM expression of NEDD9-201 isoform from cell line RNA-Seq analysis in vehicle control (DMSO), 100nM TGFβ, or 100nM TGFβ plus 10μM TGFβ inhibitor SB431542 seven day treated MDA-MB-231 suspension cultures. NEDD9 was perturbed using commercial SiRNA. Error bars represent ±s.d. of three biological replicates for no-treatment control and four for siRNA conditions.

(figure **2.10b**). In this analysis, only 15 of the 16 known transcript variants were detected, with the protein-coding variant NEDD9-202 not found. What is evident is that isoform expression overall was comparable but higher in some of the protein-coding isoforms highlighted in yellow. A similar pattern of expression was seen in both the normal and tumour cohorts, although the number of samples in the normal group is small. The isoforms with the highest expression appear to be NEDD9-201, and NEDD9-208 (variants two and three). NEDD9-208 encodes a full-length protein product but more surprisingly, the variant with the highest expression levels is NEDD9-201. This variant lacks several exons and encodes a product of only three exons with a distinct C-terminus and truncated SH3 domain, resulting in a protein with a predicted molecular weight of 19kDa. Unfortunately, this protein is not detected by the commercial antibody used in my analyses (ab18056, Abcam), as the epitope recognised by this antibody is not present in the truncated protein. Because of this, it was not possible to evaluate NEDD9-201 protein expression by western blot using this antibody. Similarly, when I returned to the full proteome analysis data to query isoform expression, only peptides common to all four isoforms were identified, therefore it is not possible to use this data to evaluate differential isoform expression.

To investigate if this pattern of expression is mirrored in MDA-MB-231 mammosphere cultures, and if any changes were observed in response to TGFβ treatment, I returned to the RNA-seq analysis. Figure **2.9c** shows the total normalised counts for NEDD9 isoform expression in the control and scramble siRNA condition where NEDD9 expression was not perturbed. Expression across the sixteen known transcripts is shown with protein-coding variants highlighted in yellow. The overall detection of NEDD9 transcripts was relatively low with only three transcripts identified, but the predominant isoform detected was the protein-coding NEDD9-201 isoform. Interestingly, neither analysis identified NEDD9-202 (transcript variant 1) which is the other isoform known to encode the full-length protein. Further investigation of NEDD9-201 expression in all sample conditions (figure **2.10d**) shows that this isoform is induced by TGFβ treatment and expression is abrogated by treatment with the TGFβ inhibitor, indicating TGFβ dependent expression. This isoform is targeted by the siRNA treatment, however, the reduction in expression by NEDD9 siRNA treatment in the TGFβ condition is not statistically significant (paired two-sample t-test $p=0.122$).

What is not known from this analysis is whether the failure to detect all NEDD9 isoforms is because they are genuinely not expressed, or insufficient read depth in the sequencing protocol. Illumina recommends a read depth of between 100-200 million reads to differentiate isoform expression; in this experiment, we obtained a depth of between 35 million and 56 million reads per sample. However, from these data, it can be seen that the expression of the NEDD9 isoforms is not uniform in normal tissue or breast cancer and

that expression of the NEDD9-201 isoform is both predominant and regulated by TGFβ in mammosphere culture from the Claudin-low subtype cell line MDA-MB-231. How this data relates to the previous proteomic investigations is not clear. For the proteomic analyses involving immunoprecipitation of NEDD9, the commercial antibody which does not recognise the NEDD9-201 isoform was used. This means that this isoform which is recognised as predominantly expressed by mRNA-seq analysis was not represented in the NEDD9 phospho-proteome analysis, however, the signalling pathways suggested by the phospho-proteome analysis were validated in the RNA-seq analysis. The significance of the NEDD9-201 truncated transcript warrants further investigation.

2.9 Summary

In this chapter, I used several different experimental techniques to fully characterise NEDD9, its expression and post-translational modification in a Claudin-low breast cancer subtype cell line in response to TGFβ signalling. I demonstrated that NEDD9 is frequently overexpressed in many cancer types, but that this overexpression does not correlate to breast cancer subtype or overall survival in the METABRIC cohort. I demonstrated that TGFβ induces NEDD9 expression only in Claudin-low cell lines prototypic of TGFβ oncogenic effects. Both the proteomic and genomic investigations highlighted similar signalling pathways which may be responsible for NEDD9 dependent TGFβ effects in the Claudin-low subset, principally the Wnt signalling pathway and mechanisms downstream of Src kinase activity. NEDD9 dependent TGFβ activation of Src kinase activity suggests cross-talk between the MAPK/ERK and PI3K/AKT signalling pathways may be occurring, and coordination between these pathways is becoming the focus for an increasing number of studies involving the use of dual AKT/ERK inhibitors in cancer treatment (Cao et al. 2019). Finally, I explored the differential expression of NEDD9 isoforms both in our cell culture system, and in the PBCP cohort of breast cancer patients, and found that surprisingly the truncated isoform NEDD9-201 appeared to be the predominantly expressed isoform in both cohorts. In the following chapter, we will build upon these findings to identify partner proteins binding with NEDD9 in response to TGFβ treatment.

3. Identification of NEDD9 interacting proteins

3.1 Preface

In the previous chapter, I thoroughly characterised the NEDD9 protein by mass spectrometry (MS), identified differential phosphorylation induced by TGFβ treatment, and explored the expression of NEDD9 isoforms in suspension culture conditions. I also evaluated the extent of NEDD9 gene mutations and isoform expression in both pan-cancer and breast cancer-specific clinical datasets. As previously discussed, NEDD9 is a non-catalytic scaffolding protein containing docking sites for proteins involved in the regulation of numerous signalling pathways (Nikonova et al. 2014; Tikhmyanova, Little, and Golemis 2010). Via this docking mechanism, it is understood that NEDD9 can enable signalling transduction pathways to converge, regulating downstream effects.

This chapter aims to further understand these complex interactions, and identify how these relationships might govern the TGFβ specific effects we observe in Claudin-low breast cancer, I performed a series of analyses to identify potential interactors. The analyses in the previous chapter identified several potential NEDD9 interacting proteins, however, none of the previous methods used was quantitative. To confirm that these interactions are mediated by TGFβ, I moved on to using the following semi-quantitative and quantitative techniques.

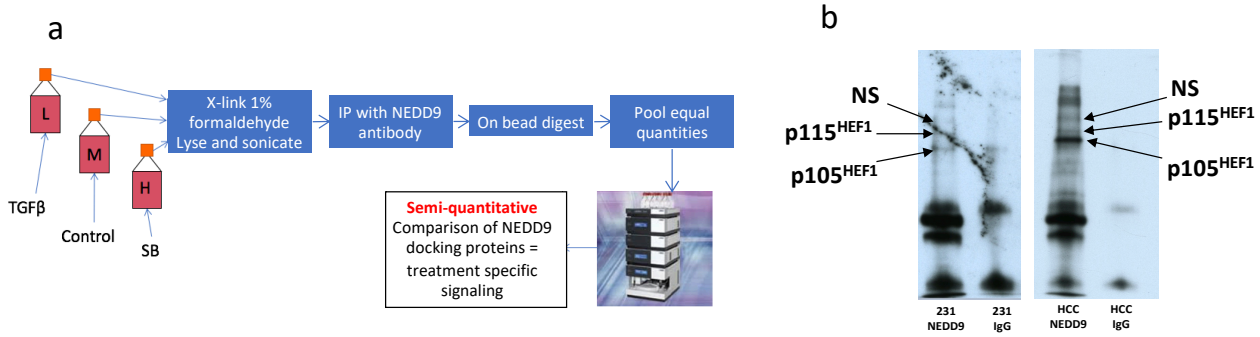
3.2 Semi-quantitative analysis of NEDD9 interacting proteins

In section 2.4, I demonstrated the successful immunoprecipitation of NEDD9. Using the same antibody, I went on to perform a semi-quantitative analysis of NEDD9 interacting proteins using Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) on SILAC (Stable isotope labelling by amino acids in cell culture) labelled cells. The RIME method (Mohammed et al. 2016) allows for the study of protein complexes by MS. Endogenous protein complexes are stabilised using formaldehyde, and are then co-immunoprecipitated with your protein of interest for MS analysis. SILAC is a protocol in which cell lines are grown in media containing amino acids labelled with stable (non-radioactive) isotopes of amino acids, typically L-Lysine and L-Arginine. In the case of L-Lysine, the replacement of $^{12}C_6$ and $^{14}N_2$ with the heavier isotopes of $^{13}C_6$ and $^{15}N_2$ results in a mass shift in

metabolites of these amino acids, of +4Da and +8Da respectively. For L-Arginine, replacement of $^{12}C_6$ and $^{14}N_4$ with $^{13}C_6$ and $^{15}N_4$ results in a mass shift of +6Da and +10Da respectively. Stable incorporation of these isotopes from labelled cell culture media allows for the generation of three distinct cell lines - typically labelled "light" for the native amino acids, "medium" for incorporation of ^{13}C, and "heavy" for the incorporation of both ^{13}C and ^{15}N - the peptides from which can be separated by MS based on their relative mass. These three differentially labelled cell lines can then be treated independently before being pooled for semi-quantitative and comparative MS analysis. The two cell lines, MDA-MB-231 and HCC1954 were each cultured separately in media containing isotopes of L-Lysine and L-Arginine to produce "light" "medium" and "heavy" stable cell lines [methods 7]. These cell lines were previously demonstrated to be representative of TGFβ oncogenic effects, where TGFβ treatment increases breast tumour initiating cell (BTIC) formation (in MDA-MB-231) and reduces BTIC formation (in HCC1954). SILAC-labelled cells were cultured as previously described [methods 10] to create 7-day mammosphere cultures, which were then treated for three hours with either 100 nM TGFβ, 10 μM TGFβ inhibitor SB431542 (SB), or with DMSO (no treatment control), as outlined in figure **3.1a**. "Light" labelled controls were treated with TGFβ, "heavy" labelled cell lines with the TGFβ inhibitor, and the "medium" labelled cell lines were used for the no-treatment control. Following treatment, the mammospheres were then collected, cross-linked, and processed [methods 2]. Cell lysates were quantified for protein levels using a standard BCA assay (Pierce™) and equal quantities of total protein from each condition were pooled for the IP. Following LCMS, SILAC data were analysed using Proteome Discoverer v.1.3. The data were arranged in order of abundance and the top fold change values for TGFβ over SB431542 treatment were calculated.

The western blot in figure **3.1b** shows enrichment of the three NEDD9 bands corresponding to $p105^{HEF1}$, $p115^{HEF1}$ and the non-specific bands (labelled) in both MDA-MB-231 and HCC1954 when compared to the IgG isotype control, confirming successful immunoprecipitation of NEDD9. Each IP here represents a pool of all three treatments.

Proteins that were co-immunoprecipitated with NEDD9 are listed in the table (figure **3.1c)** for MDA-MB-231 and table (figure **3.1d)** for HCC1954. Values represent significant ($p<0.05$) fold change in peptides detected in the TGFβ over SB-431542 condition. In the MDA-MB-231 analysis, Breast cancer anti-estrogen resistance protein 1 (BCAR1) represents a good internal positive control, as it is known hetero-dimerise with the conserved c-terminal domain of NEDD9 (Law et al. 1999). Additionally, I also observed BCAR1 as a NEDD9 interacting protein in the phospho-MS analysis (section 2.5) and observed that its expression was differentially regulated by TGFβ treatment, and expression was reduced where



c

MDA-MB-231

Fold Ch.	Uniprot ID	
5.6	P14618	**Pyruvate kinase isozymes M1/M2**
1.8	P60709	Actin, cytoplasmic 1
1.2	P56945	**Breast cancer anti-estrogen resistance protein 1**
1.2	P16615	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2
1.2	Q13162	Peroxiredoxin-4
0.7	P04908	Histone H2A type 1-B/E

d

HCC1954

Fold Ch.	UniProt ID	
29.0	P08670	Vimentin
15.0	P13645	Keratin, type I cytoskeletal 10
7.1	P04264	Keratin, type II cytoskeletal 1
5.1	P11142	**Heat shock cognate 71 kDa protein**
2.5	P62805	Histone H4*
1.8	P68431	Histone H3.1*
0.5	Q14451	Growth factor receptor-bound protein 7

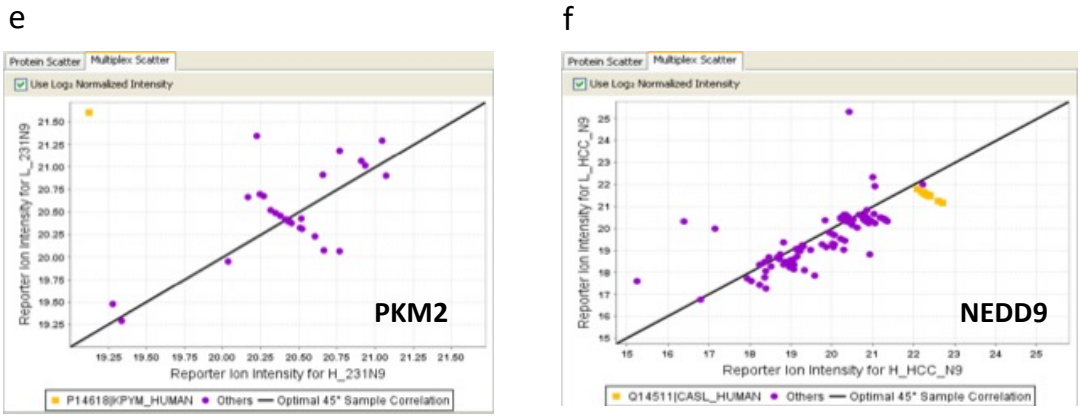


Figure 3.1. Semi-quantitative identification of NEDD9 docking proteins in response to TGFβ treatment by SILAC-RIME LCMS

(a) Schematic demonstrating the workflow for SILAC-RIME identification of NEDD9 interacting proteins in MDA-MB-231 and HCC1954 seven day mammosphere cell cultures, in either 100nM TGFβ, 10μM TGFβ inhibitor SB431542, or no treatment control conditions. Treated cultures were cross-linked with 1% formaldehyde before lysis and immunoprecipitation with NEDD9 antibody, followed by digestion and LC-MS analysis. **(b)** Western blot demonstrating the detection of NEDD9 following immunoprecipitation in MDA-MB-231(231) and HCC1954 (HCC). IgG controls for non-specific immunoprecipitation. **(c)** Fold change in the most abundant interactors for MDA-MB-231 TGFβ vs. TGFβ inhibitor (SB) **(d)** Fold change in the most abundant interactors for HCC1954 TGFβ vs. inhibitor (SB) and HCC1954 TGFβ vs. inhibitor (SB). Detection of PKM2 correlates with the TGFβ condition (L_231N) in MDA-MB-231 **(e)**, whereas NEDD9 does not in HCC1954 **(f)** .

NEDD9 was perturbed (section 2.7). The peptide with the highest fold change value was Pyruvate kinase isozyme M1/M2 (PKM2), also identified as a NEDD9 interacting protein in the phospho-proteome analysis (section 2.5). PKM2 is an isoenzyme of the glycolytic enzyme pyruvate kinase which catalyses the final step of glycolysis, the dephosphorylation of phosphoenolpyruvate (PEP) to pyruvate. However, it has also been shown to stimulate POU5F1/OCT4-mediated transcriptional activation in the control of the expression of several pluripotency genes (Qin et al. 2017). This finding could indicate that the association of NEDD9 with PKM2 upon activation of the TGFβ signalling pathway may be regulating metabolic processes involved in carcinogenesis and/or mediating the expression of genes involved in self-renewal. In the parallel analysis for HCC1954 (figure **3.1d**), the majority of proteins with the highest fold changes were keratins, likely to be due to environmental contamination. The issue of laboratory environmental contamination in MS analysis is an enduring and well-studied problem. The most common contaminants detected are keratins from hair and skin, and contaminants from reagents such as trypsin. In 2013 Hodge and colleagues published a list of the most common environmental contaminants identified in HPLC-MSMS data, alongside a series of recommendations to maintain good sample preparation (Hodge et al. 2013). Top of their list of environmental contaminants in the analysis of human samples were Keratins, type I and type II. In the analysis of HCC1954, the interactor with the highest fold change is Vimentin, a marker of epithelial to mesenchymal transition (EMT). Previously, we demonstrated an increase in vimentin expression occurs in both the MDA-MB-231 and HCC1954 cell lines in response to TGFβ, establishing that the responses in self-renewal and BTIC numbers seen in both cell lines are not simply due to the process of EMT (Bruna et al. 2012). The presence of histones in the analysis also suggests that nuclear NEDD9 may be either directly or indirectly bound to histones in the HCC1954 cell line. Also significantly enriched in response to TGFβ treatment in the HCC1954 cell line was Heat shock cognate 71 kDa protein (HSPA8), which is a repressor of SMAD-mediated transcriptional activation through inhibition of the SMAD co-activator CITED1. CITED1 is a transcriptional co-activator that enhances SMAD-mediated transcription by strengthening the functional link between SMADS and the p300/CBP transcription coactivator complex. Interestingly, CITED1 was also identified as a NEDD9 interactor in the phospho-proteome analysis (section 2.5) of the MDA-MB-231 cell line in the TGFβ treatment condition. The presence of this protein in both cell lines could indicate that CITED1 is functioning as a SMAD transcriptional co-activator in response to TGFβ treatment, but in the HCC1954 cell line, its effects are repressed by HSPA8.

The scatter plots figures **3.1e** and **f** show the reporter ion intensity values for samples from the “light” TGFβ treatment group (on the Y-axis) versus the “Heavy” TGFβ inhibitor group (on the X-axis). Figure **3.1e** confirms that PKM2 detection (yellow dots) is increased in the “light” TGFβ treatment group, with respect to all other samples (purple dots). This correlation with TGFβ treatment is not observed when we

look at the NEDD9 detection in the HCC1954 cell line. Here, the detection of NEDD9 (yellow dots) falls within the detection range for all other samples (purple dots) and, therefore, does not correlate with the treatment group (figure **3.1f**). This observation confirms, as previously demonstrated (figure **2.1**), that in the HCC1954 cell line, NEDD9 expression, although high, is not further induced by TGFβ treatment.

3.3 Quantitative analysis of NEDD9 interacting proteins

Whilst having a great advantage over non-quantitative methods, the use of SILAC-RIME is limited in the number of samples it is possible to simultaneously analyse. As the number of treatment conditions that can be compared is limited to the number of stably labelled cell lines available, it is generally possible to compare only a maximum of three conditions. Following on from my SILAC-RIME analysis, a new methodology for quantitative MS of endogenous interacting proteins was developed in the CRUK-CI, named qPLEX-RIME (Papachristou et al. 2018). This method integrates RIME as previously described together with the use of tandem mass (also known as isobaric) tags (TMT). TMT are chemical compounds that are used to label proteins for MS analysis and allow for up to 10 different peptide samples to be labelled in parallel and combined for analysis allowing quantitative comparisons to be made between sample groups. All mass tagging reagents within one labelling set have the same chemical structure and nominal mass (isobaric) but contain different combinations of ^{13}C and ^{15}N isotopes which enables them to be distinguished by MS analysis. For this analysis, I used MDA-MB-231 mammosphere cultures treated for three hours with either DMSO control, 100 nM TGFβ, or 100 nM TGFβ plus 10 μM TGFβ inhibitor SB431542. Samples were collected from three subsequent passages of mammosphere cultures to give three biological replicates and all nine samples were independently processed and TMT labelled before being simultaneously analysed by LCMS [methods 4]. The overall workflow is summarised in figure **3.2a**.

MA plots for differential protein detection show that in both the TGFβ versus control (figure **3.2b**) and the TGFβ versus TGFβ inhibitor (figure **3.2c**) conditions, NEDD9 was significantly enriched by TGFβ treatment. This result is further confirmation that TGFβ treatment induces NEDD9 expression in MDA-MB-231 mammosphere cultures. There were no peptides with significant changes in expression in the control versus TGFβ inhibitor condition. NEDD9 interacting proteins identified with significantly increased fold change values in both the TGFβ versus control, and TGFβ versus TGFβ inhibitor conditions were Hornerin (HRNR) and Arginase 1 (ARG1), and Catalase (CAT). Hornerin is a member of the S-100 fused protein family and has been demonstrated to be involved in the regulation of protein

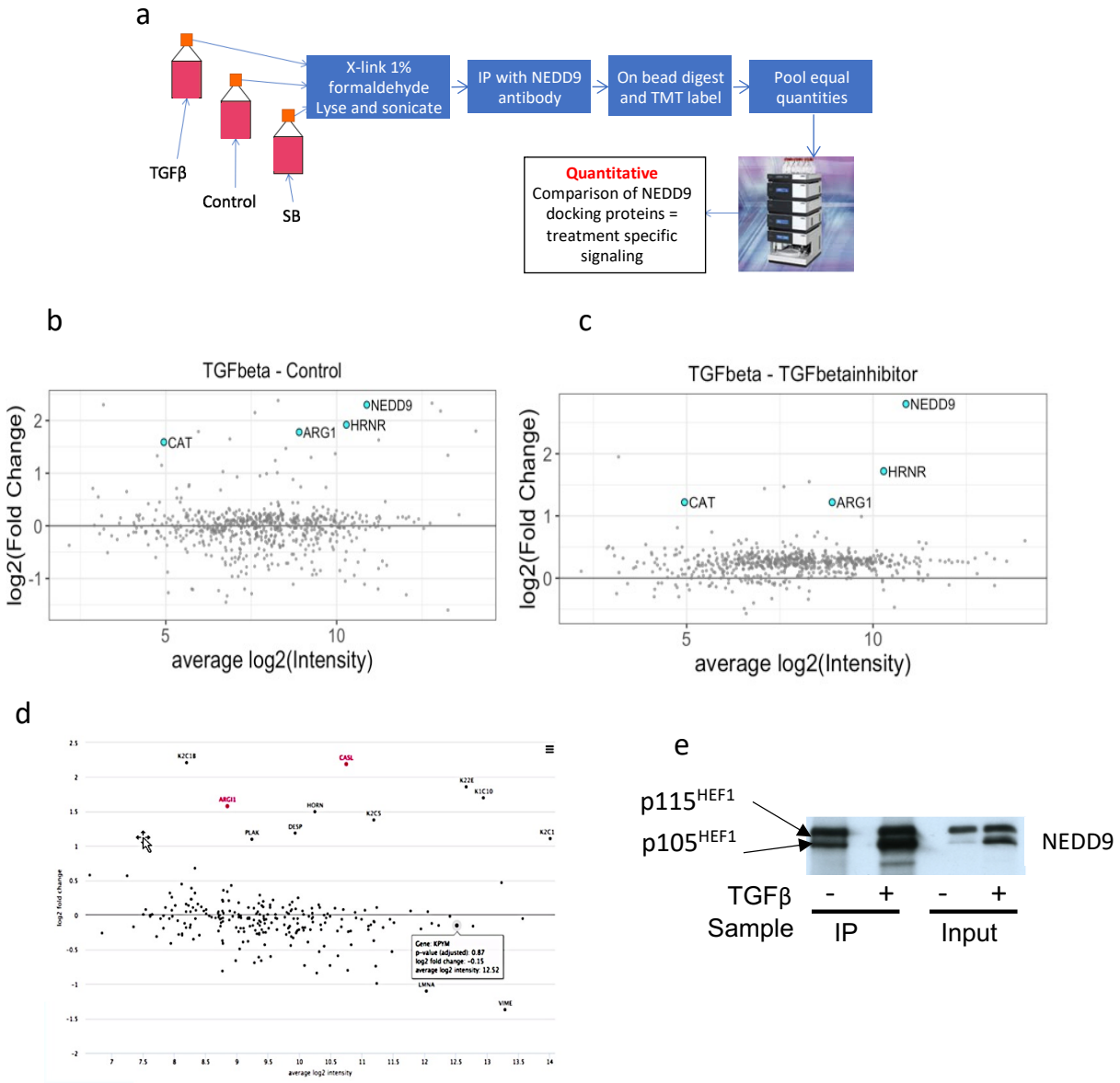


Figure 3.2. Quantitative identification of NEDD9 docking proteins in response to TGFβ treatment by TMT-RIME LCMS

(**a**) Schematic demonstrating the workflow for quantitative TMT-RIME identification of NEDD9 interacting proteins in MDA-MB-231 mammosphere cell cultures. Suspension cultures were treated for seven days with either DMSO control, 100 nM TGFβ, or 100 nM TGFβ plus 10 μM TGFβ inhibitor SB431542 (SB). Treated cultures were cross-linked with 1% formaldehyde before lysis and immunoprecipitation with a NEDD9 antibody followed by digestion and TMT labelling before LC-MS analysis. Differential expression MA plots demonstrating log fold change versus signal intensity in the TGFβ versus control (**b**) and TGFβ versus TGFβ inhibitor conditions (**c**). Proteins highlighted in blue show significant changes in expression. There are no significant proteins in the TGFβ inhibitor versus control condition. (**d**) Expanded MA plot for TGFβ versus control condition to show position of PKM2. (**e**) Western blot illustrating PKM2 immunoprecipitation probed for NEDD9 from seven day mammospheres treated with either 100 nM TGFβ or no treatment control. Samples were loaded from both the IP and input.

phosphorylation and also in the processes of cell proliferation, differentiation and apoptosis. S-100 family members have been implicated in many different cancer types and Hornerin was recently reported to be involved in breast cancer progression. Here the authors describe an increased expression of Hornerin in patients with ER-negative and metastatic disease, concluding that hornerin is involved in the processes of malignant transformation and breast cancer progression (Choi et al. 2016). Arginase 1 is involved in the regulation of cellular metabolism and has been implicated in cancer. Arginase 1 is a key element in the urea cycle, converting L-arginine to urea and L-ornithine and the subsequent generation of proline and polyamines both of which are critical for cell proliferation. Because of this important role, arginine metabolism has become an increasingly compelling target for anti-cancer therapy. Arginine is a conditionally essential amino acid that can be synthesised in the body for the purposes of basal metabolic demands, however, in times of stress the basal synthesis cannot provide sufficient amounts and exogenous sources are necessary. Some tumour types have been shown to lack expression of argininosuccinate synthase (AS), a key enzyme of the urea cycle making them severely sensitive to arginine deprivation. It is hypothesised that these tumour types would make excellent candidates for arginine deprivation therapy (Kremer et al. 2017; Zou et al. 2019). Catalase is an antioxidant enzyme involved in protecting cells from oxidative damage caused by reactive oxygen species (ROS). The down-regulation of catalase is frequently observed in cancer and has been implicated in resistance to chemotherapy, which is making it an increasingly interesting target for cancer therapy (Glorieux and Calderon 2017).

Interestingly, in the TGFβ versus control condition, PKM2 (KPYM, annotated in the box) was significantly identified (figure **3.2d)**, being one of the proteins with the highest number of peptides detected, however, there was no fold change difference in response to TGFβ treatment. This suggests that PKM2 is consistently associated with NEDD9, but contrary to the results of the previous analysis (section 3.2), the levels of bound PKM2 do not increase in response to TGFβ treatment. Because this analysis is quantitative, compared to the previous semi-quantitative method, this finding is likely to be a more accurate representation of PKM2/NEDD9 interactions. Also detected with high confidence, but not dependent on TGFβ treatment, were the Breast cancer anti-oestrogen resistance proteins 1 and 3 (BCAR1/BCAR3). As previously discussed, these proteins are other CAS protein family members known to heterodimerise with NEDD9 and were frequently detected in my previous analyses. I have consistently observed the expression of both PKM2 and BCAR1 is upregulated in response to TGFβ treatment, however, this quantitative MS experiment and in the RNA-seq analysis for BCAR1 expression (figure **2.8e**) indicate this is not the case. As we have repeatedly shown, NEDD9 expression is induced by TGFβ treatment and therefore, apparent increases in NEDD9 partner proteins may be simply due to

stoichiometry. It is only when these interactions are investigated in a fully quantitative manner that this becomes apparent.

The analyses performed so far consistently demonstrate an association between NEDD9 and PKM2 in both whole cell lysates and NEDD9 immunoprecipitates. To conclude this chapter, I performed a contrasting experiment where I immunoprecipitated PKM2 from mammosphere cultures of MDA-MB-231 cell lines treated for 7 days with or without 100 nM TGFβ. The resulting cell lysates were run as Western blots and probed for NEDD9, using a sample of the input (whole-cell lysis) as a control to confirm the protein was present in the originating sample (figure **3.2e)**. NEDD9 was present in the input sample and was increased in abundance in the IP sample, confirming an association between the two proteins.

3.4 Summary

In this chapter, I have evaluated, first by semi-quantitative methods and then using quantitative methods to confirm, the proteins which commonly interact with NEDD9 in the MDA-MB-231 mammosphere culture system in response to TGFβ signalling. These results point towards a significant association between NEDD9 and PKM2. The presence of other proteins involved in metabolic pathways and the knowledge that metabolic reprogramming is a key hallmark of cancer suggest that altered metabolism may be a mechanism by which TGFβ is regulating claudin-low breast cancer stem cells through NEDD9. As discussed, the key finding by quantitative analysis of NEDD9 interacting proteins demonstrates that PKM2 expression is not increased by TGFβ treatment, however, PKM2 is consistently associated with NEDD9. This suggests that rather than TGFβ regulating PKM2 activity through increased expression, the interaction between NEDD9 and PKM2 may be key in regulating PKM2 function. The association between NEDD9 and Arginase 1 is also an interesting finding as in 2017, Kremer and colleagues describe a link between arginine metabolism and PKM2 expression in the regulation of Warburg metabolism in cancer (Kremer et al. 2017).

In the next chapter, I further characterise the known PKM2 cellular effects in the MDA-MB-231 mammosphere cell culture system in response to TGFβ signalling.

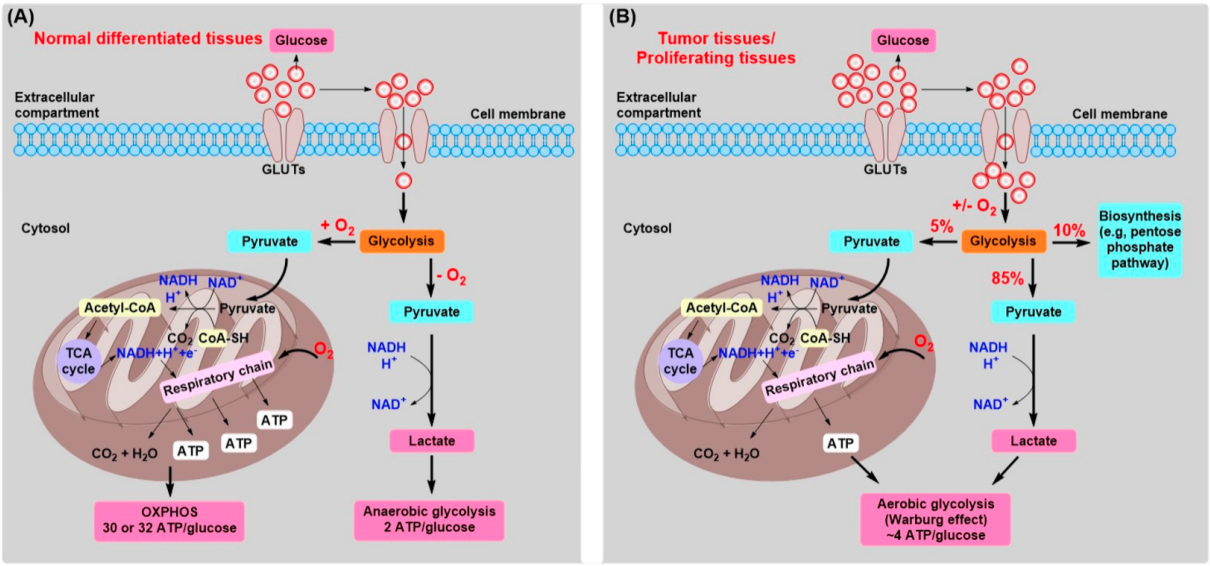
4. The role of PKM2 in Claudin-low breast cancer stem cell dynamics

4.1 Preface

In the previous chapter, I identified the metabolic enzyme PKM2 as a key NEDD9 interactor in breast tumour initiating cell (BTIC) enriched mammosphere cultures of the MDA-MB-231 cell line, a cell line representative of TGFβ oncogenic effects in Claudin-low breast cancer. As metabolic reprogramming is one of the most common hallmarks of cancer, it is not improbable that this may be one of the mechanisms by which NEDD9 mediates breast cancer stem cell dynamics in Claudin-low breast cancer in response to TGFβ signalling. As described in chapter 3, levels of PKM2 expression are not dependent upon TGFβ treatment. However, I hypothesise that in the TGFβ condition NEDD9 is orchestrating a different set of signalling pathways, and here the interaction between NEDD9 and PKM2 may be regulating different processes to those in the absence of TGFβ signalling. To investigate this hypothesis, in this chapter I aimed to explore the known cellular effects of PKM2 and to clarify its role in regulating TGFβ induced BTIC dynamics in MDA-MB-231.

4.2 Metabolic reprogramming and cancer

In 1927 the German physiologist and clinician Otto Heinrich Warburg published a paper describing how the metabolism of cancer cells differed from that of normal cells in that they were dependent upon glycolysis rather than oxidative phosphorylation to generate energy (Warburg, Wind, and Negelein 1927). This process of metabolic reprogramming termed the “Warburg effect” or aerobic glycolysis was later revealed to be one of the most common characteristics of cancer cells (Hanahan and Weinberg 2011). Compared to oxidative phosphorylation, the process of glycolysis is far less efficient in producing energy in the form of adenosine triphosphate (ATP). Cancer cells compensate for this loss in net ATP production by upregulating genes encoding glycolytic enzymes and membrane glucose transporters, resulting in an increase in overall glucose metabolism and allowing cells to survive in a hypoxic environment. This switch to aerobic glycolysis results in the increased production of intracellular intermediates and precursors of proteins and nucleotides, all of which are essential for cell proliferation (Vander Heiden, Cantley, and Thompson 2009). The transition from normal metabolism to Warburg metabolism in cancer cells is summarised in the below illustration:



The transition to Warburg metabolism in cancer cells (Fan et al. 2019). **Panel A describes the metabolism of glucose in normal cells and ane B describes how the glucose metabolic process is altered in cancer.**

This increased metabolism, present in many cancer types, has been harnessed for use in the medical imaging of cancer to both detect and monitor the presence of tumours. The radioactive glucose analogue Fluorodeoxyglucose (^{18}F-FDG) is used in patients undergoing positron emission tomography (PET) imaging to visualise increased glucose uptake and metabolism in tumours.

4.2.1 Pyruvate kinase isozyme M2 (PKM2)

Pyruvate kinase isozyme M2 (PKM2) is an isoenzyme of pyruvate kinase which catalyses the final, irreversible ATP generating step in glycolysis, the dephosphorylation of phosphoenolpyruvate (PEP) to pyruvate. The PKM gene encodes two isoforms, PKM1 and PMK2, which differ based on alternative splicing of exons 9 and 10, respectively. The PKM1 isoform is constitutively active and is expressed in terminally differentiated tissues with high energy demand, for example, brain and muscle. PKM2, however, is subject to allosteric regulation and is expressed in tissues undergoing anabolic processes such as proliferating cells in embryogenesis or aberrant cellular proliferation such as cancer. This regulation, which allows cells to switch between high and low catalytic activity states, enables cancer cells to alter their metabolic activity in response to environmental cues. The alternative splicing of the PKM gene leading to overexpression of PKM2 in cancer cells is regulated by the oncoprotein c-Myc. C-Myc is an

oncogenic transcription factor that causes the upregulation of heterogeneous nuclear ribonucleoprotein (hnRNP) proteins which in turn, leads to PKM exon 10 inclusion and expression of the PKM2 isoform (David et al. 2010). The preferential expression of PKM2 in cancer allows cells to rapidly replicate without accumulation of excessive intracellular reactive oxygen species (ROS), which would otherwise be detrimental to growth. ROS are produced in the mitochondria through the process of oxidative phosphorylation, and in normal cells, the production of ROS is tightly regulated (Perillo et al. 2020). However, in cancer cells, moderate levels of ROS activate cell survival processes including the mitogen-activated protein kinase (MAPK) and phosphoinositide-3-kinase (PIK3) pathways, at higher concentrations, excessive levels of ROS can trigger apoptosis. An additional mechanism for the regulation of ROS generation in cancer cells is the production of Reduced Nicotinamide Adenine Dinucleotide Phosphate (NADPH), which is generated as a result of PKM2 mediated promotion of oxidative pentose phosphate pathway (PPP) activation and acts to quench ROS.

4.2.2 PKM2 regulation of metabolic activity

The regulation of PKM2 kinase activity is controlled by switching between dimeric and tetrameric forms in response to environmental cues. In the tetrameric form, PKM2 demonstrates high catalytic activity associated with catabolic metabolism and the production of ATP through oxidative phosphorylation. However, in the lower activity dimeric state where the conversion of PEP to pyruvate is slowed, metabolites of glucose are directed towards the production of glycolytic intermediates, macromolecule precursors and nucleotide synthesis via several pathways including the pentose phosphate pathway (PPP) (Anastasiou et al. 2012; Cairns, Harris, and Mak 2011). Many mechanisms have been described for the regulation of PKM2 dimer-tetramer dynamics. Post-translational modifications, including phosphorylation, acetylation and oxidation have been shown to mediate the formation of the less active dimer (Christofk et al. 2008), whereas binding of FBP (a glycolytic intermediate), Phosphoribosylaminoimidazolesuccinocarboxamide (SAICAR) and serine mediate the opposite, tetramer formation and increased catalytic activity (Chaneton et al. 2012; Keller, Tan, and Lee 2012). Phosphorylation of PKM2 at tyrosine-105 has been shown to cause FBP to be released from PKM2 binding resulting in the switch from the active tetramer to the less active dimeric form. The processes involved in the regulation of dimer-tetramer dynamics are outlined in figure **4.1**.

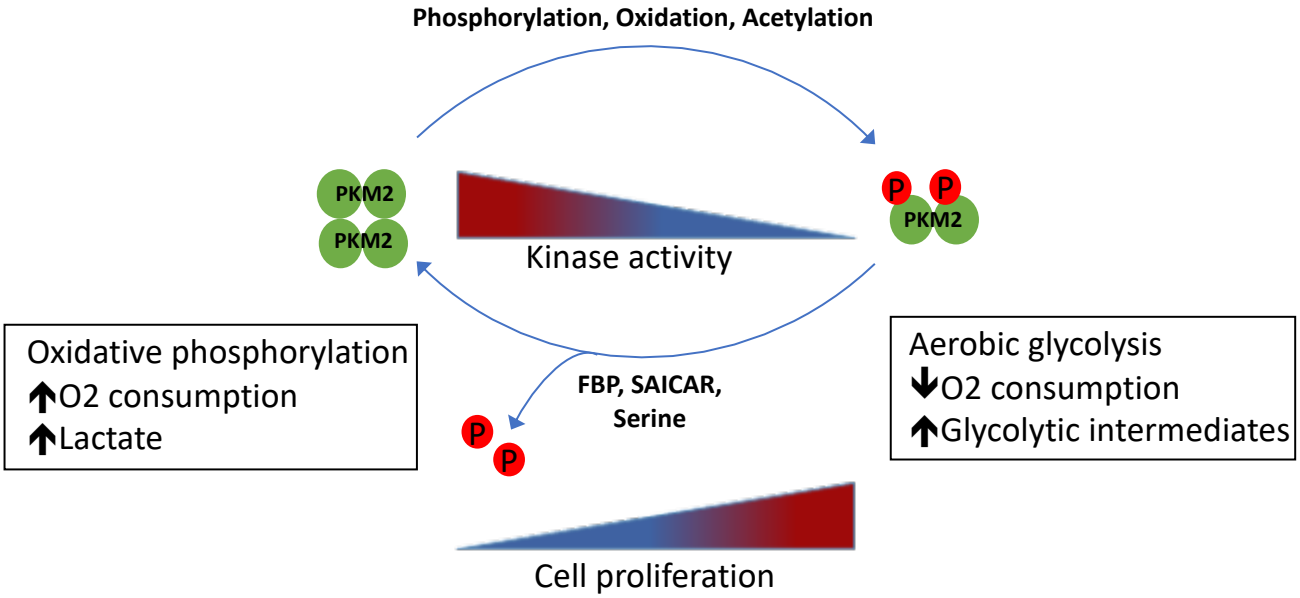


Figure 4.1. The regulation of glucose metabolism through PKM2 dimer-tetramer dynamics
In cancer cells, PKM2 exists in either the the high kinase activity tetrameric form or the low kinase activity dimeric form. In the high catalytic tetrameric form catabolic metabolism and the production of ATP through oxidative phosphorylation occurs. In the lower activity dimeric form, conversion of PEP to lactate is slowed and the accumulation of glycolytic intermediates occurs. Conversion of PKM2 tetramer form to dimer form is mainly regulated by post-translational modification, whereas conversion of the dimeric to tetramer is form is generally regulated by binding to PKM2 interacting proteins proteins such as FBP and SAICAR. Using these mechanisms cancer cells can rapidly switch between PKM2 activity states in response to environmental cues.

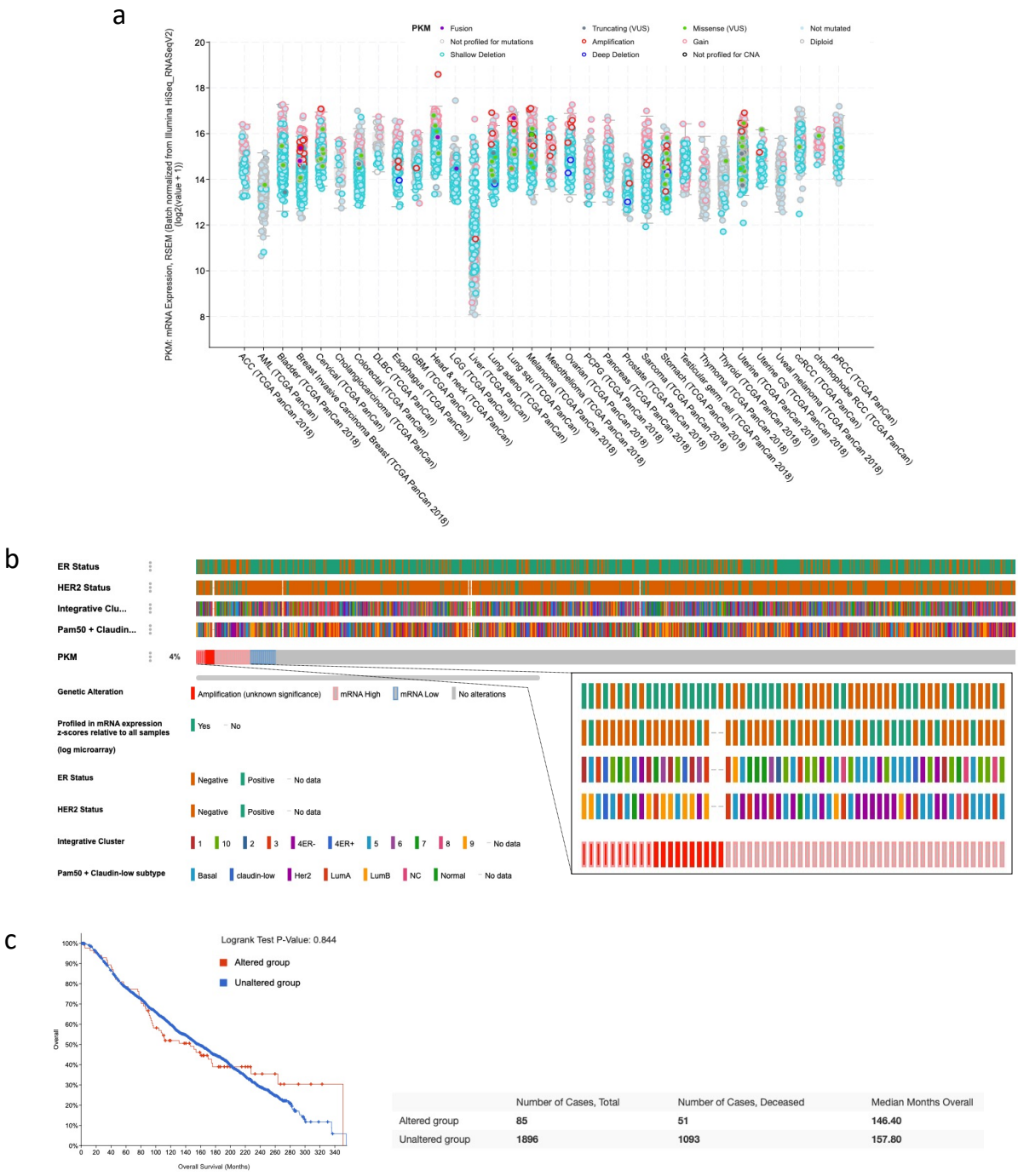


Figure 4.2. PKM expression in cancer

PKM expression by RNA-Seq (Log2 transformed) from the TCGA PanCancer Atlas (cbioportal.com) across multiple cancer types, including mutations and types where profiled. **(b)** PKM mutational profile of 2509 breast cancers from the METABRIC clinical cohort, including intrinsic subtype and Integrative cluster annotation. Zoomed insert illustrates annotation details for the 4% of patients with a PKM amplification and/or high mRNA expression. **(c)** Kaplan–Meier overall survival estimates for patients profiled for PKM alterations (red) versus the unaltered group (blue).

4.3 PKM2 gene expression in cancer

PKM2 is commonly overexpressed in many cancer types, as is evident in the TCGA PanCancer Atlas at cBioPortal (www.cbioportal.org) (Gao et al. 2013; Cerami et al. 2012). Figure **4.2a** details RSEM normalised (B. Li and Dewey 2011) relative expression levels of the PKM gene and detected mutations across a wide range of cancer types. The data were not stratified by the expression of either PKM1 or PKM2 but by the overall expression of the PKM gene. However, due to the preferential expression of PKM2 in cancer cells, the data likely represents PKM2 expression. The level of PKM expression was observed to be consistently raised across a range of cancer types. More in-depth analysis of the expression and mutational profile of PKM in the METABRIC breast cancer cohort (figure **4.2b**) reveals that PKM amplification demonstrates an increased correlation with ER-positive HER2-negative disease (14/20 and 15/18 respectively). However, in patients with high PKM expression with no genetic alterations, this was not the case, here more cases were ER-negative (23/39) and HER2-negative (25/39). Samples from patients with PKM amplifications correlate more closely with the better prognosis Luminal A and luminal B PAM50 subtypes (9/18) than the poorer prognosis basal and Claudin-low subtypes (4/18 and 1/18 respectively). Kaplan–Meier estimates (figure **4.2c**) on patients with PKM copy number amplifications show no significant effect on overall survival, with a median survival of 146.4 months in the altered group compared with 157.8 months in the unaltered group.

These data are contrary to those reported by others and suggest that like NEDD9, PKM2 gene amplification or mRNA overexpression alone is not sufficient to explain worse overall survival. These results indicate that in the MDA-MB-231 mammosphere model of Claudin-low breast cancer, the TGFβ dependent interaction between NEDD9 and PKM2 may be regulating PKM2 metabolic effects.

4.4 PKM2 protein expression in response to TGFβ treatment

Before investigating PKM2 cellular effects in the MDA-MB-231 cell line in response to TGFβ signalling, it was first necessary to understand PKM2 expression in this cell line. To assess expression, whole-cell extracts of MDA-MB-231 mammosphere cultures where NEDD9 or PKM expression was perturbed using commercial siRNA were treated for three hours with either vehicle control (DMSO), 100 nM TGFβ or 100 nM TGFβ plus 10 μM TGFβ inhibitor SB431542. PKM1 and PKM2 differ by only 23 amino acids, so the siRNA used in this analysis targeted both PKM1 and PKM2. The cell lysates were processed for Western blotting [methods 11] and probed for NEDD9, PKM1, PKM2 and two different

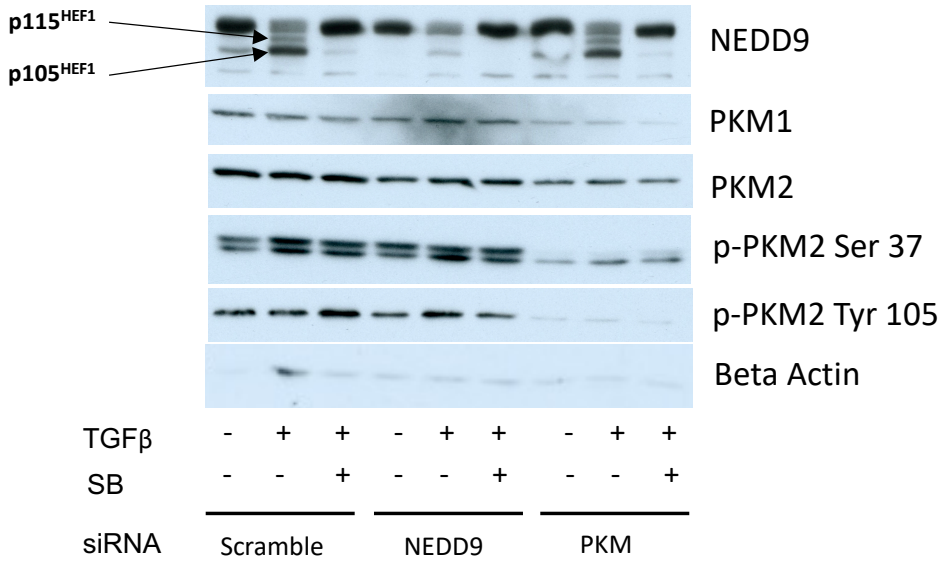


Figure 4.3. PKM2 expression in response to TGFβ treatment
Western blot of whole cell protein lysates from mammosphere cultures of MDA-MB-231. Cultures were treated for seven days +/- 100 nM TGFβ. Expression of NEDD9 or PKM was perturbed by use of a commercial SiRNA with a scramble SiRNA as a control. Cultures were treated for seven days +/- 100 nM TGFβ or 100 nM TGFβ plus 10 μM TGFβ inhibitor SB431542 10 μM (SB). The membrane was probed for NEDD9, PKM1 and PKM2 protein expression, and the expression of the two phosphorylated forms of PKM2, p-PKM2 Ser37, p-PKM2 Tyr105. Beta Actin expression was used as a loading control. The position of the NEDD9 phospho-forms $p105^{HEF}$ and $p115^{HEF}$ is located as indicated by arrows. Western blot figure is representative of at least three biological replicates.

phosphorylated forms of PKM2 (figure **4.3**). Reduction in expression of both the $p105^{HEF1}$ and $p115^{HEF1}$ NEDD9 phospho-forms was achieved by the use of the NEDD9 siRNA treatment as previously described. PKM siRNA treatment reduced the expression of both PKM1 and PKM2, but PKM expression levels were not altered by TGFβ treatment. PKM2 expression was marginally decreased in response to NEDD9 siRNA treatment, but PKM1 expression was not affected. This suggests that PKM2 expression is influenced by NEDD9 abundance, but not in response to TGFβ treatment and agrees with the findings from the TMT-RIME analysis in section 3.3.

Analysis of PKM2 post-translational modification revealed that phosphorylation of PKM2 Tyrosine 105 was decreased by NEDD9 perturbation, mirroring the decrease seen in PKM2 expression when compared to the scramble siRNA control. TGFβ treatment in the NEDD9 siRNA condition resulted in a slight increase in pTyr105. Levels of PKM2 phosphorylation at Serine 37 in the NEDD9 siRNA condition remained largely unchanged, and as expected, levels of both phospho-forms were reduced in the PKM2 siRNA condition. Both the Tyr105 and Ser37 phosphorylations have been highlighted as important regulators of metabolism in cancer (Z. Zhang et al. 2019). Phosphorylation of PKM2 at Tyr105 regulates PKM2 catalytic activity. PKM2 Ser37 is directly phosphorylated by ERK1/2 which mediates nuclear translocation and expression of Oct-4 mediated pluripotency genes, and c-Myc expression resulting in the upregulation of metabolic genes (Yang et al. 2012). As activation of the ERK pathway is known to occur in response to TGFβ signalling, these results suggest that ERK-mediated phosphorylation of PKM2 may be occurring. However, my previous RNA-seq analysis (chapter 2.7) indicated that activation of ERK signalling was not dependent upon NEDD9.

These results suggest that both NEDD9 abundance and TGFβ treatment are involved in the regulation of PKM2 dimer-tetramer dynamics, the resulting effects on PKM2 catalytic activity and the regulation of glucose metabolism.

4.5 TGFβ induced, NEDD9 dependent changes in metabolic flux

As the previous investigations into PKM2 gene expression and protein expression indicated, in the MDA-MB-231 mammosphere model of Claudin-low breast cancer, a TGFβ dependent interaction between NEDD9 and PKM2 may be regulating PKM2 dependent metabolic effects. Therefore, to investigate any TGFβ induced metabolic changes, I performed a series of analyses starting with a simple ELISA capture a snapshot of cellular metabolism in response to TGFβ treatment.

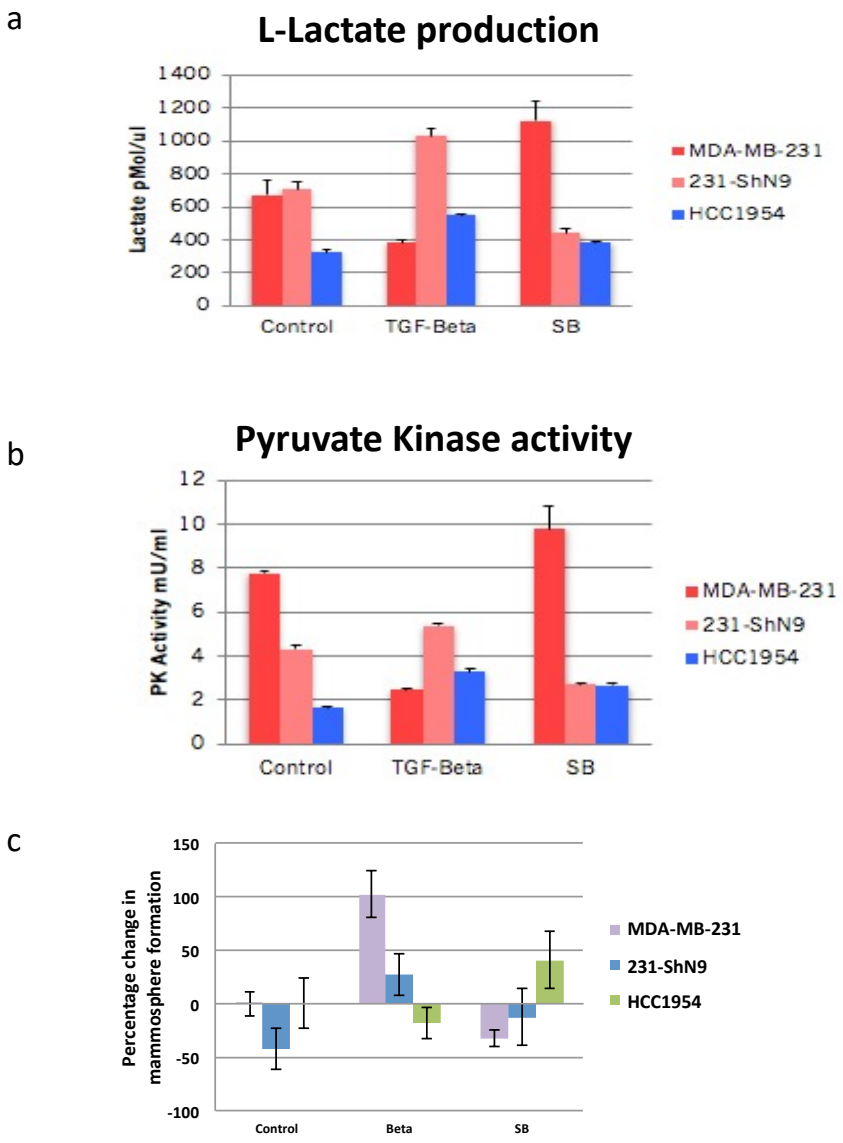


Figure 4.4. End point metabolic assays for glycolysis and pyruvate kinase activity
Seven day suspension mammosphere cultures of cell lines MDA-MB-231 and HCC1954 were treated for 24 hours with TGFβ 100nM, no treatment control or TGFβ inhibitor SB431542 10μM plus TGFβ 100nM (SB). NEDD9 expression was perturbed using a stable knockdown of NEDD9 in MDA-MB-231 expressing a short-hairpin generated against the NEDD9 coding sequence. Cell lysates were collected at day 7 (24 hours after treatment) and assayed for L-Lactate concentration (**a**) and pyruvate kinase activity (**b**). (**c**) Second generation mammosphere assays of cell lines MDA-MB-231 and HCC1954 treated at M1 with 100nM TGFβ, 100nM TGFβ plus 10μM TGFβ inhibitor SB431542 (SB) or no treatment control. NEDD9 expression was perturbed by stable knockdown of NEDD9 in MDA-MB-231 expressing a short-hairpin generated against the NEDD9 coding sequence. Error bars represent ±s.d. of at least three independent experiments for all three assays.

I then moved on to more complex real-time metabolic assays, and finally assays to track metabolic flux using ^{13}C labelled substrates.

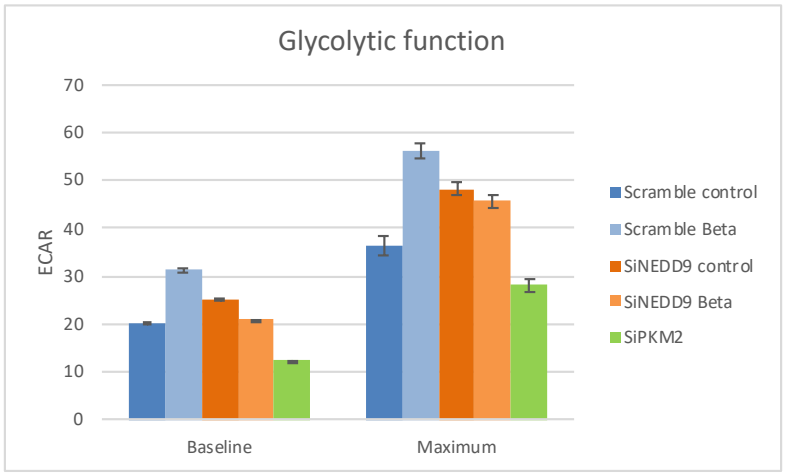
4.5.1 ELISA assays of metabolic profiles

I first performed commercial ELISA assays to measure the levels of L-Lactate, and pyruvate kinase activity (Abcam) as per the manufacturer's recommendations. Seven-day suspension mammosphere cultures from either the Claudin-low cell line MDA-MB-231, or the non-Claudin-low cell line HCC1954 were treated for 24 hours with either 100 nM TGFβ, 100 nM TGFβ plus 10 μM TGFβ inhibitor SB431542, or DMSO control. NEDD9 expression was perturbed in the Claudin-low MDA-MB-231 cell line by incorporation of a stable short-hairpin as previously described. Samples of cell culture medium were collected from each condition following the 24-hour treatment to perform the ELISA. Figure **4.4a** demonstrates that, in the Claudin-low MDA-MB-231 cell line, L-lactate production was reduced by TGFβ treatment but this reduction was reversed where NEDD9 expression was perturbed. In the HCC1954 cell line, the pattern of L-Lactate production mirrored that of the NEDD9 knock-down condition but to a lower extent. This result indicates that in MDA-MB-231 cultures, TGFβ treatment induced a reduction in glycolysis and that this switch was dependent upon NEDD9. The results of the pyruvate kinase ELISA performed on the cell pellets from the same experiment support this hypothesis (figure **4.4b**), as TGFβ treatment induced a reduction in PK activity in the parental MDA-MB-231 cell line and this pattern was again reversed where NEDD9 was perturbed. The results of the second generation mammosphere assay from the same treatment conditions (figure **4.4c**) support the theory that in the MDA-MB-213 Claudin-low cell type, TGFβ treatment induces a switch from catabolic type energy-producing metabolism to anabolic Warburg type metabolism with the synthesis of glycolytic intermediates and increased cell self-renewal. The consistent reversal in these trends in the presence of NEDD9 perturbation indicates that this switch is dependent on NEDD9. In all three experiments, the observation that this trend towards TGFβ/NEDD9 dependent Warburg metabolism, which was not seen in the type B cell line HCC1954, indicates that this effect is specific to the Claudin-low cell line subtype.

4.5.2 Real-time metabolic flux assays

To further investigate these metabolic effects, I then advanced to the use of real-time metabolic flux assays on the Agilent Seahorse analyser. This instrument is capable of analysing the real-time bioenergetics of live cells in a 96 well format. Using label-free sensors, both pericellular pH and oxygen concentration are simultaneously measured as a function of time, and from these measurements, the cellular glycolytic activity and mitochondrial respiration rate are calculated.

a



b

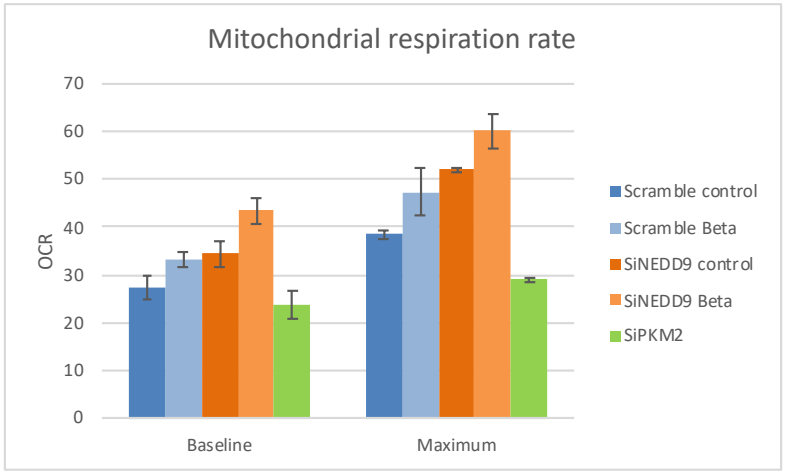


Figure 4.5. Real-time metabolic analysis for glycolytic function and mitochondrial respiration
Two day suspension mammosphere cultures of MDA-MB-231 were treated for 24 hours with TGFβ 100nM or no treatment control. NEDD9 and PKM2 expression were perturbed using a commercial si-RNA directed at NEDD9 or PKM2, knock-down was performed at day 0. Plated cells were assayed at day 3 (24 hours after treatment and 72 hours after knock-down) for glycolytic capacity (**a**) and oxidative phosphorylation (**b**). Error bars represent ±s.d. of at least three independent experiments for both assays.

The extracellular acidification rate (ECAR) is calculated as a read-out of lactic acid and bicarbonate accumulation in the media causing changes in pH and is indicative of glycolytic activity. Similarly, the oxygen consumption rate (OCR) measures mitochondrial respiration by detecting changes in extracellular oxygen levels. To perform these assays, suspension mammosphere cultures of MDA-MB-231 were prepared and commercial siRNA was used to knock down either NEDD9 or PKM2. Cultures were treated with or without 100 nM TGFβ for 24 hours before analysis on the Seahorse instrument. Glycolytic capacity was measured using the Seahorse XF glycolysis stress test kit (Agilent) and oxidative phosphorylation was measured using the Seahorse XF Mito stress test kit (Agilent) as per the manufacturer's recommendations.

Contrary to the results observed in the ELISA assays, the real-time assays indicated that TGFβ treatment increased both basal and maximum glycolytic respiration, and this increase was reduced where NEDD9 was perturbed by siRNA treatment (figure **4.5a**), indicating that the TGFβ induced glycolysis observed was NEDD9 dependent. Analysis of mitochondrial function showed oxidative phosphorylation was not significantly increased in the presence of TGFβ but was increased in the absence of NEDD9 (figure **4.5b**). This indicates that the NEDD9/PKM2 interaction is necessary for glycolysis to occur and by inhibiting it, metabolism is redirected towards oxidative phosphorylation. In these analyses, PKM2 inhibition resulted in a reduction of both glycolytic function and mitochondrial respiration, serving as a positive control for inhibition. Clear from these preliminary assays is that TGFβ treatment is inducing changes in metabolic flux and that these changes are dependent upon the presence of NEDD9. What is not clear, however, is the role the interaction between NEDD9 and PKM2 plays in the regulation of these changes in metabolic flux.

4.5.3 Metabolic flux tracking assays

To further investigate the previously observed changes in metabolic flux in response to TGFβ treatment and NEDD9 perturbation, and to expand the analysis to include the other key metabolic nutrient, glutamine, I performed metabolic tracking experiments in collaboration with Dr Mariia Yueva at the Francis Crick Institute in London. Suspension mammosphere cultures of MDA-MB-231 with or without stable knockdown of NEDD9 by short hairpin were prepared as previously described. After six days in suspension culture, the mammospheres were processed to single cells and seeded into suspension culture in media containing either [U-^{13}C]glucose or [U-^{13}C]L-glutamine, and treated for 24 hours with 100 nM TGFβ, or left untreated as a control. Cell pellets were collected and processed for gas-chromatography mass-spectrometry (GCMS, Agilent 7890A-5975C) [methods 8]. Figure **4.6** demonstrates the percentage

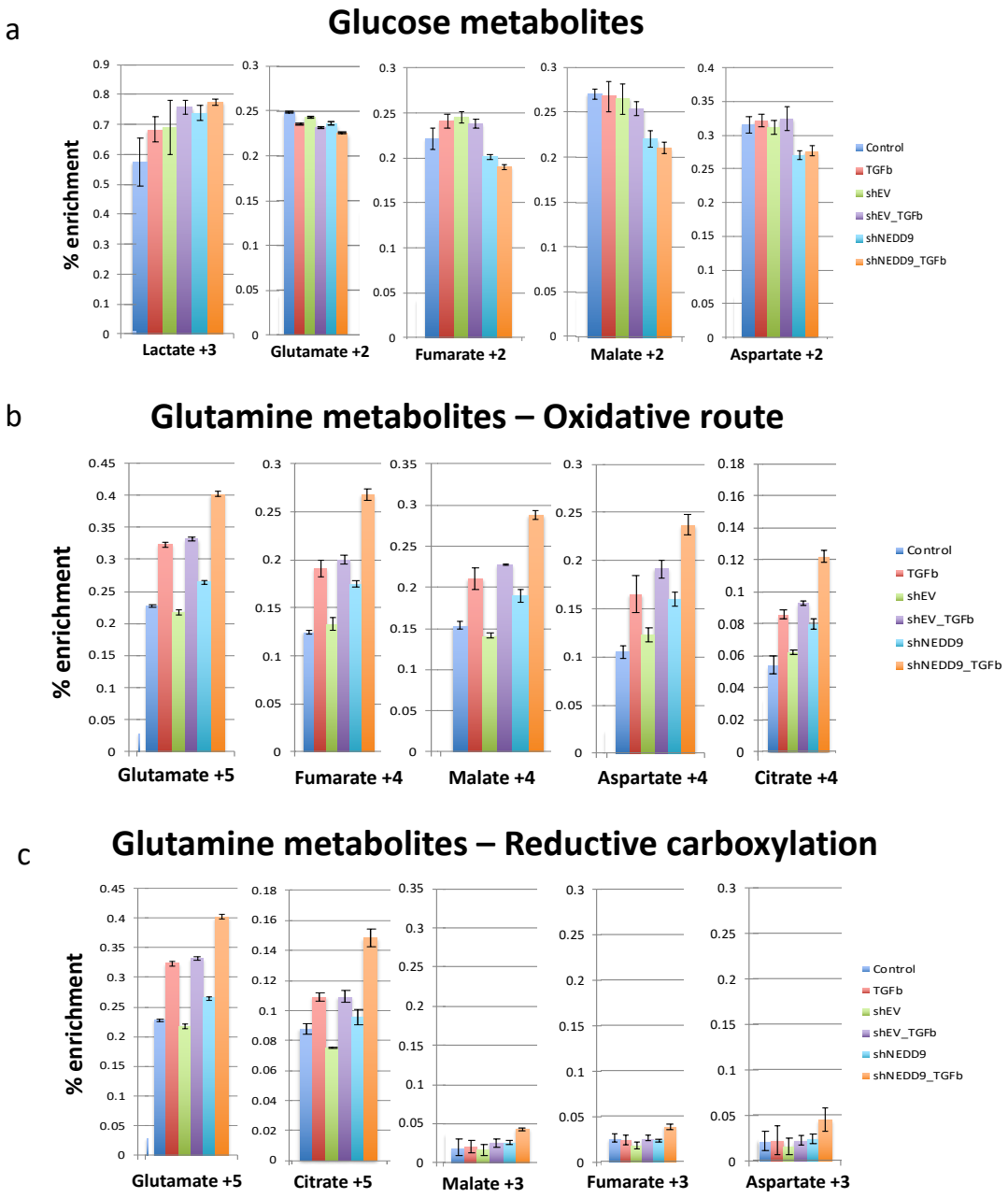


Figure 4.6. Metabolic tracking assays for glucose and glutamine metabolites
Six day suspension mammosphere cultures of MDA-MB-231 were transferred to media containing either [U-13C]-glucose or [U-13C]-glutamine and treated for 24 hours with TGFβ 100nM or no treatment control. NEDD9 expression was perturbed using a stable knockdown of NEDD9 in MDA-MB-231 expressing a short-hairpin generated against the NEDD9 coding sequence (shNEDD9), an MDA-MB-231 cell line with a stable incorporation of the empty vector was used as a negative control (shEV). Cell pellets were collected after 24 hours treatment and processed for GC-MS identification of metabolites of glucose (**a**) and glutamine via the oxidative route (**b**) and glutamine via reductive carboxylation (**c**). Percent enrichment represents the total percentage of each metabolite and related isotopolougues generated from cells containing fully-labelled glucose or glutamine. Error bars represent the SE of three technical replicates.

enrichment of fully labelled metabolites generated from both ^{13}C labelled glucose (figure **4.6a**) and ^{13}C labelled glutamine (figure **4.6b**). Neither TGFβ treatment nor NEDD9 perturbation showed much effect on glucose metabolism (figure **4.6a**). TGFβ treatment slightly decreased the incorporation of glucose into glutamate+2 (the first turn of the Krebs cycle) in all conditions, and NEDD9 perturbation further decreased this incorporation. There was, however, a significant reduction of glucose incorporation into fumarate+2, malate+2 and aspartate in the sh-NEDD9 condition, this could indicate an effect on pyruvate carboxylase conversion of pyruvate derived from glucose to oxaloacetate. These results indicate that NEDD9 is necessary for this step in glucose metabolism, however, because there was no difference observed between the control and TGFβ conditions, it is unlikely that this effect is mediated by TGFβ.

Of more interest were the striking changes observed in glutamine flux (figures **4.6b,c**). In this analysis, both TGFβ treatment and NEDD9 perturbation independently increased the catabolism of glutamine into the tricarboxylic acid (TCA), or Krebs cycle, and together showed an additive effect. This increased flux appears to have occurred via the usual oxidative route (figure **4.6b**), where fully labelled glutamine was metabolised to glutamate+5 and downstream metabolites +4. However, it also appears that glutamine was metabolised via the reverse process of reductive carboxylation (figure **4.6c**), giving rise to citrate+5 and subsequently into malate, fumarate and aspartate+3. Both of these processes occur independently in response to TGFβ treatment and NEDD9 perturbation, but the effect is additive in the presence of both. Reductive carboxylation is a novel mechanism by which cancer cells with metabolic defects redirect the glutamine metabolic pathway to support cell growth. Considering that the activation of this alternative pathway is the most pronounced in the TGFβ plus sh-NEDD9 condition, this indicates that the metabolic response to TGFβ is dependent upon NEDD9. Together these results mirror the effect seen in the Seahorse real-time metabolic experiments where both TGFβ treatment and NEDD9 perturbation increase mitochondrial metabolism.

When considered all together, the results of these functional metabolic analyses indicate that in the Claudin-low cell line, MDA-MB-231, TGFβ treatment increases both glycolysis and oxidative phosphorylation based metabolism, and that perturbation of NEDD9 expression reduces only glycolytic function. The analyses of PK function and glucose metabolism indicate that NEDD9 is important in regulating this process and I hypothesise that the association between NEDD9 and PKM2 may be important in regulating PKM2 dimer-tetramer dynamics, and therefore, its catalytic activity. When NEDD9 expression is reduced, this may in effect release the brake on PKM2 activity allowing tetramer formation and subsequently increased activity in the Krebs cycle of both glucose and glutamine metabolites.

a

Amino acid	Location	Sequence	Modification	Motif prediction
Tyrosine	315 - 317	DA**Y**	[E/D]X**pY**	SHP1 phosphatase substrate motif
Tyrosine	317 - 318	**Y**D	**pY**[A/G/S/T/E/D]	Src kinase substrate motif
Serine	572 - 576	**S**IMNS	**pS**XXX[pS/pT]	MAPKAPK2 kinase substrate motif
Serine	572 - 576	**S**IMNS	**pS**XXXpS*	GSK3 kinase substrate motif
Serine	666 - 668	L**S**Q	X**pS**Q	DNA dependent Protein kinase substrate motif
Serine	667 - 668	**S**Q	**pS**Q	ATM kinase substrate motif

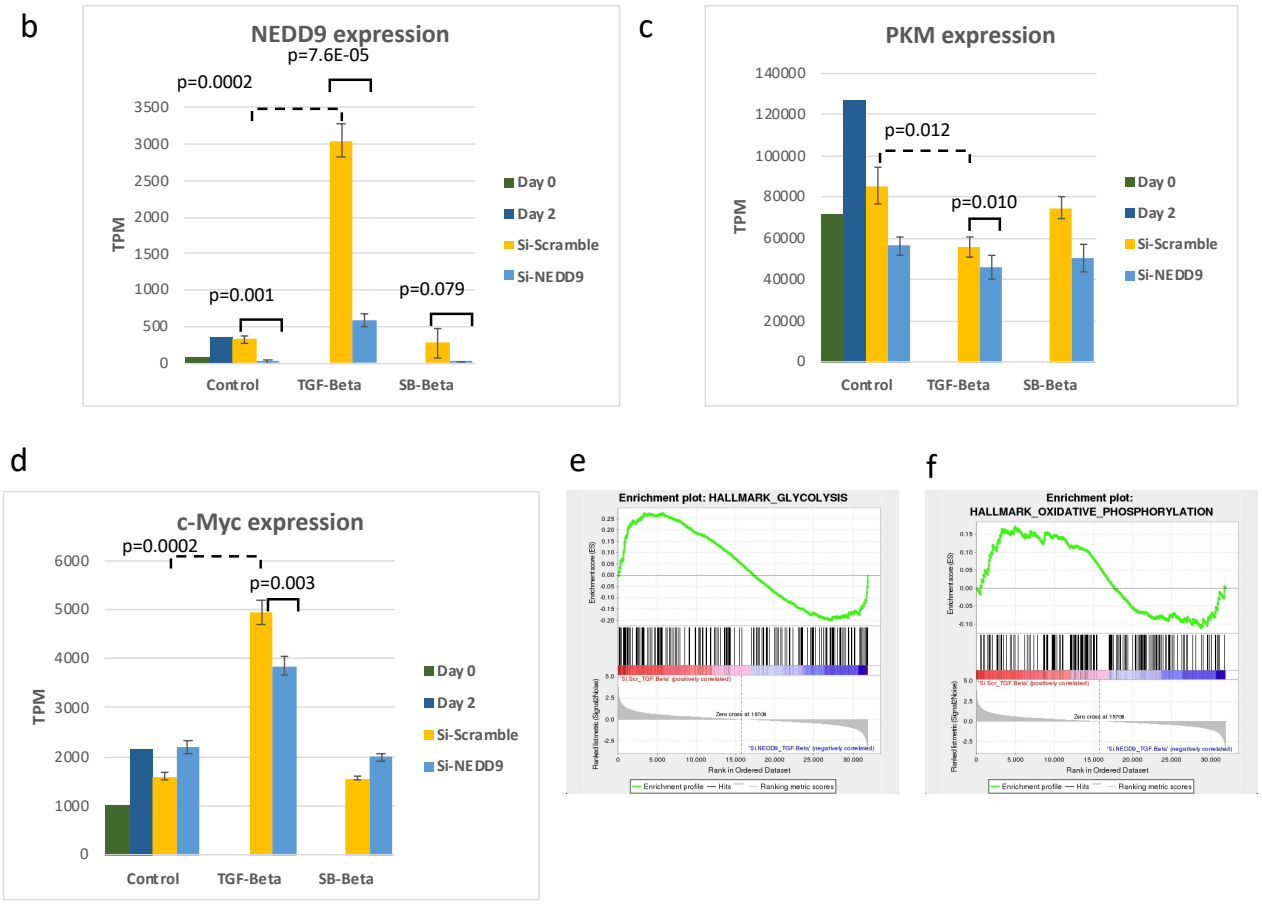


Figure 4.7. Phospho-proteome and RNA-seq analysis for metabolic profiles
NEDD9 phosphorylation modifications identified at high confidence (99-100%) by Ti02 enrichment LCMS specifically phosphorylated in the TGFβ treatment condition only. (**a**) Phopho-motif prediction of these modifications using the PhophoSitePlus® online analysis tool. RNA expression for NEDD9 (**b**), PKM1/2 (**c**) and c-Myc (**d**) in transcripts per million (TPM) from day 2 mammosphere cultures treated for 24 hours with TGFβ 100nM, no treatment control or TGFβ inhibitor SB431542 10μM plus TGFβ 100nM (SB-Beta). NEDD9 expression was perturbed using commercial si-RNA from day 0. Error bars represent ±s.d. of four independent biological replicates. Enrichment plots for correlation with hallmark gene sets from the molecular signatures database (MSigDB, www.gsea-msigdb.org) for (**e**) glycolysis and **(f)** oxidative phosphorylation in the si-Scr TGFβ condition versus the si-NEDD9 TGFβ condition.

4.6 Phospho-motif and RNA-seq prediction for metabolic profiles

Following on from the functional metabolic assays, I returned to the phospho-mass spectrometry and RNAseq data to identify both potential NEDD9 phospho-sites involved in the regulation of metabolic processes, and metabolic gene expression signatures. The NEDD9 phospho-MS analysis identified with high confidence (99-100%) three phosphorylated amino acids present in the TGFβ condition alone; namely pTyrosine317, pSerine572 and pSerine667 (figure **4.7a)**. As described in chapter 2.5, I performed phospho-motif prediction of these modifications using the PhophoSitePlus® online analysis tool. As previously discussed, the pTyr317 residue was identified as a Src kinase substrate motif and associations between NEDD9 Src, and the PI3K and MAPK pathways are well documented (Izumchenko et al. 2009). More recently, a role for nuclear PKM2 in the regulation of gene expression programmes regulating cellular proliferation in glioma was discovered (Yang et al. 2011). Here, the authors describe a mechanism by which EGFR activation leads to PKM2 binding to c-Src phosphorylated β-Catenin leading to translocation to the nucleus and activation of gene expression programmes regulating cellular proliferation via non-canonical Wnt/β-Catenin signalling. However, also included in this gene signature was c-Myc, which as previously mentioned in this chapter is responsible for the preferential expression of PKM2.

Also involved in the regulation of metabolic processes is ATM protein kinase, a member of the PIKK (PI3K-like protein kinase family), a substrate motif for this kinase was identified at Ser667. Additionally, at Ser572 a substrate motif for GSK3 (Glycogen Synthase Kinase) was found which, among one of its many roles, regulates glucose homeostasis. GSK3 is also important in the regulation of canonical Wnt/β-Catenin signalling so could link with the above observation of a NEDD9 Src kinase substrate motif becoming activated in response to TGFβ treatment. These results suggest that when TGFβ signalling is activated, NEDD9 may be coordinating pathways involved in the regulation of metabolic processes; this hypothesis is consistent with the known roles for NEDD9 acting as a signalling hub.

When we returned to the RNA-seq analysis (chapter 2.7), it was first clear that both NEDD9 and PKM2 expression were influenced by both TGFβ treatment and NEDD9 perturbation shown here in transcripts per million (TPM). As shown previously at the protein level, NEDD9 RNA expression is increased by TGFβ treatment (paired two-sample t-test p=0.0002, dashed bracket) (figure **4.7b**) and this expression is significantly reduced, even in the presence of TGFβ treatment by the use of commercial siRNA (paired two-sample t-test p=7.6E-05, solid bracket). When analysing the RNA expression of PKM (figure **4.7c**), it was not possible to distinguish between the PKM1 and PKM2 isoforms due to the very small difference in sequence between the two, but overall the levels mimic those seen in the pyruvate kinase activity ELISA.

TGFβ treatment significantly reduces PKM expression (paired two-sample t-test p=0.012, dashed bracket) and this is further significantly reduced where NEDD9 expression is perturbed by commercial siRNA (paired two-sample t-test p=0.010, solid bracket). Following on from the high confidence identification of NEDD9 Src and GSK3 kinase substrate motifs in the TGFβ treatment condition alone, I also looked in the RNA-seq data at the levels of downstream targets of Wnt/β-Catenin signalling. C-Myc is a target of Wnt/β-Catenin signalling in several different cancer types, including colorectal (Rennoll and Yochum 2015) and basal-like breast cancer (Xu et al. 2016). Analysis of c-Myc expression in my RNA-seq dataset (figure **4.7d**) showed that it was significantly increased by TGFβ treatment (paired two-sample t-test p=0.0002, dashed bracket), and this increase in expression was significantly reduced by the use of NEDD9 siRNA (paired two-sample t-test p=0.003, solid bracket). The c-Myc oncogene governs a host of cellular effects in cancer, however, it is known to be key in regulating both proliferation and metabolism in cancer (Miller et al. 2012), and is known to directly regulate both glucose and glutamine metabolism (Dang 2010). The link between c-Myc and glutamine catabolism is especially interesting due to the observation in the previous chapter that both TGFβ treatment and NEDD9 perturbation demonstrated profound effects on the metabolism of ^{13}C labelled glutamine.

Finally, GSEA analysis from the molecular signatures database (MSigDB, gsea-msigdb.org) for hallmark gene sets involved in metabolic processes revealed that both the glycolysis (figure **4.7e**) and oxidative phosphorylation (figure **4.7f**) hallmark gene sets were enriched in samples from the scramble siRNA TGFβ treatment condition but not in the NEDD9 siRNA TGFβ treatment condition.

Together, this data provides further evidence that NEDD9 is key in regulating metabolic processes induced by TGFβ signalling in the Claudin-low subtype of cell lines, and that for this regulation to occur, the interaction between NEDD9 and PKM2 is necessary. The data also suggest a role for Src and Wnt/β-Catenin signalling, possibly through Wnt/β-Catenin induced c-Myc expression, which as previously discussed is key in the regulation of glucose and glutamine metabolic processes.

4.7 PKM2 transcriptional regulation

As discussed previously, in addition to known effects in the direct regulation of glucose metabolism, PKM2 is also involved in the transcriptional regulation of metabolic gene programs, and the phospho-proteome analysis (chapter 2.5) indicates this may be occurring in the Claudin-low MDA-MB-231 cell line.

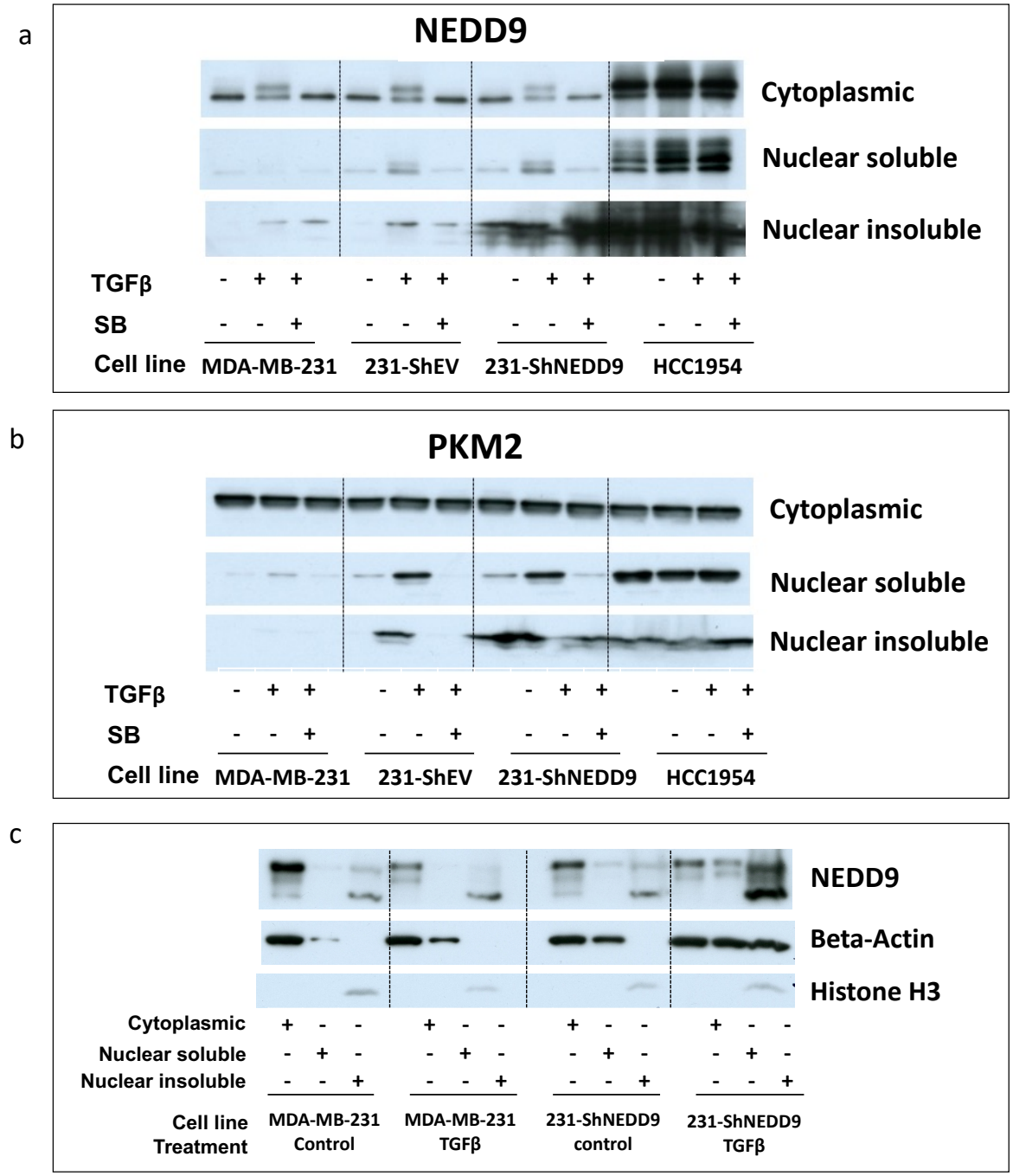


Figure 4.8. Cellular fractionation experiments to identify PKM2 cellular localisation
Seven day suspension mammosphere cultures of cell lines MDA-MB-231 and HCC1954 treated for 24 hours with vehicle control (DMSO), 100 nM TGFβ, or 100 nM TGFβ plus 10μM TGFβ inhibitor SB431542. NEDD9 expression was perturbed using a stable knockdown of NEDD9 in MDA-MB-231 expressing a short-hairpin generated against the NEDD9 coding sequence (231-ShNEDD9) or stable incorporation of the empty vector control (231-ShEV). Cell pellets were collected at day 7 for lysis and fractionation into the cytosolic, nuclear soluble and nuclear insoluble (chromatin) fractions. Lysates from each fraction were then separated using SDS PAGE and probed for NEDD9 (**a**) and PKM2 (**b**). Additionally, samples from each fraction were probed against Beta-Actin and Histone H3 as a control for nuclear soluble and nuclear insoluble/chromatin fractions (**c**). All western blot figures are representative of at least three biological replicates.

To investigate any potential PKM2 nuclear translocation, and whether it was NEDD9 dependent, I performed cellular fractionation experiments [methods 12] by isolating the cytoplasmic, nuclear soluble and nuclear insoluble or chromatin-associated cellular fractions. Seven-day suspension mammosphere cultures of MDA-MB-231, HCC1954 or MDA-MB-231 with stable expression of either a short-hairpin generated against the NEDD9 coding sequence (231-ShNEDD9) or the empty vector control (231-ShEV) [methods 9] were prepared. Cell pellets were collected from cultures treated for 24 hours with either vehicle control (DMSO), 100 nM TGFβ or 100 nM TGFβ plus 10 μM TGFβ inhibitor SB431542. The resulting lysates were separated by SDS PAGE and immunoblotting for NEDD9 and PKM2 was performed. Additional membranes were also probed for Beta-Actin and Histone H3 as positive controls for the cytoplasmic, nuclear soluble and nuclear insoluble fractions.

NEDD9 immunoblotting (figure **4.8a**) revealed that NEDD9 cellular location was predominantly cytoplasmic, with the upper phosphorylation band form induced by TGFβ treatment as previously described. Some nuclear translocation of NEDD9 to the chromatin in the parental MDA-MB-231 cell line was evident in the TGFβ and TGFβ plus inhibitor conditions. Unfortunately, however, the signal in the chromatin fraction for the Sh-NEDD9 was masked by overspill from the strong signal in the neighbouring HCC1954 lane. PKM2 protein expression at the cytoplasmic level remained consistent across all treatment conditions as previously observed (figure **4.8b**), but increased nuclear translocation of PKM2 was observed in response to TGFβ treatment in the parental MDA-MB-231 cell line. However, this TGFβ induced increase in nuclear PKM2 (both soluble and chromatin-associated) was further increased in both cell lines expressing the stable short-hairpin, suggesting this effect may be related to the presence of the short-hairpin. Consistent expression of NEDD9 and PKM2 across all conditions and cellular locations in the HCC1954 cell line demonstrated that the changes in expression and cellular location observed in MDA-MB-231 are phenotypic of the Claudin-low cell type. Additionally, the positive controls (figure **4.8c**) demonstrate that the fractionation experiment was successful, with β-actin and histone H3 being enriched in the cytoplasmic/nuclear soluble and chromatin fractions respectively.

To make any definitive conclusion on PKM2 cellular localisation in the Claudin-low cell line MDA-MB-231 and changes that may occur in response to TGFβ treatment, it would be necessary to repeat this experiment using the commercial siRNA, as used in the majority of previous analyses. The analysis of cellular fractionations in the cell lines containing the short-hairpin (both empty vector control and NEDD9 targeting) indicate that the presence of the short-hairpin alone may be influencing PKM2 cellular localisation. Additionally, the NEDD9 RNA sequence targeted by the short-hairpin is different from that of the commercial siRNA pool, so a direct comparison between the two conditions cannot be

made. However, these preliminary results suggest that in the parental MDA-MB-231 cell line, nuclear translocation of PKM2 was induced by TGFβ treatment.

4.8 Summary

Together, the results in this chapter indicate that, like NEDD9, levels of PKM2 expression are not sufficient to predict worse outcomes in a breast cancer clinical cohort. However, both NEDD9 and PKM2 are key in regulating metabolic cellular processes in the Claudin-low subtype in response to TGFβ signalling. It appears that the interaction between NEDD9 and PKM2 may be significant in modulating PKM2 dimer/tetramer dynamics and, therefore, the balance between anabolic and catabolic metabolism, the formation of glycolytic intermediates and increased self-renewal of cancer stem cells. Although in response to TGFβ signalling NEDD9 mediates a modest increase in glucose metabolism, it is the effect on oxidative phosphorylation and glutamine metabolism that are the most striking. Analysis of NEDD9 post-translational modifications identified kinase substrate motifs for both Src and GSK3, indicating that Wnt/β-Catenin signalling may be important in regulating the metabolic effects observed. Previously published links between PKM2 and c-Myc, alongside the TGFβ dependent increase in c-Myc expression observed in the MDA-MB-231 cell line suggest that c-Myc may also be involved in the regulation of TGFβ induced metabolic changes. Analysis of a nuclear role for PKM2 in the MDA-MB-231 Claudin-low cell line suggested PKM2 nuclear translocation may be occurring in response to TGFβ treatment, but further experiments are necessary to confirm this.

5. NEDD9/TGFβ oncogenic activity as a biomarker in Claudin-low breast cancer

5.1 Preface

In the previous three chapters, I have extensively characterised NEDD9 expression and function in response to TGFβ signalling and the role NEDD9 plays in the regulation of breast cancer stem cell self-renewal in the Claudin-low subtype of breast cancer. I investigated NEDD9 dependent gene expression programs initiated by TGFβ signalling and the functional output of these. I also investigated the regulation and function of the key NEDD9 interacting protein PKM2.

In this final chapter, I aim to investigate the translational application of these findings and whether NEDD9 dependent TGFβ pathway activation is a biomarker for cancer subtype. I will also investigate if any correlation exists between NEDD9 dependent pathways and response or resistance to cancer therapy in *ex-vivo* models. In the preceding chapters, I have demonstrated that the expression levels of NEDD9 and PKM2 alone are insufficient to predict breast cancer stem cell dynamics in response to TGFβ treatment *in vitro*, or patient prognosis in large breast cancer patient cohorts. Therefore, to derive a biomarker of TGFβ oncogenic effects, it is necessary to look at downstream effects of NEDD9 dependent TGFβ signalling and PKM2 activity.

5.2 The CRUK-CI biobank of patient-derived tumour xenograft models

To perform the analyses outlined in this chapter, I used a combination of data from large breast cancer clinical cohorts and experiments using patient-derived tumour xenograft (PDTX) models. As previously discussed in the introduction, the use of PDTX models in both elementary oncology, and translational research is rapidly increasing. Unlike cell line or GEMM, PDTX models retain the intra-tumour heterogeneity, genetic, molecular and histological features seen in the originating tumour even following serial passage making them a more faithful representation of cancer in the patient. Over the past ten years, the Caldas lab has generated a large biobank of PDTX models from both primary and metastatic breast cancer, a project I have been heavily involved in under the supervision of Dr Alejandra Bruna.

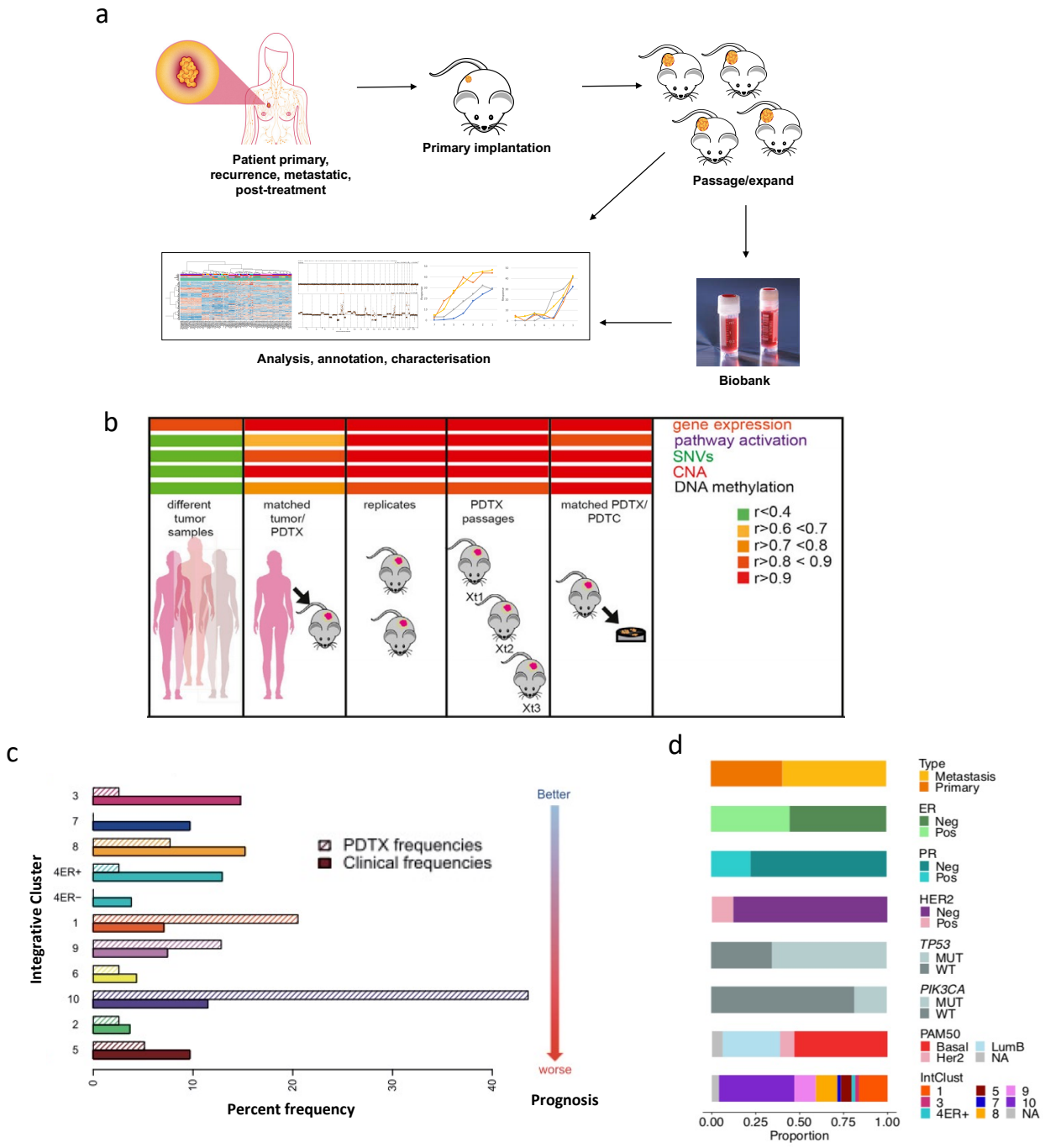


Figure 5.1. The CRUK-CI breast cancer PDTX biobank

(**a**) Schematic demonstrating the workflow for PDTX generation, expansion and downstream application. (**b**) Heatmap or Pearson correlation scores across different molecular data types (Bruna et al. 2016). (**c**) Bar plots showing the IntClust distribution of PDTX models (n = 40; shadowed) and for comparison primary breast cancers from METABRIC (n = 1,980; dense) (Bruna et al. 2016). (**d**) Histological and molecular features of 53 PDTX models profiled by mass cytometry (Calliari M et al. 2021)

As of January 2021, the PDTX biobank comprises 64 models that have been generated in-house and extensively characterised at the molecular and genomic level, and have also undergone analysis of drug response to a panel of commonly used anti-cancer agents. The biobank includes models representing all breast cancer subtypes and most integrative clusters and additionally, contains two pairs of matched primary and metastatic models. We have developed a method to derive short-term *in vitro* cultures (PDTC) from PDTX models for use in high-throughput drug screening for preclinical studies (Bruna et al. 2016). The generation of the PDTX biobank represents a remarkable collaboration between numerous teams both within the CRUK-CI and Cambridge Breast Unit, Cambridge University Hospital NHS Foundation Trust, but first and foremost, the incredible generosity of the patients who donate their tissue. To create the PDTX biobank, surgically resected primary breast cancer tissue, biopsies from metastatic sites including, brain, skin, liver, bone, axilla and lymph node, pleural effusions or ascites were obtained from patients following informed consent. The research was performed under the appropriate approval by the National Research Ethics Service, East of England-Cambridge Central. (A full list of participating clinical trials is outlined in Appendix 1). Tissue segments were encased in Matrigel® matrix (Corning) before implantation into highly immune-compromised NSG mice (NOD.Cg-Prkdcscid Il2rgtm1Wjl/SzJ, www.jax.org) as described (Bruna et al. 2016).

Figure **5.1a** outlines the workflow used to generate the biobank from primary implantation through expansion, biobanking and analysis. As previously mentioned, the established biobank is highly annotated and extensively characterised at both the genomic and molecular level, using a combination of genome sequencing, RNA expression analysis, DNA methylation analysis, immunohistochemistry, mass cytometry, and drug response analysis. Figure **5.1b** (from Bruna et al. 2016) is a heatmap of Pearson correlation scores demonstrating the clear variation seen between tumour samples from different patients. This variation is much reduced however when comparing the originating tumour to the PDTX model, PDTX replicates from the same model passage, different PDTX passages from the same model and PTDX tumours compared to the short-term PTDC cultures generated from them. Figure **5.1c** (also from Bruna et al. 2016) demonstrates the percentage frequency of IntClust subtypes arranged by prognosis, across the 40 PDTX models analysed for this publication. Hatched bars represent the distribution of the PDTX models and solid bars represent the IntClust subtype in 1,980 primary breast cancers from the METABRIC cohort for comparison. Poor prognosis, IntClust 10 tumours are over-represented in the biobank, but most other subtypes are represented. The over-representation of IntClus 10 is likely a reflection of the fact that these tumours grow more quickly and are therefore more likely to engraft in the limited timescale of the lifetime of a mouse. In an attempt to overcome the obstacle of slow implantation rates from less aggressive tumour types, PDTX from slower-growing models that demonstrate some growth at 12 months following primary implantation are routinely re-implanted into a new host.

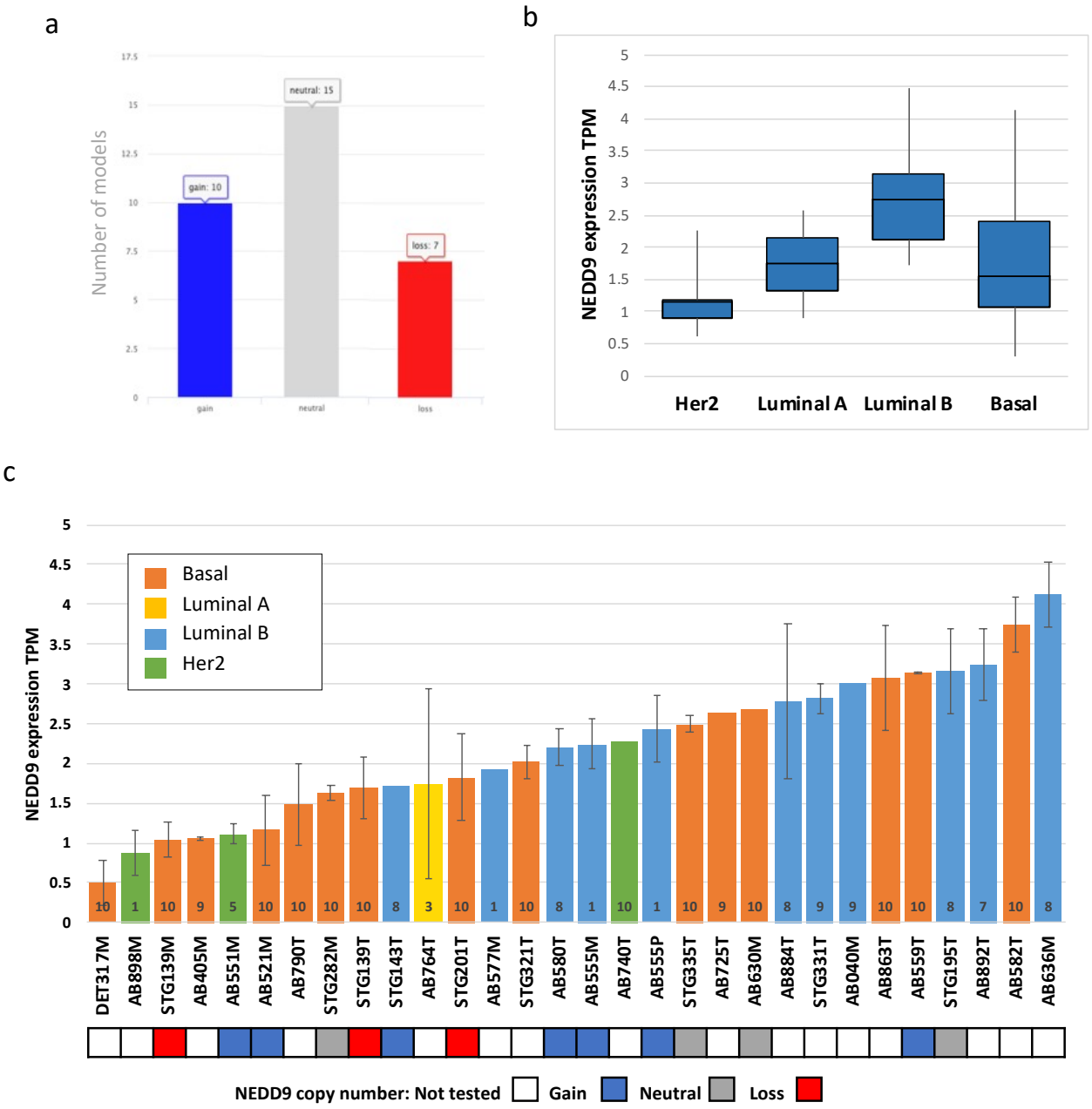


Figure 5.2. NEDD9 copy number and gene expression in the PDTX biobank
(**a**) NEDD9 copy number analysis in BCaPE (http://caldaslab.cruk.cam.ac.uk/bcape), representing number of models that are CN loss, CN neutral and CN gain. (**b**) Box and whisker plot for NEDD9 RNA expression transcripts per million (TPM) across breast cancer subtype. (**c**) Mean NEDD9 RNA expression TPM in the 30 most extensively characterised PDTX models, including primary sample and different PDTX passage. Bar colours indicate PAM50 breast cancer subclass and internal numbers represent Integrative cluster. Coloured squares represent NEDD9 copy number, where tested. Error bars represent ±s.d. of at least two different passages.

Figure **5.1d** (Callari et al. 2021) represents the histological and molecular features of 53 PDTX models (including the 40 analysed in Bruna et al. 2016) profiled using mass cytometry, demonstrating the molecular heterogeneity present in the PDTX biobank.

5.3 NEDD9 copy number and gene expression in the PDTX biobank

To gain an understanding of NEDD9 expression and copy number variation in the PDTX biobank, I queried the Breast cancer PDTX encyclopedia (BCaPE) (http://caldaslab.cruk.cam.ac.uk/bcape). Almost half (15/32) of PDTX models analysed were NEDD9 copy number neutral (figure **5.2a)**, and a smaller proportion displayed copy number gain (10/32) or copy number loss (7/32). The box and whisker plot (figure **5.2b**) demonstrated a wide variation in NEDD9 mRNA expression (TPM) across the models tested. However, median expression levels for NEDD9 were lower in models from the Her2 PAM50 subtype and higher in models from the Luminal B PAM50 subtype, with models from the Basal and Luminal A subtypes demonstrating similar levels of expression. Next, I analysed individual PDTX models in more detail (figure **5.2c**), the IntClust subtype (number inside bar) were distributed across all models but a greater number of models were classified as the poor prognosis IntClust subtype 10 in models with lower expression of NEDD9 (6/15 in the top 50% expression, versus 8/15 in the bottom 50%). Models with NEDD9 copy number loss (red squares) were clustered at the lower range of NEDD9 expression, whereas models with copy number gain (blue squares) were distributed across the range. Unfortunately, the PDTX cohorts tested for copy number and mRNA expression have a limited overlap, and only with future full testing of the PDTX biobank will it be possible to determine if a true correlation between NEDD9 copy number and NEDD9 gene expression exists. This analysis mirrors the finding from section 2.1, where I found a wide range of NEDD9 gene expression and no correlation with either poor prognosis subtypes or correlation between poor outcome and NEDD9 copy number state in patient sample cohorts.

These data illustrate similar diversity exists, in terms of NEDD9 copy number and gene expression in the PDTX cohort, when compared to the patient clinical cohort in section 2.1.

5.4 PDTX protein expression as a biomarker of pathway activation

To utilise PDTX models as translational tools in this thesis, it was first important to understand if the signalling mechanisms and biological processes that were seen in cell lines representative of the

Claudin-low subset of tumours were recapitulated in PDTX models. To investigate this, I first selected a refined panel of PDTX models. As with the selection of the cell line panel in chapter 2.3, I used the Claudin-low score to define two PDTX categories; (i) Claudin-low, and (ii) "others" (non-Claudin-low). From the range of scores calculated, models in the Claudin-low category were models with the highest Claudin-low scores, whereas the models in the "others" category had the lowest. The Claudin-low score was calculated by Dr Oscar Rueda (MRC Biostatistics Unit, Cambridge) using the method described by Prat and colleagues (Prat et al. 2010). Table **5.1** below lists the models selected along with their originating sample type, PAM50 classification, Integrative cluster classification and Claudin-low score.

Model	Source	PAM50	Integrative cluster	Claudin-low score	NEDD9 expression	NEDD9 CN
STG316T	Primary breast	Basal	4	0.302	1.3	GAIN
AB630M	Brain metastasis	Basal	10	0.214	2.7	NEUTRAL
STG139T	Primary breast	Basal	10	0.186	1.8	LOSS
STG139M	Lung metastasis	Basal	10	0.151	1.1	LOSS
AB580T	Primary breast	Luminal B	8	-0.213	2.1	GAIN
AB551M	Brain metastasis	Luminal B	5	-0.219	1.1	GAIN
STG195M	Pleural effusion	Luminal B	8	-0.294	3.2	NEUTRAL
STG143T	Primary breast	Luminal B	8	-0.318	1.7	NEUTRAL

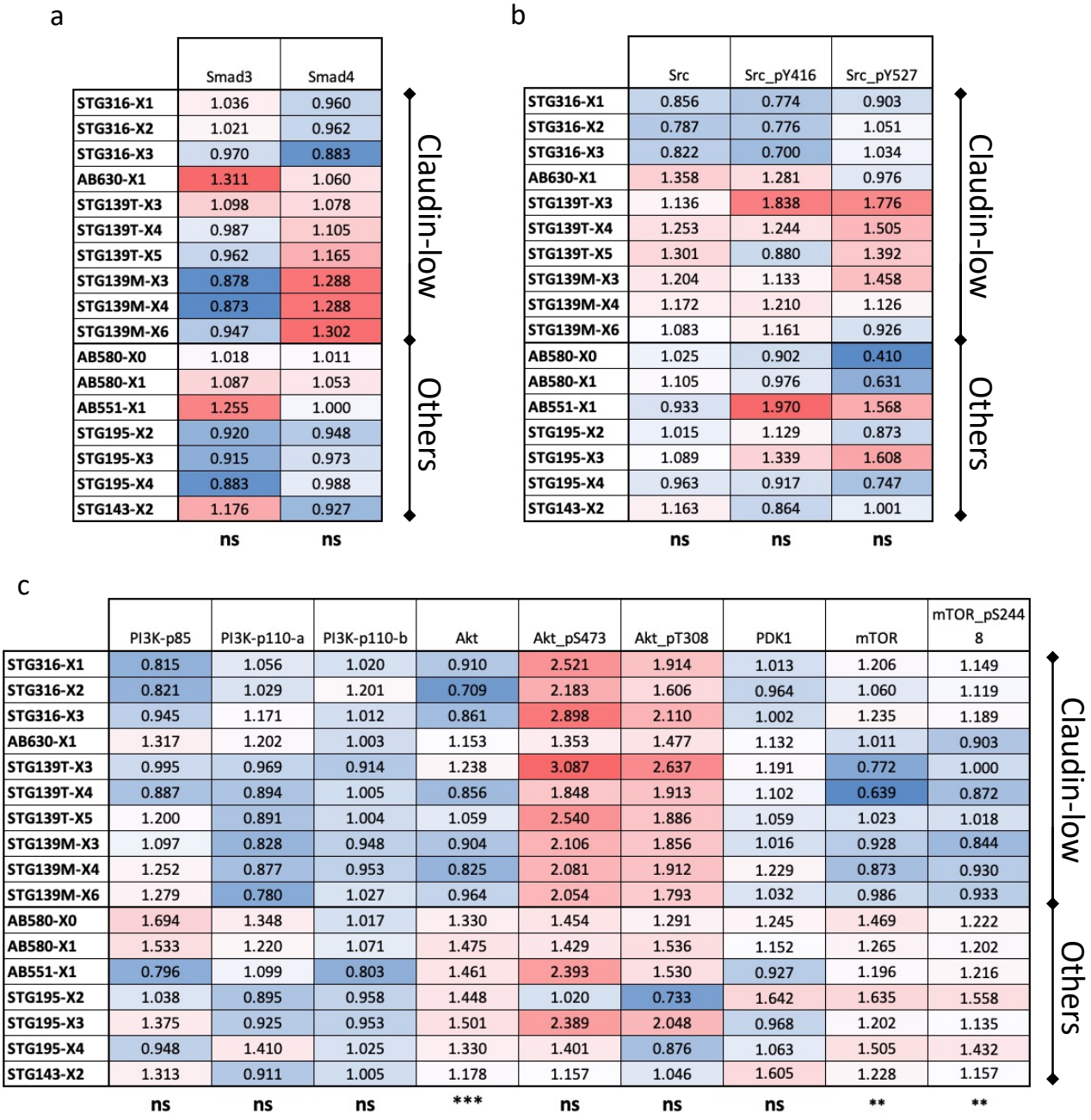
Table 5.1. A panel of PDTX models representing the Claudin-low and "others" breast cancer group. The upper section represents models belonging to the Claudin-low group and the lower section represents those belonging to the "others" group, as classified by the Claudin-low score. Higher Claudin-low scores correlate with poorer prognosis subtypes, but not with NEDD9 expression or copy number status.

The models with higher Claudin-low scores belonged to the more aggressive basal PAM50 subtype and poor prognosis integrative clusters 10 and 4, whereas models classified as "others" belonged to the less aggressive Luminal B subtype and better prognosis integrative clusters 8 and 5. As observed in figure **5.2c**, NEDD9 RNA expression did not correlate with Claudin-low classification, however, models which were NEDD9 copy number neutral and gain were represented in both groups but models with NEDD9 copy number loss were present in the Claudin-low group only. There was also no correlation observed between originating sample type and Claudin-low score, with two primary and two metastatic samples in each group. These observations agree with the heterogeneity observed in breast cancer, even within the Claudin-low subtype as recently highlighted by Pommier and colleagues (Pommier et al. 2020). Using this refined panel, I queried the reverse-phase protein array (RPPA) data generated in collaboration with the laboratory of Professor Gordon Mills at the Oregon Health and Science University (USA). Where available, multiple passages from each model were used as biological replicates except for AB630, AB551 and STG143 where only one passage was available for analysis. Multiple passages from the same

model used as biological replicates are represented in the figures by the suffix Xn and were included in the statistical analysis. Heatmaps were generated to visualise differences in normalised linear transformed protein expression and, for each protein analysed, a two-tailed Student's t-test was performed to assess the statistical significance of the difference in expression between the Claudin-low and "others" group. Protein expression was displayed by the sample group, with the PDTX models belonging to the Claudin-low group in the upper section of each heatmap and models belonging to the "others" group in the lower section.

As previously discussed, neither NEDD9 nor PKM2 expression levels are representative of TGFβ oncogenic effects, so I first looked at the protein expression of TGFβ pathway components in the data. Figure **5.3a** demonstrates that there were no significant differences between SMAD expression and PDTX subtype in this analysis, and therefore no conclusions could be made on TGFβ signalling activity in these models. Unfortunately, an antibody against phospho-SMAD2 was not included in the RPPA panel, the inclusion of this antibody would have proven more indicative of TGFβ signalling activity as is seen in western blots.

I next investigated signalling pathways downstream of TGFβ signalling which I had previously identified as being NEDD9 dependent. First, the expression of Src and the Src-pY416 and Src-pY527 phosphorylated tyrosine forms of Src (figure **5.3b**). Phosphorylation at these tyrosine residues has been linked to the regulation of Src kinase activity by influencing Src conformation (Irtegun et al. 2013; Spassov et al. 2018). Phosphorylation of Src-Y527 stabilises the closed conformation of Src, suppressing kinase activity, and Src-Y416 phosphorylation promotes kinase activity by stabilising an open protein structure allowing for substrate binding. Figure **5.3b** indicates that although within the group of models belonging to the Claudin-low phenotype, both STG139T and STG139M expressed higher levels of Src and both phospho-forms than STG316, as a whole there was no statistical difference between the Claudin-low and "others" group in terms of Src and phosphorylated-Src expression. Next, I moved to look at the regulation of pathways downstream of Src signalling, focussing first on the PI3K/AKT/mTOR pathway. This pathway is a signalling pathway involved in cell cycle regulation, and cell proliferation and is frequently hyperactive in cancer, specifically in the regulation of cancer stem cell dynamics (Bahmad et al. 2018; P. Xia and Xu 2015). Figure **5.3c** shows no significant difference was identified in the protein expression levels of the PI3K-p85 regulatory subunit and both catalytic PI3K-p100 subunits α and β between the Claudin-low and "others" group. Levels of AKT expression were significantly lower in the Claudin-low subgroup than the "others" group, but no significant difference was observed in the levels of both the AKT-pS473 and AKT-pT308 phospho-forms.



a	Smad3	Smad4
STG316-X1	1.036	0.960
STG316-X2	1.021	0.962
STG316-X3	0.970	0.883
AB630-X1	1.311	1.060
STG139T-X3	1.098	1.078
STG139T-X4	0.987	1.105
STG139T-X5	0.962	1.165
STG139M-X3	0.878	1.288
STG139M-X4	0.873	1.288
STG139M-X6	0.947	1.302
AB580-X0	1.018	1.011
AB580-X1	1.087	1.053
AB551-X1	1.255	1.000
STG195-X2	0.920	0.948
STG195-X3	0.915	0.973
STG195-X4	0.883	0.988
STG143-X2	1.176	0.927
	ns	ns

b	Src	Src_pY416	Src_pY527
STG316-X1	0.856	0.774	0.903
STG316-X2	0.787	0.776	1.051
STG316-X3	0.822	0.700	1.034
AB630-X1	1.358	1.281	0.976
STG139T-X3	1.136	1.838	1.776
STG139T-X4	1.253	1.244	1.505
STG139T-X5	1.301	0.880	1.392
STG139M-X3	1.204	1.133	1.458
STG139M-X4	1.172	1.210	1.126
STG139M-X6	1.083	1.161	0.926
AB580-X0	1.025	0.902	0.410
AB580-X1	1.105	0.976	0.631
AB551-X1	0.933	1.970	1.568
STG195-X2	1.015	1.129	0.873
STG195-X3	1.089	1.339	1.608
STG195-X4	0.963	0.917	0.747
STG143-X2	1.163	0.864	1.001
	ns	ns	ns

c	PI3K-p85	PI3K-p110-a	PI3K-p110-b	Akt	Akt_pS473	Akt_pT308	PDK1	mTOR	mTOR_pS2448
STG316-X1	0.815	1.056	1.020	0.910	2.521	1.914	1.013	1.206	1.149
STG316-X2	0.821	1.029	1.201	0.709	2.183	1.606	0.964	1.060	1.119
STG316-X3	0.945	1.171	1.012	0.861	2.898	2.110	1.002	1.235	1.189
AB630-X1	1.317	1.202	1.003	1.153	1.353	1.477	1.132	1.011	0.903
STG139T-X3	0.995	0.969	0.914	1.238	3.087	2.637	1.191	0.772	1.000
STG139T-X4	0.887	0.894	1.005	0.856	1.848	1.913	1.102	0.639	0.872
STG139T-X5	1.200	0.891	1.004	1.059	2.540	1.886	1.059	1.023	1.018
STG139M-X3	1.097	0.828	0.948	0.904	2.106	1.856	1.016	0.928	0.844
STG139M-X4	1.252	0.877	0.953	0.825	2.081	1.912	1.229	0.873	0.930
STG139M-X6	1.279	0.780	1.027	0.964	2.054	1.793	1.032	0.986	0.933
AB580-X0	1.694	1.348	1.017	1.330	1.454	1.291	1.245	1.469	1.222
AB580-X1	1.533	1.220	1.071	1.475	1.429	1.536	1.152	1.265	1.202
AB551-X1	0.796	1.099	0.803	1.461	2.393	1.530	0.927	1.196	1.216
STG195-X2	1.038	0.895	0.958	1.448	1.020	0.733	1.642	1.635	1.558
STG195-X3	1.375	0.925	0.953	1.501	2.389	2.048	0.968	1.202	1.135
STG195-X4	0.948	1.410	1.025	1.330	1.401	0.876	1.063	1.505	1.432
STG143-X2	1.313	0.911	1.005	1.178	1.157	1.046	1.605	1.228	1.157
	ns	ns	ns	***	ns	ns	ns	**	**

Figure 5.3. RPPA protein expression as a biomarker of TGFβ and Src signalling pathway activation in PDTX models

Normalised linear transformed protein expression of (**a**) SMAD3/4 (**b**) Src and phosphorylated Src and (**c**) members of the PI3K/AKT/mTOR signalling pathway downstream of Src activation in a panel of PDTX models. Models in the upper section of each heatmap represent models categorised as Claudin-low by correlation with the Claudin-low centroid (Prat et al. 2010). Models in the lower section of each heatmap represent those classified as non-claudin –low/others. Multiple rows for the same model represent biological replicates from consecutive passages as indicated, where available. For each heatmap the two-tailed Student's t-test was used for statistical analysis of difference in expression between the two groups, ns indicates not significant, ** indicates P=<0.01, *** indicates P=<0.001

Phosphorylation of the Akt S473 and T308 residues is necessary for AKT activation and phosphorylation of downstream substrates (Vanhaesebroeck and Alessi 2000). Tyrosine308 located in the AKT activation loop is phosphorylated by PDK1 (downstream of PI3K) and Serine473, located in the hydrophobic motif, is activated by the mTORC2 complex. Interestingly, levels of PDK1 expression were not significantly different between the groups, but the expression of mTOR (and phosphorylated mTOR-pS2448, which regulates mTORC complex formation) was significantly lower in the Claudin-low group. This suggests that in the "others" group, AKT activation is occurring via the usual canonical route, however, in the Claudin-low group, AKT activation may be occurring via an alternative, non-canonical mechanism. Non-canonical AKT signalling has been widely studied and there is published evidence for several alternative mechanisms (Nim et al. 2015).

PI3K/AKT signalling is understood to enhance Wnt signalling, as this pathway was regulated by NEDD9 in both the whole proteome (chapter 2.6) and gene expression (chapter 2.7) analyses, I looked at the levels of β-Catenin and GSK3 in the RPPA analysis. β-Catenin is a crucial transcription factor in Wnt signalling and aberrant activation of this pathway leads to the accumulation of β-Catenin in the nucleus resulting in the expression of many oncogenes (Shang, Hua, and Hu 2017), including c-Myc which as previously discussed is key in the regulation of PKM2 expression (chapter 4.2.1). Glycogen synthase kinase-3 (GSK3) is a serine/threonine kinase with known roles in regulating numerous cellular processes, including the PI3K/ATK/mTOR, Ras/Raf/MEK/ERK and Wnt/β-Catenin signalling pathways, and aberrant activity of GSK3 has been implicated in a wide range of human diseases (McCubrey et al. 2014). The GSK3 family consists of GSK3α and GSKβ, which are structurally similar but demonstrate different substrate specificity. GSK3 activity is regulated by phosphorylation, with inactivation of kinase activity caused by phosphorylation at serine 21 and serine 9 in GSK3α and GSKβ, respectively. In its active, non-phosphorylated form, GSK3β forms a complex with APC or Axin to phosphorylate and inactivate β-Catenin. Figure **5.4a** illustrates how β-Catenin levels are significantly higher in PDTX from the Claudin-low group compared to "others" but this does not correlate with β-Catenin phosphorylation, nor does AKT phosphorylation (figure **5.3c)** correlate with β-Catenin phosphorylation as would be expected in canonical Wnt/β-Catenin signalling. However, Yang and colleagues describe a mechanism of Src mediated β-Catenin phosphorylation at the alternative Y333 residue, not represented in the RPPA data so its expression cannot be evaluated (Yang et al. 2011). When considered along with the findings of Li and colleagues that in colorectal cancer NEDD9 is a target of aberrant Wnt/β-Catenin signalling (Y. Li et al. 2011), this suggests that in the Claudin-low breast cancer subtype Wnt/β-Catenin signalling may be important in regulating self-renewal. However, how NEDD9 is implicated in the regulation of non-canonical Wnt/β-Catenin signalling in Claudin-low breast cancer is yet to be determined.

a

	β-Catenin	β-Catenin_ pT41_S45	GSK3a/b	GSK-3a-b_pS21_S9	
STG316-X1	2.040	0.957	1.125	1.261	Claudin-low
STG316-X2	1.902	1.025	1.070	0.969	
STG316-X3	2.250	0.980	1.086	1.009	
AB630-X1	1.001	0.974	1.468	0.631	
STG139T-X3	1.675	1.010	1.188	2.299	
STG139T-X4	1.408	1.061	0.982	1.735	
STG139T-X5	1.521	1.036	1.128	1.121	
STG139M-X3	1.376	0.988	1.047	1.500	
STG139M-X4	1.247	1.014	1.020	1.537	
STG139M-X6	1.588	0.950	1.121	1.181	
AB580-X0	1.257	0.989	1.091	0.771	Others
AB580-X1	0.729	0.986	1.108	0.626	
AB551-X1	0.740	0.999	1.021	0.945	
STG195-X2	1.103	0.958	1.070	2.286	
STG195-X3	0.371	0.946	0.908	1.132	
STG195-X4	0.864	1.060	1.020	1.332	
STG143-X2	1.143	0.898	1.049	1.476	
	**	ns	ns	ns	

b

	N-Ras	A-Raf	MEK1	MEK1_pS217_S221	
STG316-X1	1.034	0.933	1.008	0.936	Claudin-low
STG316-X2	0.923	0.879	1.206	0.928	
STG316-X3	0.999	1.151	1.008	0.914	
AB630-X1	0.962	0.791	0.980	0.903	
STG139T-X3	0.997	1.137	1.154	0.954	
STG139T-X4	0.982	0.909	1.016	1.013	
STG139T-X5	1.000	1.047	1.052	1.004	
STG139M-X3	1.055	1.034	1.051	1.106	
STG139M-X4	1.172	1.177	1.143	1.210	
STG139M-X6	1.050	0.977	1.053	1.008	
AB580-X0	1.056	1.527	0.656	1.157	Others
AB580-X1	1.023	1.325	0.619	1.151	
AB551-X1	0.898	1.372	0.868	1.172	
STG195-X2	0.907	1.892	1.003	1.050	
STG195-X3	1.033	1.156	0.998	1.034	
STG195-X4	1.012	0.920	0.916	0.929	
STG143-X2	0.920	1.585	0.895	1.133	
	ns	**	*	ns	

c

	PKM2	LDHA	Hif-1-alpha	Hexokinase-II	Glutaminase	FoxM1	
STG316-X1	0.593	1.086	1.267	1.039	0.744	1.627	Claudin-low
STG316-X2	0.729	1.638	1.407	1.175	0.780	1.227	
STG316-X3	0.640	0.937	1.042	0.982	0.813	2.060	
AB630-X1	0.907	1.173	1.464	0.865	0.887	1.068	
STG139T-X3	1.105	1.453	4.133	0.908	0.797	1.451	
STG139T-X4	0.955	1.376	3.290	0.949	0.766	1.316	
STG139T-X5	1.156	1.695	2.053	0.948	0.792	1.084	
STG139M-X3	1.056	1.325	2.866	1.017	0.763	1.130	
STG139M-X4	1.002	1.143	2.232	0.923	0.788	1.193	
STG139M-X6	1.229	1.221	2.329	0.872	0.801	1.061	
AB580-X0	0.725	0.588	0.857	0.957	0.645	0.642	Others
AB580-X1	0.906	0.569	0.706	0.854	0.645	0.619	
AB551-X1	0.750	1.190	0.867	1.116	0.724	0.750	
STG195-X2	1.311	1.110	0.549	0.983	0.694	0.614	
STG195-X3	0.771	0.830	0.774	1.447	1.140	0.962	
STG195-X4	0.851	0.674	0.587	0.878	1.013	1.346	
STG143-X2	1.011	1.083	1.609	0.814	0.765	0.835	
	ns	*	*	ns	ns	*	

Figure 5.4. RPPA protein expression as a biomarker of β-Catenin and metabolic signalling pathway activation in PDTX models

Normalised linear transformed protein expression of (**a**) Wnt/β-Catenin/GSK signalling (**b**) the Ras/Raf/MEK/Erk signalling pathway and (**c**) glucose metabolic pathway components in a panel of PDTX models. Models in the upper section of each heatmap represent models categorised as Claudin-low by correlation with the Claudin-low centroid (Prat et al. 2010). Models in the lower section of each heatmap represent those classified as non-claudin –low/others. Multiple rows for the same model represent biological replicates from consecutive passages as indicated, where available. For each heatmap the two-tailed Student's t-test was used for statistical analysis of difference in expression between the two groups, ns indicates not significant, * indicates P=<0.05, ** indicates P=<0.01

As mentioned earlier in this chapter, the Ras/Raf/MEK/ERK pathway downstream of Scr activation is also regulated by GSK, and the gene expression analysis in chapter 2.7 identified enrichment of gene sets involved in MEK signalling in response to TGFβ treatment. Figure **5.4b** demonstrates little difference in the RPPA expression data between PDTX categories in N-Ras expression, but increased expression of A-Raf in the "others" category. Activation of MEK1 occurs through phosphorylation of serine 217 and serine 221 by Raf, and the lower expression levels of A-Raf and phosphorylated MEK1 in the Claudin-low subgroup suggests that Ras/Raf/MEK/ERK pathway activation is lower in the Claudin-low subgroup compared to the "others" group.

Finally, I analysed the expression of proteins involved in the glucose metabolism pathway, outlined in figure **5.4c**. As previously observed in chapter 4.3, the expression of PKM2 is variable across all PDTX models and no significant difference was seen between the groups. However, expression of Lactate dehydrogenase Alpha (LDHA), Hypoxia-inducible factor 1-alpha (Hif1α), and Forkhead box protein M1 (FOXM1) was significantly increased in the Claudin-low group when compared with the "others" group. LDHA is a metabolic enzyme that catalyses the reversible conversion between pyruvate and lactate, and tumours are known to express higher levels of LDHA compared to normal tissues as a reflection of the transition to Warburg metabolism. Levels of serum LDH have been shown to correlate with poor overall survival and reduced progression-free survival in breast cancer patients (D. Liu et al. 2019) and many other cancer types. The significant increase in LDHA expression in the Claudin-low group mirrors the effects seen in cell lines in chapter 4.5, where I demonstrated increased lactate production in the Claudin-low cell line MDA-MB-231 compared with the non-Claudin-low cell line HCC1954. HIF-1 is a transcription factor involved in the regulation of gene programs controlling glucose metabolism, cell proliferation and migration, and angiogenesis, and its activity is frequently upregulated in cancer (Semenza 2001; Masoud and Li 2015). The rapid growth of solid tumours often leads to a hypoxic environment, triggering the activation of HIF-1. In hypoxic tumour cells, HIF-1 causes a shift in glucose metabolism from oxidative phosphorylation towards glycolysis (the Warburg effect). To become active, a complex between HIF-1 alpha and HIF-1 beta subunits forms; this complex then recruits further basic helix-loop-helix proteins enabling HIF-1 DNA binding and activation of target gene transcription. One of the many metabolic genes regulated by HIF-1 is LDHA, and this regulation has been demonstrated as necessary for the growth and migration of cancer cells in pancreatic cancer (Cui et al. 2017) and gastric cancer (Hao et al. 2019). Another HIF-1 target gene involved in the regulation of cancer cell growth and survival via reprogramming of glucose metabolism is the proliferation specific transcription factor FOXM1 (L. Xia et al. 2012). FOXM1 is a target of c-Myc and drives the expression of LDHA but not Hexokinase 2 (HK2). Recently Ros and colleagues described how in ER-positive breast cancer,

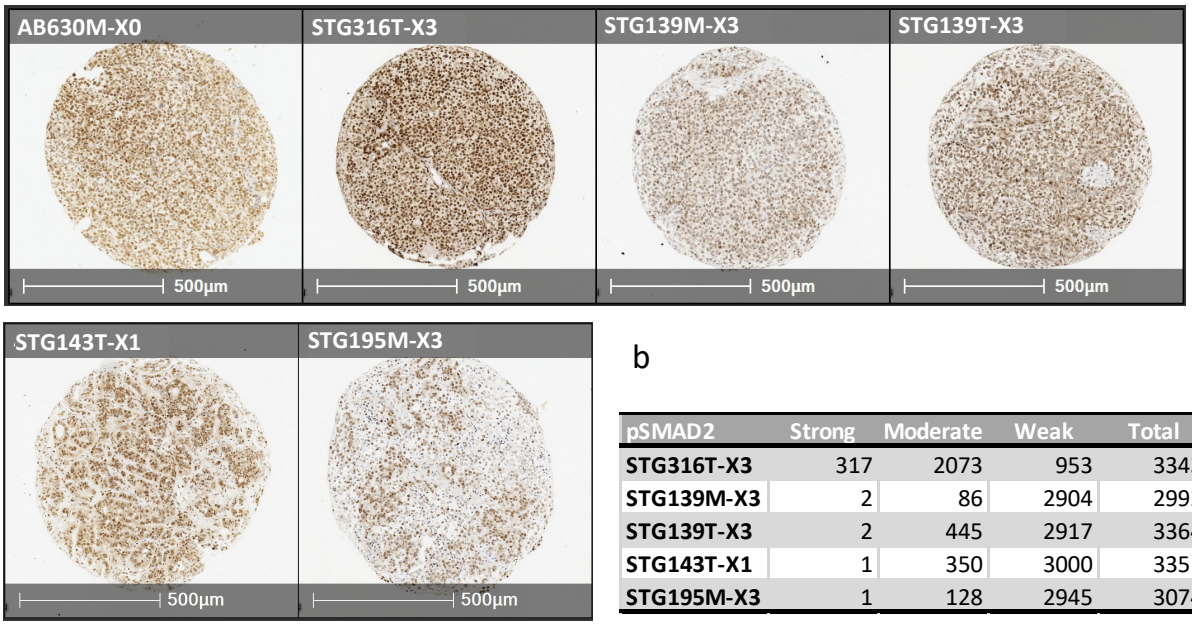
downstream metabolic effects of FOXM1 expression could be utilised for metabolic imaging and rapid assessment of patient treatment response to compounds targeting PI3K signalling (Ros et al. 2020). FOXM1 expression also appears to be key in the development of acquired resistance to therapy in ER-positive breast cancer, as PDTX models with induced resistance to PI3K inhibitors show increased levels of FOXM1 expression (Ros et al. 2020). The expression of HK2, which as mentioned is not directly targeted by increased c-Myc and FOXM1 expression, was not significantly different between the Claudin-low and "others" groups, and neither was the level of Glutaminase. Glutaminase is the enzyme responsible for the conversion of glutamine to glutamate, and the metabolic flux experiments in chapter 4.5 identified that TGFβ induced increased catabolism of glutamine into glutamate through both the oxidative and reductive carboxylation routes. However, no significant differences in Glutaminase expression between the subgroups are evident in this analysis. Significant differences in PI3K/AKT signalling and the increased expression of Hif1α, LDHA and FOXM1 in the Claudin-low subgroup indicates significant metabolic reprogramming and altered glucose metabolism in this group of PDTX, and also suggests that models belonging to this group may be sensitive to compounds targeting PI3K/AKT signalling.

Results of this analysis illustrate that protein expression indicating differential regulation of signalling pathways in the Claudin-low PDTX group compared to the "others" group, strengthen the observations of signalling pathways highlighted as significant in previous cell-line analyses. Pathways with significant differences in activity between the Claudin-low subgroup compared to the "others" were the PI3k/AKT/mTOR pathway, the Wnt/β-Catenin pathway and pathways regulating glucose metabolism. These findings strengthen previous data indicating that regulation of these pathways in response to TGFβ treatment is NEDD9 dependent in the Claudin-low subtype. However, as it was not possible to directly assess TGFβ activity in this analysis due to the absence of a pSMAD2 antibody, I cannot directly ascribe the observed changes to TGFβ signalling activity.

5.5 Drug response in PDTX models stratified by TGFβ signalling

The previous analysis demonstrated significant differences in cell signalling pathway activation in PDTX models belonging to the Claudin-low subset when compared to models from the "others" category. Whilst this is an indicator that in these models we see the activation of cellular processes similar to those observed as dependent upon NEDD9 in response to TGFβ signalling in Claudin-low cell lines, the activation of these pathways cannot be directly attributed to TGFβ signalling.

a



b

pSMAD2	Strong	Moderate	Weak	Total
STG316T-X3	317	2073	953	3343
STG139M-X3	2	86	2904	2992
STG139T-X3	2	445	2917	3364
STG143T-X1	1	350	3000	3351
STG195M-X3	1	128	2945	3074

c

Growth factor signalling	Target
AZ12601011	ALK5
AZD3463	IGFR1/ALK
BMS-754807	IGF-1R/IR
Gefitinib	EGFR
AZD8931	ERBB1/2/3
lapatinib	HER2

MEK/ERK PI3K signalling	Target
AZD6244	MEK1/2
AZ13776197	ERK1/2
BEZ235	PI3K
GDC-0941	PI3K
AZD8835	PI3K alpha
AZD8186	PI3Kb&d
GDC-0032	PI3K-β sparing
MK-2206	AKT1, AKT2
AZD2014	mTORC1/2

Cell cycle progression	Target
ABT-263	BCL2, BCL-XL, BCL-W
AZD4320	BCLL2/XL
AZD5991	MCL1
AZD4573	CDK9
AZD5576	RNA-pol/CDK9
CX-5461	quadruplexes

DDR and methylation	Target
AZD1390	ATM
AZD1775	WEE1
AZD5153	BRD4i
AZD7648	DNAPK
AZD7762	CHEK1/2
AZD2281	PARP1
BMN-673	PARP1

d

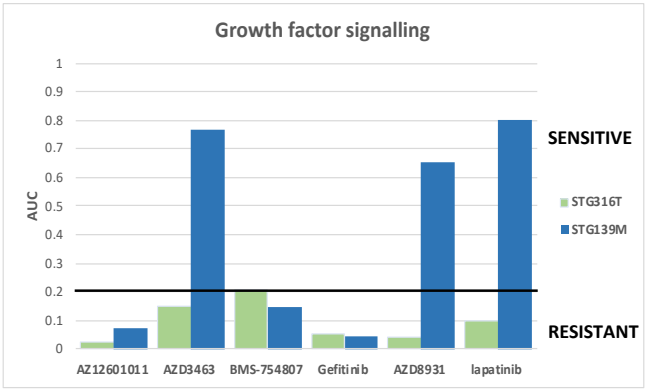


Figure 5.5. Protein expression by IHC as a biomarker of signalling pathway activation in PDTX models

(**a**) Phospho-SMAD2 IHC expression in a panel of PDTX models belonging to the Claudin-low subtype (AB630M, STG316T, STG139M and STG139 top panel) and "others" (STG143T and STG195M, bottom panel). An established passage X3 tumour was used for each model except in the case of AB630M and STG143T, where this was not available. (**b**) Quantification of IHC pSMAD2 nuclear staining intensity. (**c**) Individual drugs and drugs groups used for ex-vivo high throughput drug screens. (**d**) AUC values from high throughput drug screens using drugs targeting growth factor signalling pathways. Drug screen comparisons were performed in two PDTX models from the Claudin-low group with high TGFβ pathway activity (STG316T) and lower TGFβ pathway activity (STG139M). The bold line represents the arbitrary AUC threshold value of 0.2, above which models are considered sensitive.

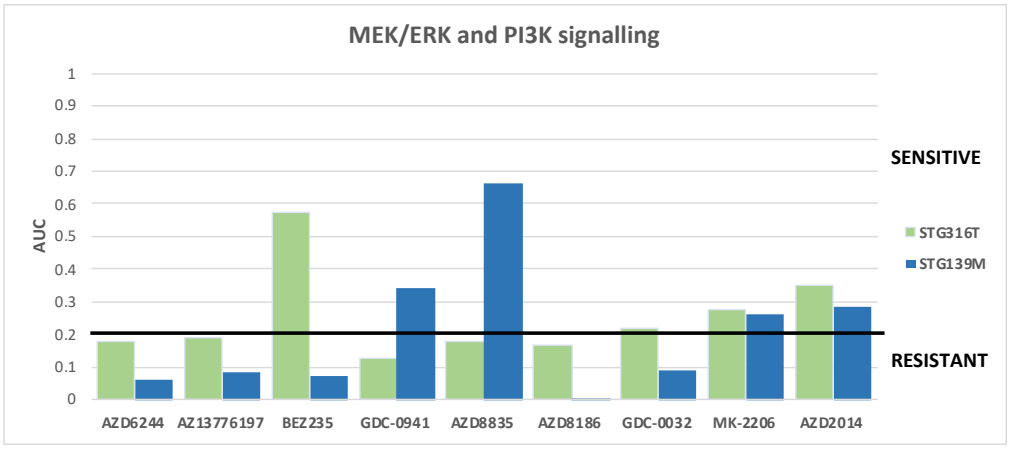
To assess TGFβ signalling pathway activity in the panel of PDTX models used for the RPPA analysis, I analysed phospho-SMAD2 (pSMAD2) expression levels by immunohistochemistry (IHC) in tissue microarrays (TMA) which had been prepared from formalin-fixed, paraffin-embedded PTDX tissue. For each of the models analysed, I used an established passage X3 tumour, except in the case of AB630M, and STG143T which had not reached this passage level at the time of testing. Figure **5.5a** illustrates the staining achieved for the four models belonging to the Claudin-low subset (upper section) and two models representative of the "others" group (lower section). The degree of pSMAD2 expression is indicated by the intensity of brown 3,3'-diaminobenzidine (DAB) staining; a hematoxylin counterstain was also used to identify nuclei, but as pSMAD2 expression is predominantly nuclear, the hematoxylin staining was somewhat obscured. Expression of pSMAD2 and, therefore, TGFβ signalling pathway activity was variable across all six models analysed. The intensity of nuclear pSMAD2 staining was quantified using the HALO multiplex IHC software (Indica Labs) and the same intensity threshold settings were used for all models across the TMA. Thresholds were set to quantify nuclei with weak, moderate and strong pSMAD2 staining. AB630M was not included in this analysis as this core was from a different TMA, and therefore a direct comparison of staining intensity could not be made. Quantification of pSMAD2 staining (figure **5.5b**) demonstrated that STG316T had the highest number of nuclei with strong and moderate staining whereas STG139M had the lowest, therefore I chose these models to investigate high throughput drug response data. Interestingly, although in cell lines we demonstrated NEDD9 expression as an indicator of TGFβ pathway activation through a positive feedback loop, these two models demonstrate very similar levels of NEDD9 mRNA expression (Table 5.1). However, STG316T is a model with NEDD9 copy number gain and STG139M has a NEDD9 copy number loss.

Using short-term *ex-vivo* cultures prepared from STG316T and STG139M, two PDTX models which exhibit high and low TGFβ pathway activation respectively, I investigated drug response data in a panel of compounds spanning a range of cellular targets, and which were either approved cancer treatments or drugs targeting key cancer pathways. The compounds were grouped into four categories, targeting growth factor signalling, the MEK/ERK and PI3K pathways, cell cycle progression and apoptosis, and DNA damage repair processes and DNA methylation. A full list of the drugs and their targets can be seen in figure **5.5c**. The data generation for this analysis was performed by Dr Elham Esmaeilishirazifard also from the Caldas laboratory, CRUK-CI, using the method described [methods 16] (Bruna et al. 2016). Additional drug screen data generated in collaboration with Dr Matthew Garnett at the Wellcome Sanger Institute, Cambridge, UK is also publicly available at https://caldaslab.cruk.cam.ac.uk/bcape/. For each model tested, a dose-response curve was generated using a range of drug concentrations based on the published IC50 for each compound. Seven days after drug treatment, cell viability was assessed using the

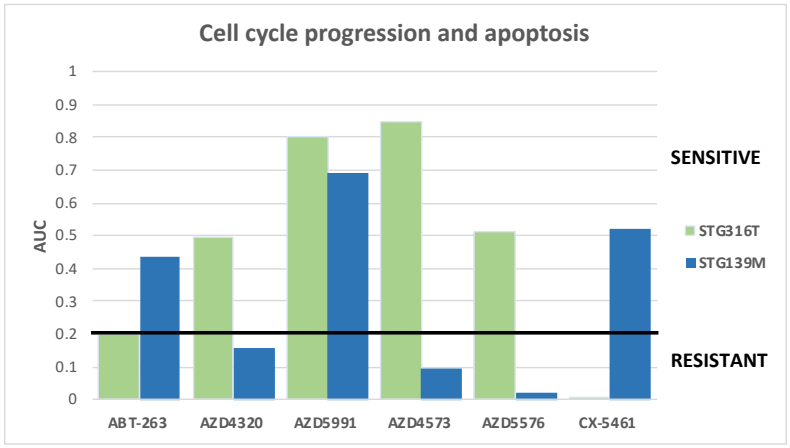
CellTiter-Glo® 3D assay (Promega), a bioluminescent assay that measures ATP generated by live cells as an indicator of viability. From the viability measure at each drug concentration, a dose-response curve was generated and the area under the curve (AUC) was calculated as an indicator of drug sensitivity. An AUC of 1 indicates a drug that is toxic to cells at all concentrations tested and an AUC of 0 indicates a drug that is not toxic at any concentration. Based on the range of data we observed in our PDTX models, we set an arbitrary threshold of AUC 0.2, above which models are considered sensitive to a drug. In analysing the drug screen data, I aimed to identify compounds that showed distinctly different drug responses between the STG316T and STG139M models, and especially to identify drugs where resistance was observed in one model over the other. These differences would indicate that in these models, both from the same Claudin-low subtype, oncogenic TGFβ signalling may be responsible for observed variations in drug response.

The first drug panel I analysed was for drugs targeting growth factor signalling. Figure **5.5d** illustrates that although STG139M was extremely sensitive to drugs targeting IGFR1/ALK signalling and ERBB/HER2 signalling, STG316T was resistant to all drugs in this panel. Interestingly, neither model was sensitive to inhibition of TGFβ signalling by the ALK5 inhibitor AZ12601011, although TGFβ signalling is occurring in both models as evident from the IHC analysis of pSMAD2 expression. Analysis of drugs targeting the MEK/ERK and PI3K signalling pathways (figure **5.6a**) demonstrated major differences in drug response in response to the PI3K targeting drug BEZ235, where only STG316T was sensitive, and also the PI3K-alpha inhibitor AZD8835 and pan-PI3K inhibitor GDC-0941, where the opposite response was observed with only STG139M demonstrating sensitivity. Analysis of compounds targeting cell cycle progression and apoptosis (figure **5.6b**) revealed that STG139M was sensitive to the ribosomal RNA synthesis inhibitor CX-5461 and the apoptosis-inducing B-cell lymphoma-2 (BCL2) protein family inhibitor ABT-263, whereas STG316T was not. Conversely, STG316T was sensitive to AZD4320, another BCL2 protein family inhibitor, and to AZD4573 and AZD5576 both of which are cyclin-dependent kinase 9 (CDK9) inhibitors. Analysis of compounds targeting DNA damage repair processes and DNA methylation (figure **5.6c**) demonstrated that overall both models are less sensitive to these classes of compounds, STG316T was sensitive to the BET-bromodomain-4 inhibitor AZD5153, the checkpoint kinase CHEK1/2 inhibitor AZD7762 and the PARP1 inhibitor BMN-673. When comparing resistance to drugs in both models, STG316T was more resistant to drugs targeting growth factor signalling, and DNA damage repair processes, whereas STG139M demonstrated resistance to drugs targeting MEK/ERK signalling and cell cycle progression.

a



b



c

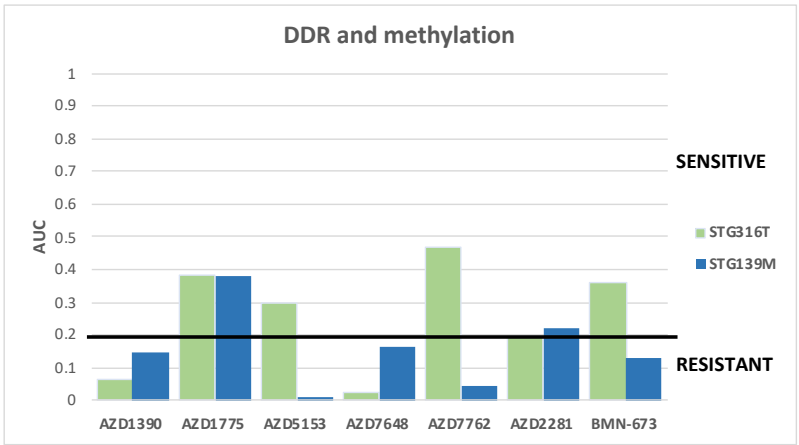


Figure 5.6. High throughput drug screens in Claudin-low group models with high versus low TGFβ pathway activity.

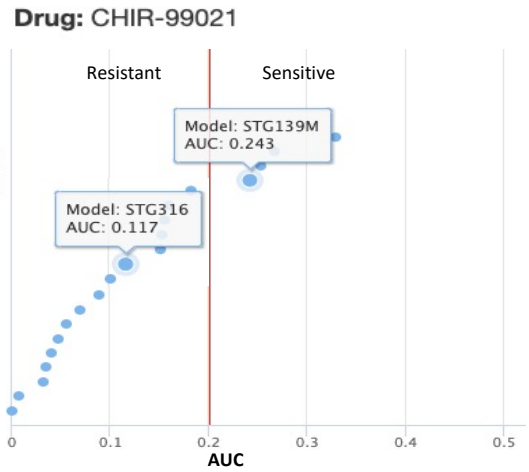
AUC values from high throughput drug screens using drugs targeting **(a)** the MEK/ERK and PI3K signalling pathways, **(b)** Cell cycle progression and apoptotic processes and **(c)** DNA damage repair and methylation processes. Drug screen comparisons were performed in two PDTX models from the Claudin-low group with high TGFβ pathway activity (STG316T) and lower TGFβ pathway activity (STG139M). The bold line represents the arbitrary AUC threshold value of 0.2, above which models are considered sensitive.

The results from high-throughput drug screens demonstrate diversity in drug response in PDTX models. However, when models classified as Claudin-low were stratified by pSMAD2 expression, an indicator of TGFβ signalling, differences in drug sensitivity emerged. STG316T, a model with high TGFβ signalling demonstrated sensitivity to a wide range of drugs targeting PI3K/AKT/mTOR signalling and cell cycle regulation, however, this model was resistant to compounds targeting growth factor signalling and DNA damage repair processes. In STG139M, a model demonstrating lower levels of TGFβ signalling, sensitivity to compounds targeting IGFR1/ALK, ERBB2 and PI3K signalling was observed, but these models were more resistant to drugs targeting MEK/ERK signalling and cell cycle progression.

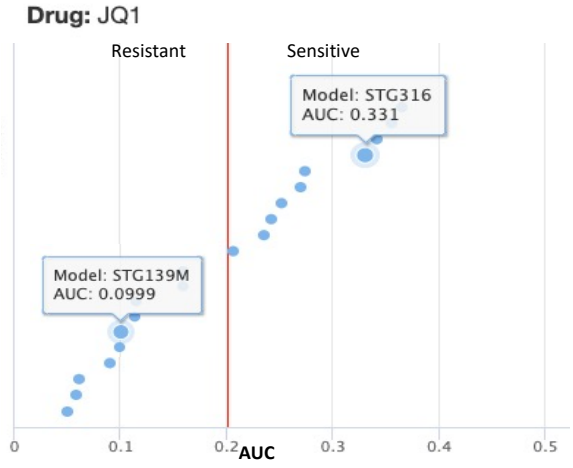
5.6 Drug sensitivity and Wnt/β-Catenin signalling

Due to observations made earlier in this chapter and in chapter 2 on the significance of Wnt signalling in the Claudin-low subtype, I also analysed response data to drugs targeting the Wnt signalling pathway. For this analysis, it was necessary to look at the previous analysis performed in collaboration with Dr Matthew Garnett (https://caldaslab.cruk.cam.ac.uk/bcape/), as drugs targeting the Wnt pathway are not included in the currently used drug panel. Figure **5.7a** demonstrates that STG139M was sensitive to the GSK3 inhibitor CHIR-99021, whereas STG316T was resistant. Although CHIR-99021 is a drug that targets the Wnt/β-Catenin signalling pathway, the direct target of this drug is GSK3. In canonical Wnt/β-Catenin signalling, GSK3 regulates β-Catenin transcriptional activity by phosphorylation and this phosphorylation targets β-Catenin for ubiquitylation and proteasomal degradation (Wu and Pan 2010). Where GSK3 is inhibited, either via PI3K/AKT signalling or by the use of a GSK3 inhibitor drug, this triggers β-Catenin mediated transcriptional activation. As mentioned in chapter 2.5, GSK3 is a serine/threonine kinase known to regulate a multitude of cellular processes, it is possible the sensitivity to CHIR-99021 observed in STG139M is via one of these other mechanisms, separate to Wnt/β-Catenin signalling. Another drug recently shown to inhibit Wnt signalling is the Bromodomain and extra terminal domain family (BET) inhibitor JQ1. Bromodomain-containing proteins regulate the processes of chromatin remodelling and transcriptional regulation through binding to acetylated histone lysine residues and acting as scaffolds for the recruitment of transcriptional regulatory complexes, this process is frequently dysregulated in cancer (Andreoli et al. 2013). JQ1 competes with BET proteins for binding to histone acetylation sites thus inhibiting transcription of target genes. However, recent studies into the mechanism of action of JQ1 have revealed that it causes cell cycle arrest and downregulation of genes involved in self-renewal, including WNT (Alghamdi et al. 2016; Y. Zhang et al. 2018).

a



b



c

	β -Catenin	c-Myc
STG316-X1	2.040	1.102
STG316-X2	1.902	1.039
STG316-X3	2.250	1.120
STG139M-X3	1.376	0.892
STG139M-X4	1.247	0.874
STG139M-X6	1.588	0.896
	***	*

d

c-Myc	Strong	Moderate	Weak	Total
STG316T-X3	89	2003	1303	3396
STG139M-X3	1	15	3646	3662
STG195M-X3	17	300	3754	4071
STG143T-X1	0	1	3772	3773

e

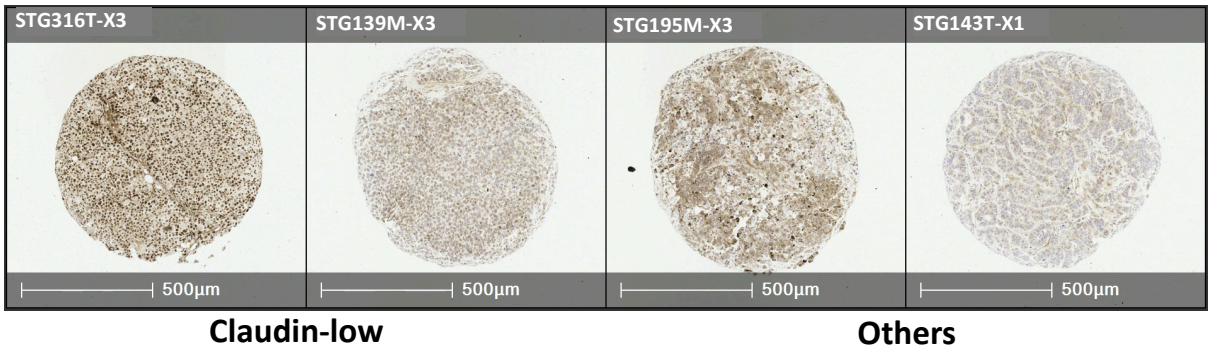


Figure 5.7. Drug sensitivity and Wnt/β-Catenin signalling activity
Drug sensitivity analysis in AUC values for **(a)** CHIR-99021 (GSK3 inhibitor) and **(b)** JQ1 from https://caldaslab.cruk.cam.ac.uk/bcape/ **(c)** Normalised linear transformed protein expression of β-Catenin and the Wnt target gene c-Myc in PDTX models STG316T (high TGFβ pathway activity) and STG139M (low TGFβ pathway activity). A two-tailed Student's t-test was used for statistical analysis of difference in expression between the two groups, * indicates P=<0.05, *** indicates P=<0.001 **(d)** Quantification of IHC c-Myc staining intensity from TMA staining. (**e**) c-Myc IHC expression in a panel of PDTX models belonging to the Claudin-low subtype (STG316T and STG139M) and others category (STG195M and STG143T).

Analysis of the drug response data from BCaPE (figure **5.7b**) revealed that STG316T was sensitive to JQ1 whereas STG139M was resistant.

The RPPA analysis (figure **5.4a**) showed that β-Catenin expression was significantly increased in PDTX models from the Claudin-low subgroup, however, closer analysis revealed that within the Claudin-low subgroup, the STG316T model expressed significantly higher levels of both β-Catenin and the Wnt target gene c-Myc than the STG139M model (figure **5.7c**). To further investigate this diversity of expression, I returned to the TMA IHC analysis and looked at the levels of c-Myc in the same cores used for the analysis of pSMAD2 staining. Figure **5.7e** demonstrated that in the models tested, two from the Claudin-low subgroup and two from the "others" subgroup, diversity in c-Myc expression was evident. However, when the images were quantified using the HALO multiplex IHC software as previously, STG316T, the Claudin-low model that demonstrated the highest levels of oncogenic TGFβ signalling had a higher number of cells staining strong and moderate for c-Myc than the other three models. As with the pSMAD2 analysis, stain-specific intensity thresholds for c-Myc were established and used across all cores of the TMA.

These results demonstrate that using a combination of IHC stains it is possible to identify PDTX models where oncogenic TGFβ signalling activity is occurring and that the model STG316T which falls into this category was sensitive to the drug JQ1. This observation indicates that Claudin-low PDTX models where oncogenic TGFβ signalling activity is occurring may be reliant on the Wnt/β-Catenin pathway for self-renewal. Conversely, Claudin-low models with low levels of oncogenic TGFβ signalling activity, of which STG139M is an example, are less reliant upon Wnt/β-Catenin signalling and more reliant upon other mechanisms for promoting self-renewal. The reliance upon Wnt/β-Catenin signalling in the STG316T model could also explain why this model was particularly sensitive to inhibitors of the PI3K/AKT pathway as observed in chapter 5.5, as PI3K/AKT signalling inhibits GSK3 allowing for β-Catenin mediated transcription. Inhibition of this pathway would lead to increased GSK3 and therefore increased proteasomal degradation of β-Catenin.

5.7 Drug sensitivity and NEDD9 expression

Finally, to reinforce my hypothesis, it was necessary to link the drug response data in Claudin-low PDTX models stratified by classification for oncogenic TGFβ signalling and c-Myc expression with the abundance of NEDD9. As previously mentioned, increased levels of NEDD9 expression alone were not

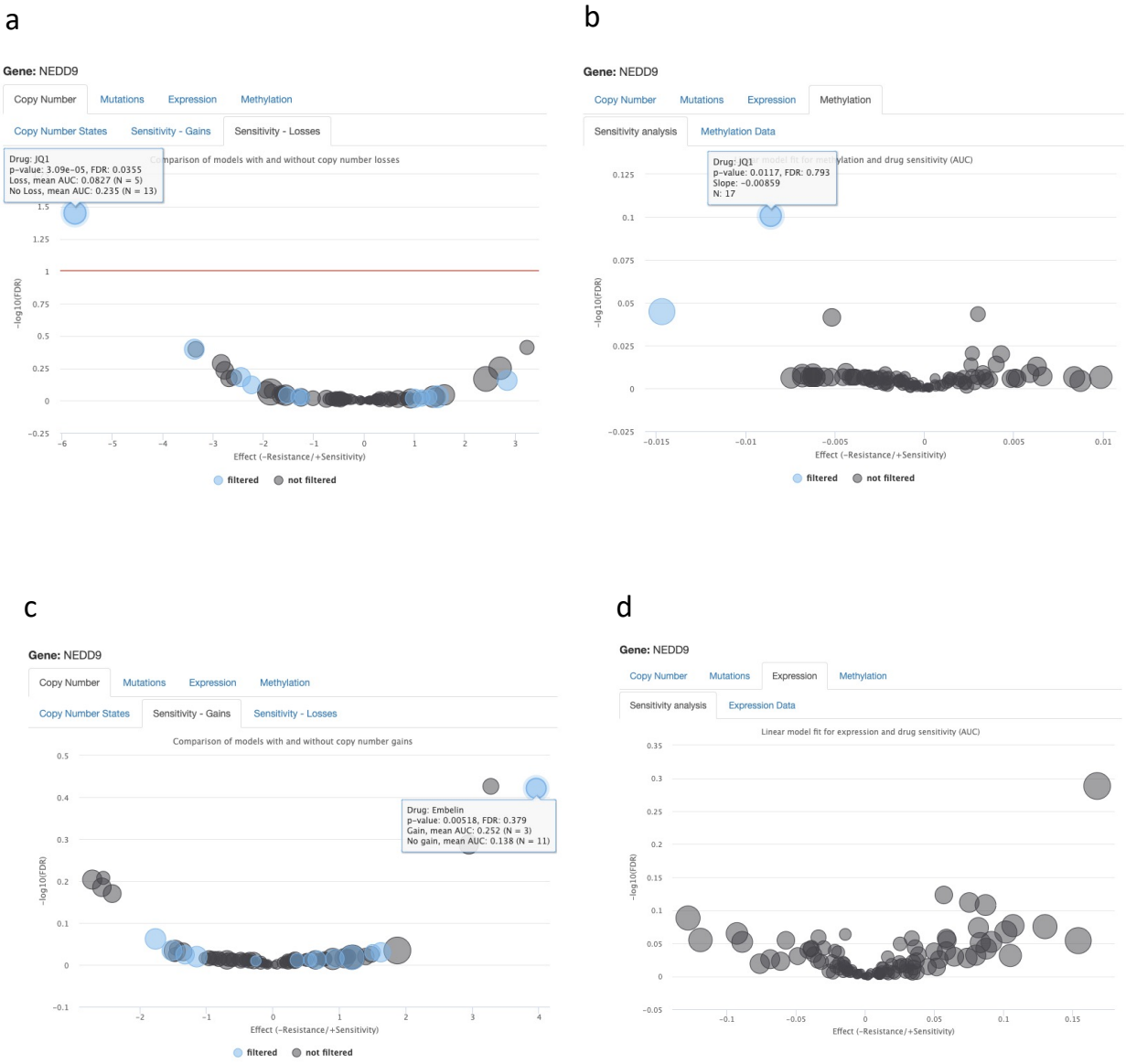


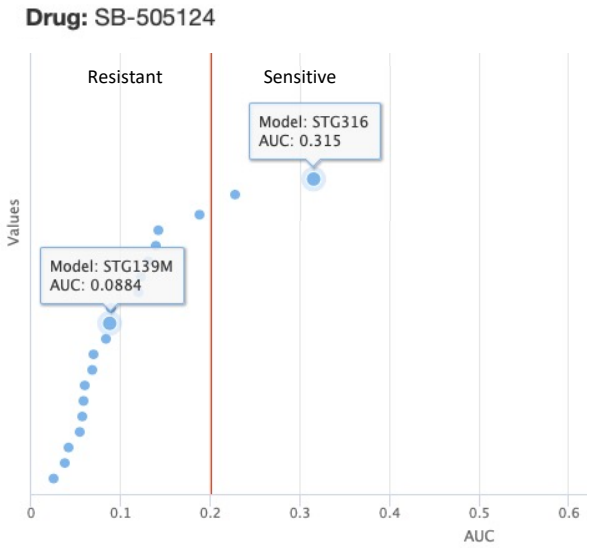
Figure 5.8. NEDD9 copy number state, promoter methylation and response to JQ1
Analysis of high throughput drug screen sensitivity data in PDTX models with (**a**) NEDD9 copy number loss, (**b**) NEDD9 promoter methylation and (**c**) NEDD9 copy number gain. (**d**) Analysis of drug sensitivity data based on NEDD9 expression. All figures are taken from https://caldaslab.cruk.cam.ac.uk/bcape/

sufficient to predict breast cancer tumour initiating cell (BTIC) response to TGFβ treatment *in vitro*, nor did they correlate with breast tumour subtype or overall survival in patient clinical cohorts. However, for oncogenic TGFβ signalling to be active in the Claudin-low subset, NEDD9 expression is necessary as NEDD9 functions as a signalling hub coordinating TGFβ downstream pathways.

First, I queried the BCaPE database for drug response by NEDD9 copy number status. Figure **5.8a** demonstrates that models with NEDD9 copy number loss were significantly more resistant to JQ1 treatment when compared to those without copy number alteration. Drugs annotated in blue (filtered) are those which demonstrated a significant change in drug response in terms of AUC in the altered versus unaltered group. The value on the y-axis represents the -log10 scale adjusted p-value of a t-test comparing the AUC in models with copy number losses versus models with no copy number alterations, values greater than 1 are considered statistically significant. Secondly, I looked at the drug response data in models in relation to NEDD9 promoter methylation, an indicator of repression of gene transcription. This analysis (figure **5.8b**) demonstrated that models with NEDD9 promoter methylation, and therefore repression of NEDD9 gene expression, were significantly more resistant to JQ1 treatment. Here, the filtered drugs (blue) were those which demonstrated a statistically significant drug response ($p=<0.05$) with a promoter methylation score of greater than 25% in all models. Y-axis values represent the -log10 scale of a linear model comparing AUC to promoter methylation and values greater than 1 are considered statistically significant. When analysing the data for models with NEDD9 copy number gain (figure **5.8c**), no drugs demonstrated a significant change in drug sensitivity, likewise with NEDD9 expression data (figure **5.8d**), levels of expression did not correlate with drug sensitivity. Models with NEDD9 copy number gain were more sensitive to the drug Embelin than those without (figure **5.8c**), however, this difference was not significant. Embelin is a naturally occurring para-benzoquinone and has a wide range of cellular effects.

To confirm the reliance of the Claudin-low, TGFβ high model STG316T on TGFβ signalling, I queried the drug response data for drugs inhibiting the TGFβ pathway. The previous analysis (figure **5.5d**) included the selective ALK4/ALK7 and TGFβR1 inhibitor AZ12601011, a drug which inhibits phosphorylation of SMAD2 and therefore downstream TGFβ signalling, in this analysis no significant differences in drug sensitivity between the Claudin-low, TGFβ high model STG316T and the model Claudin-low, TGFβ low model STG139M were observed. However, when I queried the drug response data to an alternative TGFβ inhibitor compound in the BCaPE database, a significant difference in response was observed between the two models. This alternative compound (SB-505124) is a selective inhibitor of both ALK4 and ALK5, inhibiting phosphorylation of both SMAD2 and SMAD3 and

a



b

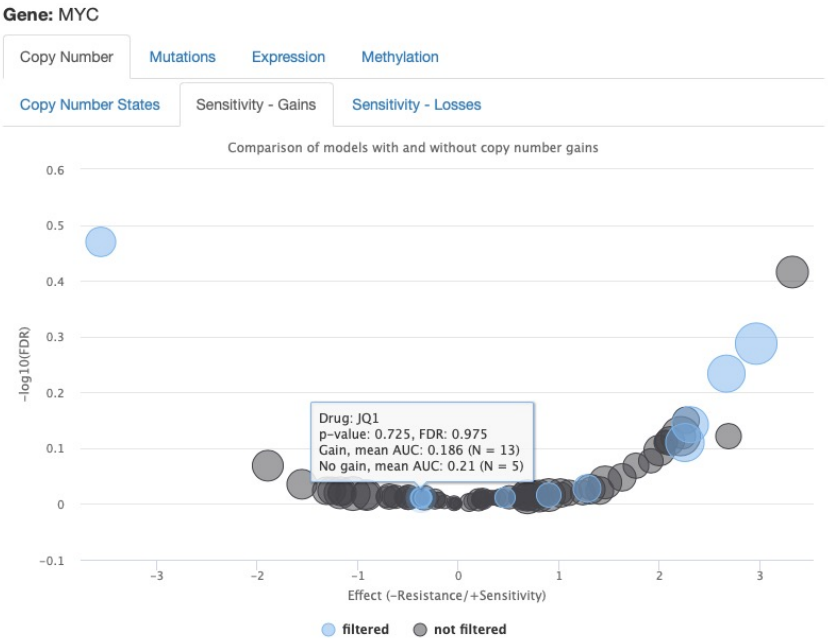


Figure 5.9. Response to inhibitors of TGFβ signalling, and JQ1 sensitivity in c-Myc overexpression

(a) Drug sensitivity analysis in AUC values for the TGFβ inhibitor SB-505124. (**b**) Analysis of high throughput drug screen sensitivity data in PDTX models with c-Myc copy number gain. Both figures are taken from from https://caldaslab.cruk.cam.ac.uk/bcape/

subsequent activation of TGFβ induced MAPK pathway activation. This analysis (figure **5.9a**) shows the Claudin-low, TGFβ high model STG316T was sensitive to SB-505124 and Claudin-low, TGFβ low model STG139M was resistant. This confirms the increased reliance of STG316T on TGFβ signalling, but also suggests that oncogenic TGFβ signalling is dependent upon MAP kinase activity, as indicated by the whole proteome analysis (chapter 2.6) and the differential gene expression analysis (chapter 2.7) in the Claudin-low cell line MDA-MB-231. As c-Myc is aberrantly expressed in over 70% of cancers, as a control to confirm that sensitivity observed in drugs targeting Wnt signalling was not purely a reflection of increased c-Myc expression in STG316T, I queried the BCaPE database for drug sensitivity in relation to c-Myc expression. This analysis (figure **5.9b**) demonstrated no significant correlation between c-Myc overexpression and sensitivity to the drug JQ1.

Analysis of drug response data when stratified by NEDD9 copy number, expression and NEDD9 promoter methylation demonstrated that models with NEDD9 copy number loss were significantly resistant to JQ1. Models with significant increases in NEDD9 promoter methylation of greater than 25% (and therefore repression of NEDD9 expression) were also resistant to JQ1 treatment, however here the effect was not statistically significant. Analysis of drug response data stratified by NEDD9 expression and NEDD9 copy number gain showed no significant differences in sensitivity. This observation strengthens the findings from chapter 5.6, that in a PDTX model with oncogenic TFGβ signalling and high expression of c-Myc the sensitivity to the Wnt/β-Catenin target JQ1 is dependent upon the availability of NEDD9, and is not merely a reflection of c-Myc overexpression. Further analysis also suggested that in this model, the mechanism of TFGβ signalling is dependent upon MAP kinase pathway activation.

5.8 Summary

In this chapter, I utilised the Cambridge PDTX biobank to build on my previous findings of the mechanisms involved in NEDD9 dependent oncogenic TGFβ signalling in the Claudin-low breast cancer subtype. I queried the biobank for correlation between breast cancer subtype, NEDD9 copy number state and NEDD9 expression, and confirmed that similar to the breast cancer patient cohort, there was no correlation between NEDD9 expression and breast cancer subtype in PDTX models. Next, I analysed RPPA data generated from multiple passages from each PDTX model for markers of the signalling pathways observed in my cell line analyses. Using the same criteria as used in cell lines, I classified the PDTX models into Claudin-low and “others" groups and demonstrated activation of similar pathways as seen in the cell line analyses.

Using pSMAD2 expression as a marker of TGFβ signalling activity I investigated the drug sensitivity of two Claudin-low PDTX models with opposing TGFβ signalling activity. I then further stratified the panel of PDTX models using c-Myc expression as a biomarker of increased Wnt/β-Catenin signalling. This stratification revealed that the Claudin-low PDTX model STG316T was sensitive to the Wnt/β-Catenin inhibitor JQ1. Finally, I looked at drug sensitivity data in models where NEDD9 expression was reduced, either through copy number loss or NEDD9 promoter methylation and confirmed that these models are uniquely resistant to JQ1, indicating that NEDD9 dependent signalling mechanisms may be necessary for the JQ1 sensitivity observed in STG316T. Control analyses confirmed that this sensitivity was not a reflection of c-Myc overexpression and that the Claudin-low/TGFβ high model STG316T was sensitive to inhibition of TGFβ signalling, possibly via MAPK pathway activity.

Together, these results indicate that in Claudin-low PDTX models exhibiting oncogenic TGFβ signalling a reliance on Wnt/β-Catenin mediated self-renewal may exist. Conversely, in Claudin-low models lacking oncogenic TGFβ signalling activity, reliance upon Wnt/β-Catenin signalling is less. Furthermore, I hypothesise that Wnt/β-Catenin signalling activity is dependent upon NEDD9 as models with NEDD9 copy number loss or NEDD9 promoter methylation are resistant to the Wnt/β-Catenin inhibitor JQ1. The reliance upon Wnt/β-Catenin signalling in the STG316T model could also explain why this model was particularly sensitive to inhibitors of the PI3K/AKT pathway.

6 . Discussion

6.1 Preface

Breast cancer is the most commonly diagnosed cancer worldwide, and in the UK it accounts for the highest number of cancer deaths. If diagnosed at an early stage 70-80% of breast cancer cases are curable. However, if breast cancer is detected at an advanced stage following metastasis, or recurrence with distant metastasis occurring following treatment, it is considered incurable. Clinically, breast cancer tumours are classified by size (grade), the extent of tumour spread (stage) and the expression of nuclear hormone receptors (ER and PR) and cell surface growth factor receptors (HER2), and this classification is used to make treatment decisions on a semi-personalised basis. Decades of breast cancer research have however highlighted the unequivocal heterogeneity of breast cancer, and the importance of this heterogeneity on tumour response, relapse and resistance to therapy. (Perou et al. 2000; Sørlie et al. 2001; Curtis et al. 2012; Pereira et al. 2016). This research has underscored the importance of understanding the complexity of breast tumour heterogeneity and its influence on breast cancer treatment and signals a shift from the current "one size fits all" approach to a more personalised model of breast cancer treatment.

Adding further complexity to the sphere of breast cancer heterogeneity was the characterisation of the Claudin-low breast cancer subtype. This class of tumours are generally more aggressive, the majority are negative for the expression of oestrogen receptors, and they are classified by their low expression of tight-junction and cell adhesion molecules (Herschkowitz et al. 2007). However, a recent publication by Pommier and colleagues described that even within the Claudin-low subtype there exists significant diversity with three distinct subgroups identified (Pommier et al. 2020). Our previous findings identified that in the Claudin-low breast cancer subtype, TGFβ regulates BTIC dynamics (Bruna et al. 2012). This research, which demonstrated NEDD9 as instrumental in regulating the mechanism of TGFβ induced BTIC self-renewal in the Claudin-low breast cancer subtype became the foundation for this PhD thesis. The overall aim of this body of work was to further clarify the mechanism by which NEDD9 regulated TGFβ dependent regulation of BTIC dynamics occurs, and to investigate the translational application of the findings, through identifying biomarkers of oncogenic TGFβ signalling and identifying drugs that may target specifically these patients with the more aggressive Claudin-low breast cancer subtype. As mentioned, even within the Claudin-low subtype heterogeneity exists. High TGFβ activity alone is not prognostic of poor survival in breast cancer, so it is important to identify patients where oncogenic TGFβ

signalling is occurring, to understand the signalling pathways regulated by TGFβ activity in these patients, and to identify drugs that may prove effective in targeting these pathways.

6.2 Summary of aims

This thesis marks a natural progression of research from my contribution to the Bruna 2012 paper. The aims were to elucidate the role of NEDD9 in oncogenic TGFβ signalling in the Claudin-low breast cancer subtype. To this aim, I performed the following investigations:

1. **The characterisation of NEDD9 RNA expression, post-translational modification and isoform expression in response to TGFβ treatment in the MDA-MB-231 Claudin-low cell line.** The proteomics analysis identified phosphorylation-based binding motifs which indicated signalling pathways activated in response to TGFβ treatment, I then compared these pathways to those indicated as active in the RNA expression analysis from the same cell line and conditions.
2. **The analysis of NEDD9 interacting proteins in the Claudin-low cell line MDA-MB-231**. This analysis identified the pyruvate kinase PKM2 as a consistent NEDD9 interactor. I then investigated the known metabolic and transcriptional regulation roles of PKM2 in response to TGFβ treatment and examined the dependence of these mechanisms on NEDD9 by perturbing NEDD9 expression.
3. **Evaluation of potential translational applications for my findings using the Caldas lab PDTX biobank.** Using a combination of IHC staining for phospho-SMAD2 and c-Myc in PDTX TMAs I was able to distinguish a Claudin-low model with oncogenic TGFβ signalling activity. Using high throughput drug screens I then compared the drug sensitivity data for this model to that of a Claudin-low model without oncogenic TGFβ signalling activity to identify compounds this model was uniquely sensitive to.

6.3 Significance and implications

NEDD9 is frequently overexpressed in a range of cancer types, however contrary to previous reports, my analyses demonstrate that in breast cancer, NEDD9 expression alone is not sufficient to predict a worse prognosis. Initial investigations identified that only in breast cancer cell lines belonging to the Claudin-low group, TGFβ treatment induced an increase in expression of both NEDD9 full-length protein forms $p105^{HEF1}$ and $p115^{HEF1}$ as detected by western blot. The $p115^{HEF1}$ phospho-form is known to result from serine/threonine phosphorylation of $p105^{HEF1}$ following TGFβ treatment (Zheng and McKeown-Longo 2002), in turn triggering proteasomal degradation of NEDD9 (X. Liu et al. 2000). The

expression of NEDD9 is cell cycle-regulated with both the $p105^{HEF1}$ and $p115^{HEF1}$ phospho-forms upregulated during cell growth (Law et al. 1998). I identified both proteasomal targeting of NEDD9 and cell cycle regulation occurring in BTIC enriched cultures of the Claudin-low cell line MDA-MB-231 in response to TGFβ treatment. This indicates that in this cell population, TGFβ is key in regulating cellular proliferation and that a delicate balance between NEDD9 expression, phosphorylation and proteasomal degradation is important in this process.

Using a combination of genomic analyses to investigate TGFβ responsive gene expression signatures and proteomic analyses to look for kinase binding and substrate motifs I identified that the Wnt signalling pathway and pathways downstream of Src were both active in response to TGFβ signalling, and dependent upon NEDD9. Wnt is one of the key signalling pathways in development and stem cell regulation, so, unsurprisingly, it is frequently dysregulated in cancer. In 2010, Vermeulen and colleagues described how high Wnt signalling activity defined the CSC population in colon cancer (Vermeulen et al. 2010). In breast cancer, Wnt signalling is frequently activated and this activity correlates with overall survival (S. Li et al. 2014), and recently Wnt signalling was shown to be fundamental in triple-negative breast cancer oncogenicity (Xu et al. 2015). Although most thoroughly described in colorectal cancer, abnormal Wnt signalling is beginning to be detailed in numerous different cancer types, however, mutations in components of the Wnt signalling pathway remain significantly fewer in breast cancer when compared with colorectal adenocarcinoma and cutaneous melanoma (Zhan, Rindtorff, and Boutros 2017). Wnt signalling is typically classified as canonical, β-Catenin dependent, or non-canonical β-Catenin independent. In canonical Wnt signalling, the pathway is activated by binding of Wnt ligands to the G protein-coupled receptor Frizzled (FZD) leading to downstream accumulation of β-Catenin which translocates to the nucleus to initiate target gene expression (MacDonald, Tamai, and He 2009). In the absence of Wnt ligand binding, β-Catenin is targeted for ubiquitination and proteolysis in the cytoplasm. In the case of non-canonical β-Catenin independent Wnt signalling, Wnt ligands bind to the FZD receptor in complex with the receptor tyrosine kinases ROR1/2 leading to activation of the Rho kinase pathway and cytoskeletal rearrangement (Kikuchi et al. 2011). Non-canonical Wnt signalling pathways are far more diverse than those of canonical Wnt signalling, and despite many recent advances in research, they remain less well understood.

NEDD9 has already been shown as a Wnt signalling target in colorectal cancer by Li and colleagues, who demonstrated direct binding of the TCF/β-Catenin regulatory complex to the NEDD9 promoter (Y. Li et al. 2011), but a similar mechanism in breast cancer is yet to be described. Evidence exists of a novel mechanism by which Src increases β-Catenin expression by Cap-dependent translation (Karni et al.

2005), and more recently, Jain and colleagues demonstrated that Src inhibition leads to a reduction in c-Myc (a Wnt/β-Catenin target gene) expression and glucose metabolism (Jain et al. 2015).

Signalling pathways observed as activated in my gene expression analysis mirror those discovered as active in Claudin-low breast cancer in a recent publication (Pommier et al. 2020). In this paper, Pommier and her team describe a further sub-classification of the Claudin-low breast cancer subtype based on gene expression and methylation data. Of these three sub-classifications, the basal-like CL3 subset displays remarkable similarity in activation of signalling pathways and biological processes to that seen in my analysis, specifically in the cellular processes of proliferation and DNA repair, and pathways regulating ROS, glycolysis, Myc targets, and PI3K/AKT/mTOR signalling. Absent in the CL3 signature is TGFβ signalling, which is enriched in the differentiation/EMT CL1 sub-classification. Enrichment in the CL1 category is likely to be a reflection of TGFβ roles in tissue homeostasis and induction of EMT processes in tumours from the normal-like breast cancer subtype. Together with my data, these observations highlight that Wnt/β-Catenin signalling is frequently upregulated in cancer, and this upregulation affects metabolic processes through influencing c-Myc expression.

In addition to the role of NEDD9 hyper-phosphorylation in cell cycle dynamics and the regulation of proteolysis, I hypothesised that changes in NEDD9 phosphorylation may lead to the binding of different partner proteins, therefore uniting different signalling networks in response to TGFβ pathway activity in the Claudin-low cell line. Proteomic assays confirmed these differences, and the metabolic pyruvate kinase PKM2 was consistently observed in association with NEDD9. A combination of metabolic analyses identified an increase in both glucose metabolism and oxidative phosphorylation, however, more striking was the effect on glutamine metabolism in response to TGFβ treatment. These changes in metabolism were shown to be dependent upon NEDD9. Historically, the focus of research in cancer metabolism has been on glucose metabolism, however, more recently the importance of glutamine metabolism in cancer has been recognised (Altman, Stine, and Dang 2016). Glutamine is considered conditionally essential amino acid, in that the demand on its synthesis is increased in response to catabolic stress, such as injury or sepsis. Rapidly dividing cells such as cancer cells are also known to rely upon glutamine as a source of carbon and nitrogen to support rapid cellular growth (DeBerardinis et al. 2007). Increased Warburg metabolism, a hallmark of cancer cells, results in the majority of glucose-derived pyruvate being converted to lactate. The complete metabolism of glucose via the traditional glycolysis route, in turn creates a need to replenish intermediates of the TCA cycle for macromolecule synthesis, and this need is fulfilled by an increased reliance upon glutamine. The metabolic flux tracking experiment in chapter 4.5.3, a method of tracking metabolites derived from both glucose and glutamine, demonstrated

modest changes in glycolysis and oxidative phosphorylation of glucose metabolites in response to TGFβ treatment. However, more striking changes were observed in the consumption of glutamine into the TCA cycle, both via oxidative phosphorylation and reductive carboxylation, and these changes were demonstrated to be dependent upon both TGFβ treatment and the presence of NEDD9. Recently, in addition to its known role in catalysing the conversion of PEP to pyruvate, PKM2 has been shown to be necessary for the promotion of reductive carboxylation of glutamine metabolites (M. Liu et al. 2018). This finding strengthens my hypothesis that an association between NEDD9 and PKM2 is important in regulating PKM2 dimer-tetramer dynamics, and therefore, its catalytic activity resulting in changes in the metabolism of glucose but more importantly the incorporation of glutamine metabolites into the TCA via both oxidative phosphorylation and reductive carboxylation.

Links between PKM2 activity and both Src and β-Catenin have been previously described. Yang and colleagues described a non-metabolic role for PKM2 where β-Catenin was phosphorylated by c-Src and PKM2 for nuclear translocation and regulation of target genes (Yang et al. 2011). In 2015 Lamb and colleagues demonstrated that in a Wnt driven model of breast cancer stem cells utilised for biomarker discovery, they observed a 3.5 fold increase in mammosphere formation and increased expression of mitochondrial proteins and glycolytic enzymes including PKM2 (Lamb et al. 2015). What remains unclear, however, is the mechanism of NEDD9 interaction with PKM2 in the Claudin-low cell line MDA-MB-231. Analysis of NEDD9 post-translational modifications identified multiple kinase substrate motifs activated on NEDD9 by TGFβ treatment, but none that were specific for PKM2.

Analysis of PDTX models representing both the Claudin-low and “others” groups, illustrated that markers of signalling pathway activation from the PDTX RPPA expression dataset mirrored those observed as active in the cell line analyses. This is confirmation, in a biologically relevant platform that also takes into account breast cancer heterogeneity, that the results of the cell line analyses are meaningful. The PDTX RPPA analysis revealed that models belonging to the Claudin-low subtype had significantly higher levels of expression for proteins involved in c-Myc driven metabolic processes and Wnt/β-Catenin signalling. The reliance on c-Myc driven metabolic processes in this subtype could be harnessed for use in drug treatment response in patients. As previously mentioned, it is possible to image FOXM1 driven changes in metabolism using magnetic resonance spectroscopy (MRS), by measuring the exchange of ^{13}C from labelled pyruvate to endogenous lactate, a reaction that is catalysed by LDHA (Ros et al. 2020).

Stratification of PDTX models by expression of phospho-SMAD2 (pSMAD2) as a marker of TFGβ signalling activity allowed for comparison of high-throughput drug screen data in two models belonging

to the Claudin-low group, one with high pSMAD2 levels and one with low pSMAD2 expression. The Claudin-low model with high TGFβ signalling activity (STG316T) demonstrated sensitivity to drugs targeting the PI3K/AKT/mTOR pathway, the Wnt/β-Catenin pathway and drugs targeting cell cycle regulation, whereas the model with lower TGFβ signalling activity (STG139M) was sensitive to drugs targeting IGFR1/ALK ERBB2 and PI3K signalling. Further investigation of the Claudin-low/TGFβ high PDTX model STG316T showed higher expression of the Wnt/β-Catenin target c-Myc than the Claudin-low/TGFβ low PDTX model STG139M in both the RPPA and IHC analyses.

Sensitivity to the BET inhibitor JQ1 (which is also known to target Wnt/β-Catenin signalling) was observed in the Claudin-low/TGFβ high PDTX model STG316T, and this sensitivity was not seen in PDTX models with NEDD9 copy number loss, indicating that mechanisms leading to the sensitivity in STG316T may be NEDD9 mediated. In 2015, da Motta and his team published that in the MDA-MB-231 cell line JQ1 showed profound effects on mammosphere growth by modulating the expression of a group of hypoxia-responsive genes, but also found that Myc expression alone was not predictive of JQ1 sensitivity (da Motta et al. 2017). Indeed, the results of the high-throughput drug screen analysis in chapter 5.7 demonstrated no significant correlation between c-Myc expression and sensitivity to JQ1. This supports my hypothesis that the dual marker of TGFβ signalling activity and c-Myc expression identifies a subgroup of Claudin-low tumours that are sensitive to this drug.

Finally, our previous work resulting in the discovery of NEDD9 as a key regulator in Claudin-low oncogenic TGFβ signalling found that Rho/Actin/SRF pathway activity was central to this mechanism (Bruna et al. 2012). Recently, He and colleagues demonstrated that in the breast cancer cell line MCF7, β-Catenin physically interacts with the MRTF-A promoter following nuclear localisation by Rho/Actin signalling, resulting in direct regulation of MRTF-A expression (He et al. 2018). This mechanism also influenced the expression of genes involved in the regulation of cellular proliferation and differentiation, including the TGFβ induced gene CCN1. Here the authors used lithium chloride to artificially increase levels of β-Catenin in the MCF7 breast cancer cell line, a model with low endogenous β-Catenin expression. This study provides an important link between my observed findings in this thesis and our previous publication.

6.4 Scientific contribution

These data suggest a novel mechanism by which oncogenic TGFβ signalling regulates cellular proliferation and self-renewal via β-Catenin/c-Myc regulation of altered metabolism in the Claudin-low breast cancer subtype, a process which is dependent upon the scaffolding protein NEDD9 and the metabolic enzyme PKM2. Breast cancers belonging to the Claudin-low subtype generally fall into the triple-negative group (ER, PR and HER2 negative), and are typically more aggressive than other subtypes. The identification of further stratification within the Claudin-low subgroup suggests that patients with NEDD9 dependent oncogenic TGFβ signalling would classify with the CL3 group identified by Pommier and her team, a group with more aggressive basal features (Pommier et al. 2020). Additionally, the use of PDTX models in this thesis strengthens the impact of the analysis, since as previously discussed these models more faithfully represent breast cancer heterogeneity than cell line models.

Together, these data suggest that a combined biomarker of TGFβ signalling and c-Myc expression may be useful in identifying a subset of Claudin-low breast cancer patients who would be sensitive to inhibition of Wnt/β-Catenin, particularly by the use of JQ1. Additionally, due to the dependence of these tumours on c-Myc driven glutamine dependent metabolic processes, the use of LDH-catalyzed hyperpolarized ^{13}C MRI may be useful for treatment monitoring in these patients.

The work contained in this thesis represents a small step towards personalised breast cancer medicine by identifying a small group of breast cancer patients who may benefit from targeted therapy. Even though this patient group is small, they are patients with aggressive disease who are more likely to relapse or metastasise.

6.5 Study limitations

This thesis suggests an interesting mechanism for the regulation of TGFβ dependent oncogenic activity in a subset of Claudin-low breast tumours. The mechanism, first observed in cell lines, was validated in PDTX models which are known to better represent tumour heterogeneity than cell lines, however, the analysis is limited by the number of cell lines and PDTX models used. To assess whether a true correlation between pSMAD2/c-Myc expression and TGFβ oncogenic effects exists it would first be

necessary to expand the IHC and drug screen analysis to include the entire PDTX biobank. Further experiments are also required to validate the signalling mechanisms observed. Unfortunately, the ability to carry out these analyses during my PhD was severely hampered by the SARS-CoV-2 coronavirus pandemic of 2020/2021.

6.6 Hypothesis and key future experiments

Figure **6.1** outlines the key published findings previously discussed in this thesis, along with my proposed model of the mechanism of TGFβ oncogenic activity in the Claudin-low subtype. This figure describes how my findings relate to those of the featured publications, in the areas of; NEDD9 phosphorylation functional consequences, PKM2 regulation of glycolysis and oxidative phosphorylation, PKM2/β-Catenin interaction and c-Myc expression, β-Catenin regulation of NEDD9 expression and finally, Rho-Actin/β-Catenin regulation of cellular proliferation. What is unclear however, is whether all of these mechanisms are concurrently at-play and if so, how are they regulated by TGFβ signalling and NEDD9 phosphorylation in the Claudin-low breast cancer subtype?

In order to answer this question, the three main areas to be addressed are;

1. What is the nature of the interaction between NEDD9 and PKM2, and does this interaction affect PKM2 dimer/tetramer dynamics as hypothesised?
2. What is the function of nuclear PKM2, does it play a role in β-Catenin mediated gene expression programs as observed by Yang *et al*?
3. What is the mechanism of Wnt/β-Catenin signalling, is it regulated by non-canonical Src phosphorylation of β-Catenin as the data suggests?

To address question 1, I would perform metabolic tracking assays using the commercial NEDD9 siRNA, due to the limitations of the short-hairpin cell lines used in the initial investigations. Additionally, *in vitro* mammosphere assays with restricted glutamine concentrations, followed by second generation assays to quantify BTIC numbers would indicate glutamine dependence if present. These experiments would clarify the role of the PKM2/NEDD9 interaction on the regulation of metabolic reprogramming, however to further investigate the role of PKM2 in the Claudin-low subtype it is also important to understand PKM2 non-metabolic activity. To address this question (2) mammosphere cultures treated with commercial siRNA towards NEDD9 and PKM2 would be assayed for expression of the Wnt signalling target gene c-Myc, along with BTIC assays to quantify proliferation and self-renewal.

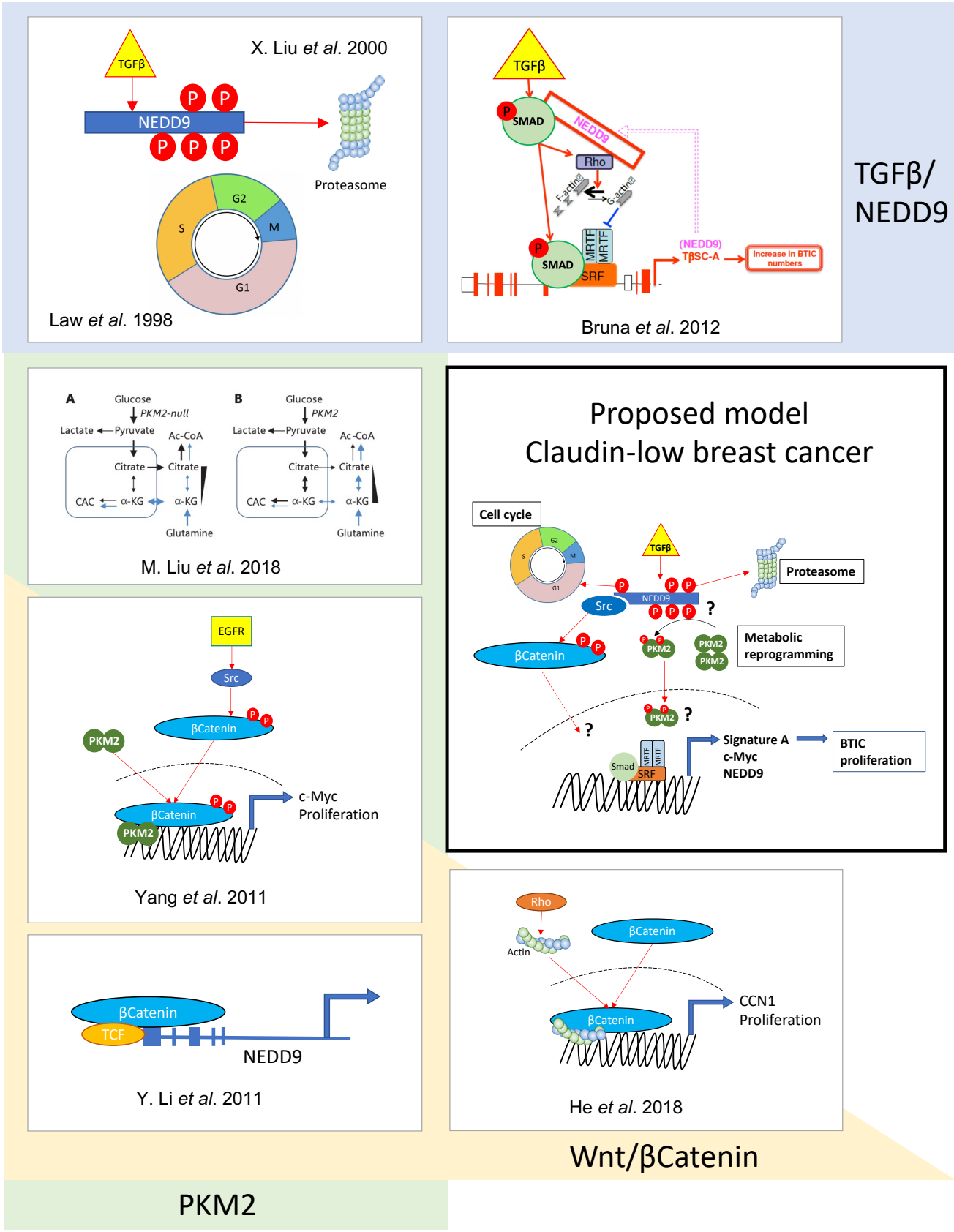


Figure 6.1. Published mechanisms and proposed hypothesis.
Oncogenic TGFβ signalling regulates cellular proliferation and self-renewal via β-Catenin/c-Myc regulation of altered metabolism in the Claudin-low breast cancer subtype, a process which is dependent upon the scaffolding protein NEDD9 and the metabolic enzyme PKM2.

Finally, understanding the mechanism of TGFβ mediated Wnt/β-Catenin signalling in the Claudin-low subtype poses the most translationally relevant question. As discussed previously, Wnt/β-Catenin signalling is one of the key pathways in development and stem cell regulation, and is frequently dysregulated in cancer. The major role of Wnt/β-Catenin signalling in cancer makes it an attractive target for cancer therapy, and there are several compounds currently in clinical trials targeting different parts of the signalling pathway (Jung and Park 2020). The data in this thesis suggest that TGFβ mediated Wnt/β-Catenin signalling in the Claudin-low subtype may be occurring via the less well understood non-canonical pathway, and a PDTX model belonging to this subtype demonstrated sensitivity to the drug JQ1, a known Wnt inhibitor.

To explore this question I would first use commercial inhibitors of canonical and non-canonical Wnt signalling pathways in cell line assays and assess expression of signature A genes, c-Myc and NEDD9. I would also perform second generation mammosphere assays to evaluate self-renewal and proliferation in the presence of these inhibitors. I would also assess for changes in NEDD9 protein expression and post-translational modification induced by Wnt inhibitor treatment.

I would then expand the IHC TMA panel to include additional Claudin-low PDTX models, then stratify these models on the basis of their p-Smad2 and c-Myc expression. Using this expanded panel I would perform *in vivo* PDTX studies comparing Claudin-low models with and without oncogenic TGFβ activity. Following drug tolerability and dose escalation studies, I would treat the mice implanted with PDTX tumours from both subtypes with JQ1, and monitor tumour growth both by manual palpation and MRS imaging using ^{13}C labelled pyruvate. It will also be possible to perform mass spectrometry metabolic tracking experiments in these models by administering ^{13}C labelled pyruvate and ^{13}C labelled lactate immediately before sacrifice and collection of the tumour.

I would also like to assess whether PDTX models with induced resistance to PI3K inhibitors lose their sensitivity to JQ1 treatment, as from my analysis the PI3K/AKT pathway appears to be important in regulating Wnt/β-Catenin signalling in this breast cancer subtype. This would be performed by generating resistance models by long-term PI3K inhibitor dosing of PDTX implanted mice, followed by high-throughput drug screens on explanted tumours.

7. Methods

For consumable information see resources chapter 8.

1. NEDD9 Immunoprecipitation

Mammosphere cell lysates were quantified using the BCA protein assay (ThermoFisher Scientific), equal quantities were added to 100 µl of magnetic beads (ThermoFisher Scientific) prebound with 10 µg HEF1 mouse monoclonal antibody (Abcam), 10 µg Smad2/3 Smad2/3 goat polyclonal antibody (Santa Cruz Biotechnology) and 10 µg IgG control antibody (Cell Signalling Technology), and immunoprecipitation (IP) was conducted overnight at 4°C

Permanent crosslinking of the antibody to the magnetic beads, by amine-to-amine coupling, was performed using the BS^3 (Bis[sulfosuccinimidyl] suberate) kit (ThermoFisher Scientific). Briefly, 2 mg BS_3 (one easy use vial) was diluted in 699 µl fresh conjugation buffer. 100 µl of overnight incubated antibody-bound beads were washed twice in fresh conjugation buffer and re-suspended in 500 µl BS_3 solution and incubated at room temperature for 30 minutes with mixing. Cross-linked beads were then quenched for 15 mins at room temperature by adding 25 µl 1 M Tris-HCl, pH 7.4. The beads were ready for use after three washes in IP wash buffer.

Mass spectrometry (MS) was performed using an LTQ Velos-Orbitrap MS (ThermoFisher Scientific)

2. RIME

RIME (Rapid Immunoprecipitation Mass spectrometry of Endogenous proteins) was performed using the Mohammed method as published (Mohammed et al. 2013, 2016). Mammospheres were washed and resuspended in 40 ml PBS. They were then cross-linked by adding 18% formaldehyde (EM grade; Polysciences inc.) to a final concentration of 1% for 20 min at room temperature. Crosslinking was quenched by adding glycine to a final concentration of 0.125 M for 20 minutes at room temperature. The cells were washed twice with ice-cold PBS, once with ice-cold PBS plus protease inhibitors and the

pellets were snap-frozen. Sonication was performed in a water bath sonicator (Diagenode Bioruptor) for 30 seconds on, 30 seconds off for 15 minutes total sonication time.

RIME IP was performed using 15 µg HEF1 mouse monoclonal antibody (Abcam), or 15 µg Smad2/3 goat polyclonal antibody (Santa Cruz Biotechnology) and 15 µg IgG control antibody (Cell Signalling Technology).

Bead processing and MS using the LTQ Velos-Orbitrap MS (ThermoFisher Scientific) were performed by the proteomics core.

3. Phospho-mass spectrometry with TiO enrichment

Samples were prepared using the NEDD9 immunoprecipitation protocol (as above 1). TiO2 enrichment for phospho-peptides was performed using the Pierce TiO_2 Phosphopeptide Enrichment and Clean-up Kit (ThermoFisher Scientific) as per the manufacturer's recommendations by the proteomics core.

Mass spectrometry and data processing was performed using an LTQ Velos-Orbitrap MS (ThermoFisher Scientific)

4. TMT-RIME

MDA-MB-231 cells were grown as mammosphere cultures as described (methods 9) The media was replaced with PBS containing 1% FA (ThermoFisher Scientific) and crosslinked for 10 min. Crosslinking was quenched by adding glycine to a final concentration of 0.1 M. The cross-linked cell pellets were lysed using a RIPA lysis buffer (50mM Tris, 150mM NaCl, 1% Igepal, 0.5% Na Deoxycholate, 0.1% SDS) including protease inhibitors. NEDD9 immunoprecipitation was performed using 50 µl of Dynabeads® Protein A (Invitrogen) and 5 µg of NEDD9 antibody (Abcam), followed by ten ice-cold RIPA washes (50mM HEPES-KOH, 500mM LiCl, 1mM EDTA, 1% NP 40, 0.7% Na Deoxycholate) and three 100mM AMBIC (ammonium bicarbonate) washes. The following steps for trypsin digestion, TMT labelling, LC-MS analysis, data processing and bioinformatics analysis were performed by the CRUK-CI

proteomics and bioinformatics core using the method as published by Papachristou et al (Papachristou et al. 2018)

Sample dissolution, TMT labelling and Reverse-Phase fractionation

Cell pellets were resuspended in lysis buffer containing 100mM Triethylammonium bicarbonate (TEAB, Sigma), 0.1% SDS (Sigma) followed by heating at 90°C for 5 min and probe sonication (Active motif). Protein concentration was estimated using Bradford assay according to the manufacturer's instructions (BIO-RAD-Quick start). 90ug of total protein was reduced with 2ul of 50mM tris-2-carboxymethyl phosphine (TCEP, Sigma) for 1 hour at 60°C followed by alkylation with 1ul of 200mM methyl methanethiosulfonate (MMTS, Sigma) for 10min at room temperature (RT). Then protein samples were digested overnight at 37°C using trypsin (ThermoFisher Scientific) solution at ratio protein/trypsin ~ 1:30. The next day, protein digest was labelled with the TMT-10plex reagents (ThermoFisher Scientific) for 1 hour. The reaction was quenched with 8 µL of 5% hydroxylamine (ThermoFisher Scientific) for 15 min at room temperature (RT). All the samples were mixed and dried with a speed vac concentrator (Eppendorf). The dry TMT mix was fractionated on a Dionex Ultimate 3000 system at high pH using the X-Bridge C18 column (3.5 µm, 2.1x150mm, Waters) with a 90min linear gradient from 5% to 95% acetonitrile contained 20mM ammonium hydroxide (Fisher Scientific) at a flow rate of 0.2 ml/min. Peptides fractions were collected between 20-55 minutes (95% of each fraction were taken for phospho analysis and remaining were used for Full proteome) and were dried using a speed vacuum concentrator. Fractions for phospho analysis were enriched for phosphopeptides using High-Select™ Fe-NTA Phosphopeptide Enrichment Kit (ThermoFisher Scientific) as per manufacturer instructions. Dried fractions of Full Proteome and Phosphoproteome were reconstituted in 0.1% formic acid (ThermoFisher Scientific) for liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis.

LC-MS/MS

Peptide fractions were analysed on a Dionex Ultimate 3000 system coupled with the nano-ESI source Fusion Lumos Orbitrap Mass Spectrometer (ThermoFisher Scientific). Peptides were trapped on a 100µm ID X 2 cm microcapillary C18 column (5µm, 100A) followed by 2h elution using 75µm ID X 25 cm C18 RP column (3µm, 100A) with 5–45% acetonitrile gradient in 0.1% formic acid at 300nl/min flow rate. In each data collection cycle, one full MS scan (380–1,500 m/z) was acquired in the Orbitrap (120K resolution, automatic gain control (AGC) setting of 3×105 and Maximum Injection Time (MIT) of 100

ms). The subsequent MS2 was conducted with a top speed approach using a 3-s duration. The most abundant ions were selected for fragmentation by collision-induced dissociation (CID). CID was performed with a collision energy of 35%, an AGC setting of 1×104, an isolation window of 0.7 Da, an MIT of 35ms. Previously analysed precursor ions were dynamically excluded for 45s. During the MS3 analyses for TMT quantification, precursor ion selection was based on the previous MS2 scan and isolated using a 2.0 Da m/z window. MS2–MS3 was conducted using sequential precursor selection (SPS) methodology with the top10 settings. HCD was used for the MS3, it was performed using 55% collision energy and reporter ions were detected using the Orbitrap (50K resolution, an AGC setting of 5×104 and MIT of 86 ms).

Data processing

The Proteome Discoverer 2.1. (ThermoFisher Scientific) was used for the processing of CID tandem mass spectra. The SequestHT search engine was used and all the spectra were searched against the Uniprot Homo sapiens FASTA database (taxon ID 9606 - Version February 2017). All searches were performed using as a static modification TMT6plex (+229.163 Da) at any N-terminus and on lysines and Methylthio at Cysteines (+45.988Da). Methionine oxidation (+15.9949Da) and Deamidation on Asparagine and Glutamine (+0.984) were included as dynamic modifications. Mass spectra were searched using precursor ion tolerance 20 ppm and fragment ion tolerance 0.5 Da. For peptide confidence, 1% FDR was applied and peptides uniquely matched to a protein were used for quantification.

5. RNA-Seq

MDA-MB-231 mammosphere cultures were prepared as methods 10, under the following treatment conditions: vehicle control (DMSO), 24 hours with 100 nM TGFβ, or 100 nM TGFβ plus 10 μM TGFβ inhibitor SB431542. NEDD9 expression was perturbed with the use of commercial SiRNA (Dharmacon) to NEDD9 with a non-targeting scramble pool as a control. siRNA treatment was carried out at day zero at the point the mammosphere cultures were generated, and TGFβ/TGFβ inhibitor treatment was added after 24 hours in culture. Total RNA was collected at 48 hours post-experiment initiation using the Qiagen miRNeasy kit [methods 13] this time point was selected to enable maximal SiRNA activity and sufficient

TGFβ/TGFβ inhibitor treatment to allow for transcriptional activation and detection of downstream effects.

Libraries for Illumina RNA-seq sequencing were prepared using TruSeq® Stranded mRNA Library Prep kit (20020594, Illumina) as per manufacturer's instructions, using an input of 500 ng total RNA per sample. Sample quality and quantity were assessed using the TapeSTation 4200 (Agilent). After 12 cycles of PCR used at the Enrichment of DNA Fragments step, all libraries were quantified using KAPA Library Quantification Kit Illumina ROX Low (Cat-ID, KK4873, KAPA Biosystems) and normalised to 10nM. Libraries were then pooled in equal volumes and pools were used for clustering on the NovaSeq6000 sequencing platform (Illumina) according to the manufacturer's instructions. Sequencing was performed using 100 bp paired-end (PE) reads to generate on average 50 million total reads per library, sufficient depth to allow for differential detection of NEDD9 transcript isoforms..

Data processing

One pool of 42 libraries sequenced on NovaSeq with paired-end and 100 bp long reads. On an average, each sample has about 40 million reads. Each sample group has 3 or 4 replicates. Reads aligned against human genome assembly GRCh38 (hg38) using STAR alignment tool (version 2.7.2) (PMID: 23104886). Reads are counted against genes using featureCounts from Subread package (version 1.5.2) (PMID: 23558742). For differential expression analysis DEseq2 (version 1.26.0) (PMID: 25516281) package used.

6. SILAC Labelling of Cell Lines for RIME

For RIME experiments, MDA-MB-231 and HCC1954 cells were grown in R/K-deficient SILAC DMEM and RPMI (ThermoFisher Scientific) and 10% dialyzed serum (Sigma-Aldrich) and supplemented with 800 µM L-Lysine $^{13}C_6{}^{15}N_2$Hydrochloride and 482 µM L-Arginine $^{13}C_6{}^{15}N_4$ hydrochloride (Sigma-Aldrich) for "heavy"-labelled media, 800 µM L-Lysine $^{13}C_6{}^{14}N_2$Hydrochloride and 482 µM L-Arginine $^{13}C_6{}^{14}N_4$ hydrochloride for "medium"-labelled media, and 800 µM L-Lysine $^{12}C_6{}^{14}N_2$-Hydrochloride and 482 µM L-Arginine $^{12}C_6{}^{14}N_4$ hydrochloride for "light"-labelled media. Cell lines were maintained in heavy, medium and light media for a period of at least two weeks and tested for incorporation of labelled amino acids prior to experiments being performed.

For mammosphere assays stably labelled cells were cultured in R/K-deficient SILAC DMEM:F12 (ThermoFisher Scientific) as outlined below.

For treatments “light” labelled cells were treated with TGFβ (100nM), “medium” labelled cells with the vehicle and “Heavy” labelled cells with the TGFβ inhibitor SB-431542 (10μM) for 3 hours at 37°C prior to collection as outlined above.

7. Metabolic tracking

Polar metabolite extraction

After labelling, cell plates were put on ice and washed twice with ice-cold PBS. Cell metabolism was quenched using 500 μL cold methanol containing 1 μL of 1 mM scyllo-inositol (1 nmol per sample) added directly to the cells in the 6-well plate. Cells were scraped into 2 mL Eppendorf tubes. A further 250 μL methanol was used to wash the well and scrape any remaining cells into the same tube. 250 μL ultrapure water and 250 μL chloroform were added to the cell pellet to give 1:3:1 chloroform:methanol:water. Cell samples were vortexed for 15 mins and incubated for 1 hour at 4°C with periodic sonication. The samples were spun for 20 mins at 4°C at 16 000 rpm. The supernatant was transferred into a new 2 mL Eppendorf and dried in a speed vac. The remaining cell pellet was re-extracted using 200 μL methanol and 100 μL ultrapure water. The samples were vortexed for 15 mins and incubated for 1 hour at 4°C with periodic sonication. The samples were spun for 20 mins at 4°C at 16 000 rpm. The supernatant was transferred into the same tube as the previously dried supernatant and dried in a SpeedVac. Once dry, the dried supernatant was resuspended in 50 μL chloroform, 150 μL methanol and 150 μL ultrapure water (1:3:3), vortexed and spun at 4°C for 20 mins at 16 000 rpm. The polar phase was transferred into GC vial inserts and dried in a SpeedVac, the lower lipid phase was discarded. Cell metabolites were derivatised before GC-MS analysis.

Derivatisation for GC-MS

For derivatisation, GC vial inserts containing extracted metabolites were washed with 60 μL methanol containing 1 μL 5 mM L-Norleucine and dried in a speed vac. The vial inserts were washed again with 60 μL methanol and dried. 20 μL 20 mg/mL methoxyamine hydrochloride in pyridine was added to each insert using a glass positive displacement pipette. The samples were vortexed briefly and incubated at room temperature overnight. 20 μL fresh BSTFA with TCMS was added to each sample and the samples were incubated at room temperature for an hour.

Protein quantification

BCA was used to quantify protein samples. Dried protein pellets were dissolved in Tris-SDS buffer (62.5 mM Tris, 2% SDS, pH6.8). If samples did not dissolve when incubated at 90°C for an hour, they were briefly sonicated, in a sonicating water bath. Samples were incubated at 90°C for a further 10 mins. The BCA protein quantification kit was used. Reagent A was mixed with reagent B at a ratio of 50:1. 4 µL of protein solution was added to 70 µL of BCA reagent. The samples were incubated at 37°C for 30 mins and absorbance at 562 nm was measured and used to calculate the concentration of protein in each sample.

GC-MS analysis and data processing

Samples were run by GS-MS using an Agilent GC-MSD (7890A-5975C). Data analysis was performed using Agilent MassHunter Quantitative Analysis software.

8. Short hairpin stable cell lines

Oligonucleotide sequences below: from within the NEDD9 coding sequence were cloned into the MSCV-miR30 plasmid. Viruses were generated and used to infect the MDA-MB-231 cell line and positive clones were selected by puromycin resistance (1µg/ml).

NEDD9 oligonucleotide sequences for MSCV-miR30 cloning

Forward

5’TCGAGAAGGTATATTGCTGTTGACAGTGAGCGAAGGAAAGGGATGGTGTTTATGTAGTGAA GCCACAGATGTACATAAACACCATCCCTTTCCTGTGCCTACTGCCTCGG3’

Reverse

5’AATTCCGAGGCAGTAGGCACAGGAAAGGGATGGTGTTTATGTACATCTGTGGCTTCACTACA TAAACACCATCCCTTTCCTTCGCTCACTGTCAACAGCAATATACCTTC3’

9. Mammosphere Assay (M1)

Cells were processed to single-cell suspension from attached culture by typsinisation (0.05% trypsin, 0.53 mM EDTA-4Na; Invitrogen) and twice washed in PBS. A further wash in supplemented DMEM:F12

media was followed by filtration through a 40µm cell strainer. Single cells were plated on ultra-low attachment plates or flasks (Corning) at a density of either 50,000 cells/ml or 100,000 cells/ml for larger-scale experiments in DMEM:F12 media supplemented with B27 (ThermoFisher Scientific), 20 ng/ml recombinant human FGF (ThermoFisher Scientific) and 20 ng/ml recombinant human EGF (ThermoFisher Scientific).

10. Immunoblotting

Cells were collected in protein lysis buffer containing 1 x Halt™ Phosphatase Inhibitor Cocktail (ThermoFisher scientific). Proteins were separated by SDS-PAGE under reducing conditions using 8% or 10% Tris-HCl gels. Electrophoretically separated proteins were transferred to nitrocellulose membranes using the Invitrogen iBlot® dry blotting system. After incubation in blocking buffer: 5% non-fat milk in PBS/0.05% Tween-20, membranes were incubated in relevant primary antibody diluted in blocking buffer. HRP conjugated anti-rabbit and anti-mouse secondary antibodies (DAKO) were used to allow for chemiluminescent detection of target proteins.

11. Cellular fractionation

Cell pellets from seven day first-generation mammosphere cultures were re-suspended in Fractionation Buffer 1 (10 mM Hepes pH 7.9, 10 mM KCl, 1.5 mM MgCl2, .34 M Glycerol, 1 mM DTT, 0.1% TritonX-100; 106 cells in 40 µl) and incubated on ice for 10 minutes. The soluble fraction was collected as the supernatant after spinning at 1,300 rcf for 10 minutes at 40C and the pellet was resuspended in Fractionation Buffer 2 (3 mM EDTA, 0.2 mM EGTA, 1 mM DTT; 106 cells in 40 µl) and incubated on ice for 30 minutes. The nuclear soluble fraction was collected as the supernatant after spinning at 1,900 rcf for 10 minutes at 4oC and the pellet representing the insoluble nuclear (chromatin) fraction was re-suspended in 2x Laemmli Buffer (Bio-rad) (106 cells in 40 µl).

12. RNA extraction

RNA extraction from mammosphere culture and adherent cell lines was performed using the Qiagen miRNeasy kit following the manufacturer's protocols.

13. Cell culture

For a full list of cell lines used please refer to the resources chapter 8. Mammospheres were cultured in either Mammary epithelial basal medium (MEBM) (Clonetics) or DMEM:F12 1:1 (Sigma Aldrich) supplemented with B27 (ThermoFisher Scientific), 20 ng/ml recombinant human FGF (ThermoFisher Scientific) and 20 ng/ml recombinant human EGF (ThermoFisher Scientific).

14. RNA-seq analysis for PDTX

Prior to alignment, sequencing quality was enforced using Trim Galore! (v0.4.2; http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Then, reads were aligned to a combined human (hg19) and mouse (mm10) reference genome using STAR (v2.5.2b). Reads were then assigned to genes using featureCounts (v1.5.2) to give counts, whereby the alignment score is used to distinguish reads as being sourced from human or mouse. Genes not expressed in any sample were removed from the dataset. The expression value of a given gene was calculated from the counts via the following transformation, log(TMM(counts)+0.5), following which batch correction, using the limma package, was performed to account for experimental batches. ER status, which was inferred from the expression, and sample type (primary vs PDX biopsy) were considered when performing the batch correction.

To give the iC10 and PAM50 classification, the counts from above were joined with those of the TCGA-BRCA dataset. All counts were then transformed using the Voom function from the limma package alongside library size normalisation. Only the human reads were considered towards a given samples library size. Batch correction was then performed, both to account for experimental batches within our biobank and those between our biobank and the TCGA dataset. Again, ER status, which was inferred from the expression for our samples and given for TCGA, as well as sample type (primary vs PDX biopsy) were considered when performing the batch correction. The biomaRt package was then used to map gene Ensembl IDs to HUGO and Entrez IDs, following which the iC10 and PAM50 classifications were generated using the iC10 and Genefu packages respectively.

15. RPPA

Tumour or cell lysates were serially diluted two-fold for 5 dilutions (from undiluted to 1:16 dilution) and arrayed on nitrocellulose-coated slides in an 11x11 format. RPPA slides were stained for 305 unique antibodies. Samples were probed with antibodies by a tyramide-based signal amplification approach and visualized by DAB colourimetric reaction. Slides were scanned on a flatbed scanner to produce 16-bit tiff images. Spots from tiff images were identified and the density was quantified by Array-Pro Analyzer then by SuperCurve Rx64 3.1.1. Relative protein levels for each sample were determined by interpolation of each dilution curve from the "standard curve" (super curve) of the slide (antibody). Supercurve is constructed by a script in R. All data points were normalized for protein loading and transformed to linear value. Normalized Linear values were transformed to Log2 values and then median-centred for hierarchical clustering analysis.

QC tests were performed for each antibody staining (slide)

16. PDTC preparation for high throughput drug screens

Single-cell suspensions were generated from a 1.5 cm3 PDTX tumour using a human tumour dissociation kit and gentleMACS instrument (Milteyni) following the manufacturer's guidelines and plated in triplicates at 40.000 cells/well into 384-well black clear bottom plates. Drugs were added after 24h incubation using an Echo acoustic liquid handler (Beckman Coulter). To quantify drug responses in PDTCs, cell viability was measured 6 days post-treatment using CellTiter-Glo 3D (Promega). The intensity was normalized against positive and negative values.

8. Resources

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-β actin	Abcam	Cat#ab8227; RRID: AB_2305186
Rabbit monoclonal anti-c-Myc	Abcam	Cat# ab32072, RRID:AB_731658
Mouse monoclonal anti-p44/42 ERK1/2	Cell Signalling Technology	Cat#4377, RRID: AB_331775
Rabbit polyclonal anti-ERK1/2	Cell Signalling Technology	Cat#9102, RRID: AB_330744
Mouse monoclonal anti-HEF1	Abcam	Cat#ab18056; RRID: AB_2149476
Rabbit polyclonal anti- Histone H3	Abcam	Cat#ab1791; RRID: AB_302613
Rabbit monoclonal anti-FOXM1	Cell Signalling Technology	Cat#5436: RRID: AB_10692483
Rabbit monoclonal anti-LDHA	Cell Signalling Technology	Cat#3582; [C4B5]; RRID: AB_2066887
Rabbit monoclonal anti-PKM1	Cell Signalling Technology	Cat#7067, RRID: AB_2715534
Rabbit polyclonal anti-PKM2	Abcam	Cat#ab85555, RRID: AB_10562282
Rabbit polyclonal anti-PKM2 (pSer37)	ThermoFisher Scientific	Cat#PA5-37684, RRID: AB_2554292
Rabbit polyclonal anti-PKM2 (pTyr105)	Cell Signalling Technology	Cat#3827, RRID: AB_1950369
Rabbit Normal IgG Control Antibody	Cell Signalling Technology	Cat#2729, RRID: AB_1031062
Goat polyclonal anti-Smad2/3	Santa Cruz Biotechnology	Cat#sc-6033, RRID: AB_656621
Mouse monoclonal anti-pSmad2	Cell Signalling Technology	Cat#3108, RRID: AB_490941
Polyclonal goat anti-rabbit HRP	DAKO/Agilent	Cat#P0448, RRID: AB_2617138
Polyclonal goat anti-mouse HRP	DAKO/Agilent	Cat# P0447, RRID: AB_2617137
Experimental models: Cell lines		
Human: MDA-MB-231	ATCC/CRUKCI Biorepository	Cat# HTB-22; RRID: CVCL_0062
Human: HCC1954	ATCC/CRUKCI Biorepository	Cat# CRL-2338, RRID: CVCL_1259
Human: MCF7	ATCC/CRUKCI Biorepository	Cat# HTB-22; RRID: CVCL_0031
Human: T47D	ATCC /CRUKCI Biorepository	Cat# HTB-133, RRID: CVCL_0553
Human: CAMA1	ATCC/CRUKCI Biorepository	Cat# HTB-21, RRID: CVCL_1115
Human: SKBR3	ATCC /CRUKCI Biorepository	Cat# HTB-30, RRID: CVCL_0033
Human: SKBR7	ATCC/CRUKCI Biorepository	RRID: CVCL_5218
Human:BT549	ATCC/CRUKCI Biorepository	Cat# HTB-122, RRID:CVCL_1092
Experimental Animals: Strains		
Mouse: NSG: NOD.CgPrkdcscidIl2rgtm1Wjl/SzJ	Charles River	Cat# 614; RRID: IMSR_JAX:005557
Chemicals		
Proteinase K	ThermoFisher Scientific	Cat#26260
Dynabeads Protein A IP Kit	ThermoFisher Scientific	Cat#10006D
Halt™ Protease Inhibitor Cocktail (100X)	ThermoFisher Scientific	Cat#78429
B-27™ Supplement (50X), serum free	ThermoFisher Scientific	Cat#17504044
Human FGF Recombinant Protein	ThermoFisher Scientific	Cat#PHG0021
Human EGF Recombinant Protein	ThermoFisher Scientific	Cat#PHG0313
Recombinant Human TGF-beta 1	R&D systems	Cat#240-B-010/CF
SB 431542 hydrate	Sigma-Aldrich	Cat#S4317

REAGENT or RESOURCE	**SOURCE**	**IDENTIFIER**
Biological samples		
Patient-derived xenografts (PDTX)	(Bruna et al., 2016)	Models as listed in Fig 5.2
Commercial assays		
BCA protein assay	ThermoFisher Scientific	Cat# 23225
BS^3 (Bis[sulfosuccinimidyl] suberate) kit	ThermoFisher Scientific	Cat#21580
miRNeasy kit	Qiagen	Cat#217004
L-Lactate Assay Kit	Abcam	Cat#ab65331
Pyruvate Kinase Assay Kit	Abcam	Cat#ab83432
Seahorse XF glycolysis stress test kit	Agilent	Cat#103017-100
Seahorse XF Mito stress test kit	Agilent	Cat#103015-100
Tumour dissociation kit, human	Milteyni Biotec	Cat#130-095-929
TiO_2 Phosphopeptide Enrichment kit	ThermoFisher Scientific	Cat#88301
Si RNA		
ON-TARGETplus Non-targeting Pool	Dharmacon/Horizon discovery	Cat#:D-001810-10-05
ON-TARGETplus NEDD9 siRNA	Dharmacon/Horizon discovery	Cat#:L-019466-01-0010
ON-TARGETplus PKM siRNA	Dharmacon/Horizon discovery	Cat#:L-006781-00-0010

9. Appendix

Samples for PDTX engraftment and/or genomic analysis were collected under the following participating studies at Cambridge University Hospitals NHS Foundation Trust and the University of Cambridge. All studies were approved under the East of England - Cambridge Central Research Ethics Committee. Samples labelled with the STG prefix were collected under the ABC trial.

REC reference	Study title	Study name	Study lead
08/H0308/178	Adult breast stem cells study	ABC	Professor Carlos Caldas
18/EE/0251	Personalised Breast Cancer Program	PBCP	Professor Carlos Caldas Dr Jean Abraham
07/Q0106/63	Identification and classification of circulating tumour cells and circulating nucleic acids in patients with breast cancer	DETECT	Professor Carlos Caldas Dr Emma Beddowes

10. Bibliography

Alghamdi, Saeed, Irfan Khan, Naimisha Beeravolu, Christina McKee, Bryan Thibodeau, George Wilson, and G. Rasul Chaudhry. 2016. “BET Protein Inhibitor JQ1 Inhibits Growth and Modulates WNT Signaling in Mesenchymal Stem Cells.” Stem Cell Research & Therapy 7 (February): 22.

Al-Hajj, Muhammad, Max S. Wicha, Adalberto Benito-Hernandez, Sean J. Morrison, and Michael F. Clarke. 2003. “Prospective Identification of Tumorigenic Breast Cancer Cells.” Proceedings of the National Academy of Sciences of the United States of America 100 (7): 3983–88.

Altman, Brian J., Zachary E. Stine, and Chi V. Dang. 2016. “From Krebs to Clinic: Glutamine Metabolism to Cancer Therapy.” Nature Reviews. Cancer 16 (11): 749.

Anastasiou, Dimitrios, Yimin Yu, William J. Israelsen, Jian-Kang Jiang, Matthew B. Boxer, Bum Soo Hong, Wolfram Tempel, et al. 2012. “Pyruvate Kinase M2 Activators Promote Tetramer Formation and Suppress Tumorigenesis.” Nature Chemical Biology 8 (10): 839–47.

Andreoli, Federico, Arménio Jorge Moura Barbosa, Marco Daniele Parenti, and Alberto Del Rio. 2013. “Modulation of Epigenetic Targets for Anticancer Therapy: Clinicopathological Relevance, Structural Data and Drug Discovery Perspectives.” Current Pharmaceutical Design 19 (4): 578–613.

Aquino, J. B., F. Lallemend, F. Marmigère, I. I. Adameyko, E. A. Golemis, and P. Ernfors. 2009. “The Retinoic Acid Inducible Cas-Family Signaling Protein Nedd9 Regulates Neural Crest Cell Migration by Modulating Adhesion and Actin Dynamics.” Neuroscience 162 (4): 1106–19.

Ashburner, Michael, Catherine A. Ball, Judith A. Blake, David Botstein, Heather Butler, J. Michael Cherry, Allan P. Davis, et al. 2000. “Gene Ontology: Tool for the Unification of Biology.” Nature Genetics 25 (1): 25–29.

Bahmad, Hisham F., Tarek H. Mouhieddine, Reda M. Chalhoub, Sahar Assi, Tarek Araji, Farah Chamaa, Muhieddine M. Itani, et al. 2018. “The Akt/mTOR Pathway in Cancer Stem/progenitor Cells Is a Potential Therapeutic Target for Glioblastoma and Neuroblastoma.” Oncotarget 9 (71): 33549–61.

Blackford, Andrew N., and Stephen P. Jackson. 2017. “ATM, ATR, and DNA-PK: The Trinity at the Heart of the DNA Damage Response.” Molecular Cell 66 (6): 801–17.

Bonnet, D., and J. E. Dick. 1997. “Human Acute Myeloid Leukemia Is Organized as a Hierarchy That Originates from a Primitive Hematopoietic Cell.” Nature Medicine 3 (7): 730–37.

Boyer, Laurie A., Tong Ihn Lee, Megan F. Cole, Sarah E. Johnstone, Stuart S. Levine, Jacob P. Zucker, Matthew G. Guenther, et al. 2005. “Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells.” Cell 122 (6): 947–56.

Briknarová, Klára, Fariborz Nasertorabi, Marnie L. Havert, Ericka Eggleston, David W. Hoyt, Chenglong Li, Arthur J. Olson, Kristiina Vuori, and Kathryn R. Ely. 2005. “The Serine-Rich Domain from Crk-Associated Substrate (p130cas) Is a Four-Helix Bundle.” The Journal of Biological Chemistry 280 (23): 21908–14.

Brinkman, A., S. van der Flier, E. M. Kok, and L. C. Dorssers. 2000. “BCAR1, a Human Homologue of the Adapter Protein p130Cas, and Antiestrogen Resistance in Breast Cancer Cells.” Journal of the National Cancer Institute 92 (2): 112–20.

Bruna, Alejandra, Rachel S. Darken, Federico Rojo, Alberto Ocaña, Silvia Peñuelas, Alexandra Arias, Raquel Paris, et al. 2007. “High TGFβ-Smad Activity Confers Poor Prognosis in Glioma Patients and Promotes Cell Proliferation Depending on the Methylation of the PDGF-B Gene.” Cancer Cell 11 (2): 147–60.

Bruna, Alejandra, Wendy Greenwood, John Le Quesne, Andrew Teschendorff, Diego Miranda-Saavedra, Oscar M. Rueda, Jose L. Sandoval, et al. 2012. “TGFβ Induces the Formation of Tumour-Initiating Cells in Claudinlow Breast Cancer.” Nature Communications 3: 1055.

Bruna, Alejandra, Oscar M. Rueda, Wendy Greenwood, Ankita Sati Batra, Maurizio Callari, Rajbir Nath Batra, Katherine Pogrebniak, et al. 2016. “A Biobank of Breast Cancer Explants with Preserved Intra-Tumor Heterogeneity to Screen Anticancer Compounds.” Cell 167 (1): 260–74.e22.

Cairns, Rob A., Isaac S. Harris, and Tak W. Mak. 2011. “Regulation of Cancer Cell Metabolism.” Nature Reviews. Cancer 11 (2): 85–95.

Callari, Maurizio, Dimitra Georgopoulou, Oscar M. Rueda, Abigail Shea, Alistair Martin, Agnese Giovannetti, Fatime Qosaj, et al. 2021. “Landscapes of Cellular Phenotypic Diversity in Breast Cancer Xenografts and Their Impact on Drug Response.” Nature Communications, March. https://doi.org/10.17863/CAM.65539.

Cao, Zhe, Qianjin Liao, Min Su, Kai Huang, Junfei Jin, and Deliang Cao. 2019. “AKT and ERK Dual Inhibitors: The Way Forward?” Cancer Letters 459 (September): 30–40.

Cassidy, John W., Carlos Caldas, and Alejandra Bruna. 2015. “Maintaining Tumor Heterogeneity in Patient-Derived Tumor Xenografts.” Cancer Research 75 (15): 2963–68.

Cerami, Ethan, Jianjiong Gao, Ugur Dogrusoz, Benjamin E. Gross, Selcuk Onur Sumer, Bülent Arman Aksoy, Anders Jacobsen, et al. 2012. “The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data.” Cancer Discovery 2 (5): 401–4.

Chaneton, Barbara, Petra Hillmann, Liang Zheng, Agnès C. L. Martin, Oliver D. K. Maddocks, Achuthanunni Chokkathukalam, Joseph E. Coyle, et al. 2012. “Serine Is a Natural Ligand and Allosteric Activator of Pyruvate Kinase M2.” Nature 491 (7424): 458–62.

Chaplin, Rebecca, Lyna Thach, Morley D. Hollenberg, Yingnan Cao, Peter J. Little, and Danielle Kamato. 2017. “Insights into Cellular Signalling by G Protein Coupled Receptor Transactivation of Cell Surface Protein Kinase Receptors.” Journal of Cell Communication and Signaling 11 (2): 117–25.

Choi, Jinhyuk, Dong-Il Kim, Jinkyoung Kim, Baek-Hui Kim, and Aeree Kim. 2016. “Hornerin Is Involved in Breast Cancer Progression.” Journal of Breast Cancer 19 (2): 142–47.

Christofk, Heather R., Matthew G. Vander Heiden, Marian H. Harris, Arvind Ramanathan, Robert E. Gerszten, Ru Wei, Mark D. Fleming, Stuart L. Schreiber, and Lewis C. Cantley. 2008. “The M2 Splice Isoform of Pyruvate Kinase Is Important for Cancer Metabolism and Tumour Growth.” Nature. https://doi.org/10.1038/nature06734.

Cui, Xin-Gang, Zhi-Tao Han, Shao-Hui He, Xing-da Wu, Tian-Rui Chen, Cheng-Hao Shao, Dan-Lei Chen, et al. 2017. “HIF1/2α Mediates Hypoxia-Induced LDHA Expression in Human Pancreatic Cancer Cells.” Oncotarget. https://doi.org/10.18632/oncotarget.15266.

Curtis, Christina, Sohrab P. Shah, Suet-Feung Chin, Gulisa Turashvili, Oscar M. Rueda, Mark J. Dunning, Doug Speed, et al. 2012. “The Genomic and Transcriptomic Architecture of 2,000 Breast Tumours Reveals Novel Subgroups.” Nature 486 (7403): 346–52.

Dang, Chi V. 2010. “Rethinking the Warburg Effect with Myc Micromanaging Glutamine Metabolism.” Cancer Research 70 (3): 859–62.

David, Charles J., Mo Chen, Marcela Assanah, Peter Canoll, and James L. Manley. 2010. “HnRNP Proteins Controlled by c-Myc Deregulate Pyruvate Kinase mRNA Splicing in Cancer.” Nature 463 (7279): 364–68.

DeBerardinis, Ralph J., Anthony Mancuso, Evgueni Daikhin, Ilana Nissim, Marc Yudkoff, Suzanne Wehrli, and Craig B. Thompson. 2007. “Beyond Aerobic Glycolysis: Transformed Cells Can Engage in Glutamine Metabolism That Exceeds the Requirement for Protein and Nucleotide Synthesis.” Proceedings of the National Academy of Sciences of the United States of America 104 (49): 19345–50.

Dias, Kay, Anna Dvorkin-Gheva, Robin M. Hallett, Ying Wu, John Hassell, Gregory R. Pond, Mark Levine, Tim Whelan, and Anita L. Bane. 2017. “Claudin-Low Breast Cancer; Clinical & Pathological Characteristics.” PloS One 12 (1): e0168669.

Dontu, Gabriela, Muhammad Al-Hajj, Wissam M. Abdallah, Michael F. Clarke, and Max S. Wicha. 2003. “Stem Cells in Normal Breast Development and Breast Cancer.” Cell Proliferation 36 Suppl 1 (October): 59–72.

Dorssers, Lambert C. J., Nicolai Grebenchtchikov, Arend Brinkman, Maxime P. Look, Simone P. J. van Broekhoven, Danielle de Jong, Harry A. Peters, et al. 2004. “The Prognostic Value of BCAR1 in Patients with Primary Breast Cancer.” Clinical Cancer Research: An Official Journal of the American Association for Cancer Research 10 (18 Pt 1): 6194–6202.

Fan, Tengjiao, Guohui Sun, Xiaodong Sun, Lijiao Zhao, Rugang Zhong, and Yongzhen Peng. 2019. “Tumor Energy Metabolism and Potential of 3-Bromopyruvate as an Inhibitor of Aerobic Glycolysis: Implications in Tumor Treatment.” Cancers 11 (3). https://doi.org/10.3390/cancers11030317.

Fashena, Sarah J., Margret B. Einarson, Geraldine M. O’Neill, Christos Patriotis, and Erica A. Golemis. 2002. “Dissection of HEF1-Dependent Functions in Motility and Transcriptional Regulation.” Journal of Cell Science 115 (Pt 1): 99–111.

Fougner, Christian, Helga Bergholtz, Jens Henrik Norum, and Therese Sørlie. 2020. “Re-Definition of Claudin-Low as a Breast Cancer Phenotype.” Nature Communications 11 (1): 1787.

Gao, Jianjiong, Bülent Arman Aksoy, Ugur Dogrusoz, Gideon Dresdner, Benjamin Gross, S. Onur Sumer, Yichao Sun, et al. 2013. “Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal.” Science Signaling 6 (269): l1.
Gene Ontology Consortium. 2021. “The Gene Ontology Resource: Enriching a GOld Mine.” Nucleic Acids Research 49 (D1): D325–34.

Glorieux, Christophe, and Pedro Buc Calderon. 2017. “Catalase, a Remarkable Enzyme: Targeting the Oldest Antioxidant Enzyme to Find a New Cancer Treatment Approach.” Biological Chemistry 398 (10): 1095–1108.

Guo, Wei, Dong Ren, Xiuting Chen, Xiang ’an Tu, Shuai Huang, Min Wang, Libing Song, Xuenong Zou, and Xinsheng Peng. 2013. “HEF1 Promotes Epithelial Mesenchymal Transition and Bone Invasion in Prostate Cancer under the Regulation of microRNA-145.” Journal of Cellular Biochemistry. https://doi.org/10.1002/jcb.24502.

Gu, Yang, Jingjing Lu, Chen Chen, and Fei Zheng. 2019. “NEDD9 Overexpression Predicts Poor Prognosis in Solid Cancers: A Meta-Analysis.” OncoTargets and Therapy 12 (May): 4213–22.

Hammerich-Hille, Stephanie, Benny A. Kaipparettu, Anna Tsimelzon, Chad J. Creighton, Shiming Jiang, Jose M. Polo, Ari Melnick, Rene Meyer, and Steffi Oesterreich. 2010. “SAFB1 Mediates Repression of Immune Regulators and Apoptotic Genes in Breast Cancer Cells.” The Journal of Biological Chemistry 285 (6): 3608–16.

Hanahan, Douglas, and Robert A. Weinberg. 2011. “Hallmarks of Cancer: The next Generation.” Cell 144 (5): 646–74.

Hao, Lang-song, Qi Liu, Chuan Tian, Dong-xing Zhang, Bo Wang, Dong-xu Zhou, Zhao-peng Li, and Zhi-xiang Yuan. 2019. “Correlation and Expression Analysis of Hypoxia-inducible Factor 1α, Glucose Transporter 1 and Lactate Dehydrogenase 5 in Human Gastric Cancer.” Oncology Letters 18 (2): 1431–41.

Harbeck, Nadia, Frédérique Penault-Llorca, Javier Cortes, Michael Gnant, Nehmat Houssami, Philip Poortmans, Kathryn Ruddy, Janice Tsang, and Fatima Cardoso. 2019. “Breast Cancer.” Nature Reviews. Disease Primers 5 (1): 66.

He, Hongpeng, Fu Du, Yongping He, Zhaoqiang Wei, Chao Meng, Yuexin Xu, Hao Zhou, et al. 2018. “The Wnt-β-Catenin Signaling Regulated MRTF-A Transcription to Activate Migration-Related Genes in Human Breast Cancer Cells.” Oncotarget 9 (20): 15239–51.

Hennessy, Bryan T., Ana-Maria Gonzalez-Angulo, Katherine Stemke-Hale, Michael Z. Gilcrease, Savitri Krishnamurthy, Ju-Seog Lee, Jane Fridlyand, et al. 2009. “Characterization of a Naturally Occurring Breast Cancer Subset Enriched in Epithelial-to-Mesenchymal Transition and Stem Cell Characteristics.” Cancer Research 69 (10): 4116–24.

Herschkowitz, Jason I., Karl Simin, Victor J. Weigman, Igor Mikaelian, Jerry Usary, Zhiyuan Hu, Karen E. Rasmussen, et al. 2007. “Identification of Conserved Gene Expression Features between Murine Mammary Carcinoma Models and Human Breast Tumors.” Genome Biology 8 (5): R76.

Hodge, Kelly, Sara Ten Have, Luke Hutton, and Angus I. Lamond. 2013. “Cleaning up the Masses: Exclusion Lists to Reduce Contamination with HPLC-MS/MS.” Journal of Proteomics 88 (August): 92–103.

Huminiecki, Lukasz, Leon Goldovsky, Shiri Freilich, Aristidis Moustakas, Christos Ouzounis, and Carl-Henrik Heldin. 2009. “Emergence, Development and Diversification of the TGF-β Signalling Pathway within the Animal Kingdom.” BMC Evolutionary Biology 9 (1): 28.

Irtegun, Sevgi, Rebecca J. Wood, Angelique R. Ormsby, Terrence D. Mulhern, and Danny M. Hatters. 2013. “Tyrosine 416 Is Phosphorylated in the Closed, Repressed Conformation of c-Src.” PloS One 8 (7): e71035.

Ishino, M., T. Ohba, H. Sasaki, and T. Sasaki. 1995. “Molecular Cloning of a cDNA Encoding a Phosphoprotein, Efs, Which Contains a Src Homology 3 Domain and Associates with Fyn.” Oncogene 11 (11): 2331–38.

Izumchenko, Eugene, Mahendra K. Singh, Olga V. Plotnikova, Nadezhda Tikhmyanova, Joy L. Little, Ilya G. Serebriiskii, Sachiko Seo, et al. 2009. “NEDD9 Promotes Oncogenic Signaling in Mammary Tumor Development.” Cancer Research 69 (18): 7198–7206.

Jain, Shalini, Xiao Wang, Chia-Chi Chang, Catherine Ibarra-Drendall, Hai Wang, Qingling Zhang, Samuel W. Brady, et al. 2015. “Src Inhibition Blocks c-Myc Translation and Glucose Metabolism to Prevent the Development of Breast Cancer.” Cancer Research 75 (22): 4863–75.

Jiang, Shaojie, Miaofeng Zhang, Yanhua Zhang, Weiping Zhou, Tao Zhu, Qing Ruan, Hui Chen, et al. 2019. “WNT5B Governs the Phenotype of Basal-like Breast Cancer by Activating WNT Signaling.” Cell Communication and Signaling: CCS 17 (1): 109.

Jiao, Qinlian, Lei Bi, Yidan Ren, Shuliang Song, Qin Wang, and Yun-Shan Wang. 2018. “Advances in Studies of Tyrosine Kinase Inhibitors and Their Acquired Resistance.” Molecular Cancer 17 (1): 36.
Jonkers, Jos, and Anton Berns. 2002. “Conditional Mouse Models of Sporadic Cancer.” Nature Reviews. Cancer 2 (4): 251–65.

Jonkers, Jos, and Anton Berns. 2002. “Conditional Mouse Models of Sporadic Cancer.” Nature Reviews. Cancer 2 (4): 251–65.

Jung, Youn-Sang, and Jae-Il Park. 2020. “Wnt Signaling in Cancer: Therapeutic Targeting of Wnt Signaling beyond β-Catenin and the Destruction Complex.” Experimental & Molecular Medicine 52 (2): 183–91.

Karni, Rotem, Yael Gus, Yuval Dor, Oded Meyuhas, and Alexander Levitzki. 2005. “Active Src Elevates the Expression of β-Catenin by Enhancement of Cap-Dependent Translation.” Molecular and Cellular Biology 25 (12): 5031–39.

Keller, Kirstie E., Irene S. Tan, and Young-Sam Lee. 2012. “SAICAR Stimulates Pyruvate Kinase Isoform M2 and Promotes Cancer Cell Survival in Glucose-Limited Conditions.” Science 338 (6110): 1069–72.

Kersten, Kelly, Karin E. de Visser, Martine H. van Miltenburg, and Jos Jonkers. 2017. “Genetically Engineered Mouse Models in Oncology Research and Cancer Medicine.” EMBO Molecular Medicine 9 (2): 137–53.

Kikuchi, Akira, Hideki Yamamoto, Akira Sato, and Shinji Matsumoto. 2011. “New Insights into the Mechanism of Wnt Signaling Pathway Activation.” International Review of Cell and Molecular Biology 291: 21–71.

Kim, Hyun Seok, Yeo-Jin Sung, and Soonmyung Paik. 2015. “Cancer Cell Line Panels Empower Genomics-Based Discovery of Precision Cancer Medicine.” Yonsei Medical Journal 56 (5): 1186–98.

Kim, Sun-Hee, Dianren Xia, Sang-Wook Kim, Vijaykumar Holla, David G. Menter, and Raymond N. DuBois. 2010. “Human Enhancer of Filamentation 1 Is a Mediator of Hypoxia-Inducible Factor-1α–Mediated Migration in Colorectal Carcinoma Cells.” Cancer Research 70 (10): 4054–63.

Kim, Wook, Yong Seok Kang, Jin Soo Kim, Nah-Young Shin, Steven K. Hanks, and Woo Keun Song. 2008. “The Integrin-Coupled Signaling Adaptor p130Cas Suppresses Smad3 Function in Transforming Growth Factor-β Signaling.” Molecular Biology of the Cell 19 (5): 2135–46.

Kirsch, Nadine, Ling-Shih Chang, Stefan Koch, Andrey Glinka, Christine Dolde, Gabriele Colozza, Maria D. J. Benitez, Edward M. De Robertis, and Christof Niehrs. 2017. “Angiopoietin-like 4 Is a Wnt Signaling Antagonist That Promotes LRP6 Turnover.” Developmental Cell 43 (1): 71–82.e6.

Kondo, Shunsuke, Satoshi Iwata, Taketo Yamada, Yusuke Inoue, Hiromi Ichihara, Yoshiko Kichikawa, Tomoki Katayose, et al. 2012. “Impact of the Integrin Signaling Adaptor Protein NEDD9 on Prognosis and Metastatic Behavior of Human Lung Cancer.” Clinical Cancer Research: An Official Journal of the American Association for Cancer Research 18 (22): 6326–38.

Kremer, Jeff Charles, Bethany Cheree Prudner, Sara Elaine Stubbs Lange, Gregory Richard Bean, Matthew Bailey Schultze, Caitlyn Brook Brashears, Megan Deanna Radyk, et al. 2017. “Arginine Deprivation Inhibits the Warburg Effect and Upregulates Glutamine Anaplerosis and Serine Biosynthesis in ASS1-Deficient Cancers.” Cell Reports 18 (4): 991–1004.

Kretzschmar, Marcus, Jacqueline Doody, and Joan Massagu. 1997. “Opposing BMP and EGF Signalling Pathways Converge on the TGF-β Family Mediator Smad1.” Nature. https://doi.org/10.1038/39348.

Kumar, S., Y. Tomooka, and M. Noda. 1992. “Identification of a Set of Genes with Developmentally down-Regulated Expression in the Mouse Brain.” Biochemical and Biophysical Research Communications 185 (3): 1155–61.

Lamb, Rebecca, Gloria Bonuccelli, Béla Ozsvári, Maria Peiris-Pagès, Marco Fiorillo, Duncan L. Smith, Generoso Bevilacqua, et al. 2015. “Mitochondrial Mass, a New Metabolic Biomarker for Stem-like Cancer Cells: Understanding WNT/FGF-Driven Anabolic Signaling.” Oncotarget 6 (31): 30453–71.

Law, S. F., J. Estojak, B. Wang, T. Mysliwiec, G. Kruh, and E. A. Golemis. 1996. “Human Enhancer of Filamentation 1, a Novel p130cas-like Docking Protein, Associates with Focal Adhesion Kinase and Induces Pseudohyphal Growth in Saccharomyces Cerevisiae.” Molecular and Cellular Biology 16 (7): 3327–37.

Law, S. F., Y. Z. Zhang, S. J. Fashena, G. Toby, J. Estojak, and E. A. Golemis. 1999. “Dimerization of the Docking/adaptor Protein HEF1 via a Carboxy-Terminal Helix-Loop-Helix Domain.” Experimental Cell Research 252 (1): 224–35.

Law, S. F., Y. Z. Zhang, A. J. Klein-Szanto, and E. A. Golemis. 1998. “Cell Cycle-Regulated Processing of HEF1 to Multiple Protein Forms Differentially Targeted to Multiple Subcellular Compartments.” Molecular and Cellular Biology 18 (6): 3540–51.

Levy, L., and C. Hill. 2006. “Alterations in Components of the TGF-β Superfamily Signaling Pathways in Human Cancer.” Cytokine & Growth Factor Reviews. https://doi.org/10.1016/j.cytogfr.2005.09.009.

Li, Bo, and Colin N. Dewey. 2011. “RSEM: Accurate Transcript Quantification from RNA-Seq Data with or without a Reference Genome.” BMC Bioinformatics 12 (August): 323.

Li, Shuguang, Shanshan Li, Ying Sun, and Li Li. 2014. “The Expression of β-Catenin in Different Subtypes of Breast Cancer and Its Clinical Significance.” Tumour Biology: The Journal of the International Society for Oncodevelopmental Biology and Medicine 35 (8): 7693–98.

Liu, Dongbo, Dingkun Wang, Cheng Wu, Linli Zhang, Qi Mei, Guangyuan Hu, Guoxian Long, and Wei Sun. 2019. “Prognostic Significance of Serum Lactate Dehydrogenase in Patients with Breast Cancer: A Meta-Analysis.” Cancer Management and Research 11 (April): 3611–19.

Liu, Miao, Yuanyuan Wang, Yuxia Ruan, Changsen Bai, Li Qiu, Yanfen Cui, Guoguang Ying, and Binghui Li. 2018. “PKM2 Promotes Reductive Glutamine Metabolism.” Cancer Biology & Medicine 15 (4): 389–99.

Liu, X., A. E. Elia, S. F. Law, E. A. Golemis, J. Farley, and T. Wang. 2000. “A Novel Ability of Smad3 to Regulate Proteasomal Degradation of a Cas Family Member HEF1.” The EMBO Journal 19 (24): 6759–69.

Li, Y., J. H. Bavarva, Z. Wang, J. Guo, C. Qian, S. N. Thibodeau, E. A. Golemis, and W. Liu. 2011. “HEF1, a Novel Target of Wnt Signaling, Promotes Colonic Cell Migration and Cancer Progression.” Oncogene 30 (23): 2633–43.

Lu, Ruijing, Ziliang Ji, Xiaoqing Li, Qingna Zhai, Chunjuan Zhao, Zhimao Jiang, Shiqiang Zhang, Liping Nie, and Zhendong Yu. 2014. “miR-145 Functions as Tumor Suppressor and Targets Two Oncogenes, ANGPT2 and NEDD9, in Renal Cell Carcinoma.” Journal of Cancer Research and Clinical Oncology 140 (3): 387–97.

MacDonald, Bryan T., Keiko Tamai, and Xi He. 2009. “Wnt/beta-Catenin Signaling: Components, Mechanisms, and Diseases.” Developmental Cell 17 (1): 9–26.

Mace, Peter D., Yann Wallez, Małgorzata K. Dobaczewska, Jeongeun J. Lee, Howard Robinson, Elena B. Pasquale, and Stefan J. Riedl. 2011. “NSP-Cas Protein Structures Reveal a Promiscuous Interaction Module in Cell Signaling.” Nature Structural & Molecular Biology 18 (12): 1381–87.

Masoud, Georgina N., and Wei Li. 2015. “HIF-1α Pathway: Role, Regulation and Intervention for Cancer Therapy.” Acta Pharmaceutica Sinica. B 5 (5): 378–89.

Massagué, Joan. 2008. “TGFbeta in Cancer.” Cell 134 (2): 215–30.

Massagué, Joan. 2012. “TGFβ Signalling in Context.” Nature Reviews. Molecular Cell Biology 13 (10): 616–30.

Massagué, Joan, Joan Seoane, and David Wotton. 2005. “Smad Transcription Factors.” Genes & Development 19 (23): 2783–2810.

Massagué, J., and D. Padua. 2009. “Roles of TGF Beta in Metastasis.” Cell Research.

Masters, John R. 2002. “HeLa Cells 50 Years on: The Good, the Bad and the Ugly.” Nature Reviews. Cancer 2 (4): 315–19.

Matsuura, Isao, Natalia G. Denissova, Guannan Wang, Dongming He, Jianyin Long, and Fang Liu. 2004. “Cyclin-Dependent Kinases Regulate the Antiproliferative Function of Smads.” Nature 430 (6996): 226–31.

McCubrey, James A., Linda S. Steelman, Fred E. Bertrand, Nicole M. Davis, Melissa Sokolosky, Steve L. Abrams, Giuseppe Montalto, et al. 2014. “GSK-3 as Potential Target for Therapeutic Intervention in Cancer.” Oncotarget 5 (10): 2881–2911.

Miller, Donald M., Shelia D. Thomas, Ashraful Islam, David Muench, and Kara Sedoris. 2012. “C-Myc and Cancer Metabolism.” Clinical Cancer Research: An Official Journal of the American Association for Cancer Research 18 (20): 5546–53.

Minegishi, M., K. Tachibana, T. Sato, S. Iwata, Y. Nojima, and C. Morimoto. 1996. “Structure and Function of Cas-L, a 105-kD Crk-Associated Substrate-Related Protein That Is Involved in Beta 1 Integrin-Mediated Signaling in Lymphocytes.” The Journal of Experimental Medicine 184 (4): 1365–75.
Miyazono, K. 2000. “TGF-Beta Signaling by Smad Proteins.” Cytokine & Growth Factor Reviews 11 (1-2): 15–22.

Mohammed, Hisham, Christopher Taylor, Gordon D. Brown, Evaggelia K. Papachristou, Jason S. Carroll, and Clive S. D’Santos. 2016. “Rapid Immunoprecipitation Mass Spectrometry of Endogenous Proteins (RIME) for Analysis of Chromatin Complexes.” Nature Protocols 11 (2): 316–26.

Motta, L. L. da, I. Ledaki, K. Purshouse, S. Haider, M. A. De Bastiani, D. Baban, M. Morotti, et al. 2017. “The BET Inhibitor JQ1 Selectively Impairs Tumour Response to Hypoxia and Downregulates CA9 and Angiogenesis in Triple Negative Breast Cancer.” Oncogene 36 (1): 122–32.

Mullen, Alan C., David A. Orlando, Jamie J. Newman, Jakob Lovén, Roshan M. Kumar, Steve Bilodeau, Jessica Reddy, Matthew G. Guenther, Rodney P. DeKoter, and Richard A. Young. 2011. “Master Transcription Factors Determine Cell-Type-Specific Responses to TGF-β Signaling.” Cell 147 (3): 565–76.

Natarajan, M., J. E. Stewart, E. A. Golemis, E. N. Pugacheva, K. Alexandropoulos, B. D. Cox, W. Wang, J. R. Grammer, and C. L. Gladson. 2006. “HEF1 Is a Necessary and Specific Downstream Effector of FAK That Promotes the Migration of Glioblastoma Cells.” Oncogene 25 (12): 1721–32.

Neuzillet, Cindy, Annemilaï Tijeras-Raballand, Romain Cohen, Jérôme Cros, Sandrine Faivre, Eric Raymond, and Armand de Gramont. 2015. “Targeting the TGFβ Pathway for Cancer Therapy.” Pharmacology & Therapeutics. https://doi.org/10.1016/j.pharmthera.2014.11.001.

Nikonova, Anna S., Anna V. Gaponova, Alexander E. Kudinov, and Erica A. Golemis. 2014. “CAS Proteins in Health and Disease: An Update.” IUBMB Life 66 (6): 387–95.

Nim, Tri Hieu, Le Luo, Jacob K. White, Marie-Véronique Clément, and Lisa Tucker-Kellogg. 2015. “Non-Canonical Activation of Akt in Serum-Stimulated Fibroblasts, Revealed by Comparative Modeling of Pathway Dynamics.” PLoS Computational Biology 11 (11): e1004505.

Nowell, P. C. 1976. “The Clonal Evolution of Tumor Cell Populations.” Science 194 (4260): 23–28.
O’Neill, G. M., and E. A. Golemis. 2001. “Proteolysis of the Docking Protein HEF1 and Implications for Focal Adhesion Dynamics.” Molecular and Cellular Biology 21 (15): 5094–5108.

Padua, David, Xiang H-F Zhang, Qiongqing Wang, Cristina Nadal, William L. Gerald, Roger R. Gomis, and Joan Massagué. 2008. “TGFβ Primes Breast Tumors for Lung Metastasis Seeding through Angiopoietin-like 4.” Cell 133 (1): 66–77.

Papachristou, Evangelia K., Kamal Kishore, Andrew N. Holding, Kate Harvey, Theodoros I. Roumeliotis, Chandra Sekhar Reddy Chilamakuri, Soleilmane Omarjee, et al. 2018. “A Quantitative Mass Spectrometry-Based Approach to Monitor the Dynamics of Endogenous Chromatin-Associated Protein Complexes.” Nature Communications 9 (1): 2311.

Patro, Rob, Geet Duggal, Michael I. Love, Rafael A. Irizarry, and Carl Kingsford. 2017. “Salmon Provides Fast and Bias-Aware Quantification of Transcript Expression.” Nature Methods 14 (4): 417–19.

Pereira, Bernard, Suet-Feung Chin, Oscar M. Rueda, Hans-Kristian Moen Vollan, Elena Provenzano, Helen A. Bardwell, Michelle Pugh, et al. 2016. “The Somatic Mutation Profiles of 2,433 Breast Cancers Refine Their Genomic and Transcriptomic Landscapes.” Nature Communications 7 (1): 11479.

Perillo, Bruno, Marzia Di Donato, Antonio Pezone, Erika Di Zazzo, Pia Giovannelli, Giovanni Galasso, Gabriella Castoria, and Antimo Migliaccio. 2020. “ROS in Cancer Therapy: The Bright Side of the Moon.” Experimental & Molecular Medicine 52 (2): 192–203.

Perou, C. M., T. Sørlie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, C. A. Rees, J. R. Pollack, et al. 2000. “Molecular Portraits of Human Breast Tumours.” Nature 406 (6797): 747–52.

Platt, Randall J., Sidi Chen, Yang Zhou, Michael J. Yim, Lukasz Swiech, Hannah R. Kempton, James E. Dahlman, et al. 2014. “CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling.” Cell 159 (2): 440–55.

Polyak, Kornelia. 2011. “Heterogeneity in Breast Cancer.” The Journal of Clinical Investigation 121 (10): 3786–88.

Pommier, Roxane M., Amélien Sanlaville, Laurie Tonon, Janice Kielbassa, Emilie Thomas, Anthony Ferrari, Anne-Sophie Sertier, et al. 2020. “Comprehensive Characterization of Claudin-Low Breast Tumors Reflects the Impact of the Cell-of-Origin on Cancer Evolution.” Nature Communications 11 (1): 3431.

Prat, Aleix, Joel S. Parker, Olga Karginova, Cheng Fan, Chad Livasy, Jason I. Herschkowitz, Xiaping He, and Charles M. Perou. 2010. “Phenotypic and Molecular Characterization of the Claudin-Low Intrinsic Subtype of Breast Cancer.” Breast Cancer Research: BCR 12 (5): R68.

Pugacheva, Elena N., and Erica A. Golemis. 2005. “The Focal Adhesion Scaffolding Protein HEF1 Regulates Activation of the Aurora-A and Nek2 Kinases at the Centrosome.” Nature Cell Biology 7 (10): 937–46.

Qin, Shengtang, Danli Yang, Kang Chen, Haolan Li, Liqiang Zhang, Yuan Li, Rongrong Le, Xiaojie Li, Shaorong Gao, and Lan Kang. 2017. “Pkm2 Can Enhance Pluripotency in ESCs and Promote Somatic Cell Reprogramming to iPSCs.” Oncotarget 8 (48): 84276–84.

Rennoll, Sherri, and Gregory Yochum. 2015. “Regulation of MYC Gene Expression by Aberrant Wnt/β-Catenin Signaling in Colorectal Cancer.” World Journal of Biological Chemistry 6 (4): 290–300.

Reynolds, B., and S. Weiss. 1992. “Generation of Neurons and Astrocytes from Isolated Cells of the Adult Mammalian Central Nervous System.” Science. https://doi.org/10.1126/science.1553558.

Ros, Susana, Alan J. Wright, Paula D’Santos, De-En Hu, Richard L. Hesketh, Yaniv Lubling, Dimitra Georgopoulou, et al. 2020. “Metabolic Imaging Detects Resistance to PI3Kα Inhibition Mediated by Persistent FOXM1 Expression in ER+ Breast Cancer.” Cancer Cell 38 (4): 516–33.e9.

Sakai, R., A. Iwamatsu, N. Hirano, S. Ogawa, T. Tanaka, H. Mano, Y. Yazaki, and H. Hirai. 1994. “A Novel Signaling Molecule, p130, Forms Stable Complexes in Vivo with v-Crk and v-Src in a Tyrosine Phosphorylation-Dependent Manner.” The EMBO Journal 13 (16): 3748–56.

Schneider, Caroline A., Wayne S. Rasband, and Kevin W. Eliceiri. 2012. “NIH Image to ImageJ: 25 Years of Image Analysis.” Nature Methods 9 (7): 671–75.

Semenza, G. L. 2001. “HIF-1 and Mechanisms of Hypoxia Sensing.” Current Opinion in Cell Biology 13 (2): 167–71.

Shackleton, Mark, Elsa Quintana, Eric R. Fearon, and Sean J. Morrison. 2009. “Heterogeneity in Cancer: Cancer Stem Cells versus Clonal Evolution.” Cell 138 (5): 822–29.

Shagisultanova, Elena, Anna V. Gaponova, Rashid Gabbasov, Emmanuelle Nicolas, and Erica A. Golemis. 2015. “Preclinical and Clinical Studies of the NEDD9 Scaffold Protein in Cancer and Other Diseases.” Gene 567 (1): 1–11.

Shah, Sohrab P., Ryan D. Morin, Jaswinder Khattra, Leah Prentice, Trevor Pugh, Angela Burleigh, Allen Delaney, et al. 2009. “Mutational Evolution in a Lobular Breast Tumour Profiled at Single Nucleotide Resolution.” Nature 461 (7265): 809–13.

Shah, Sohrab P., Andrew Roth, Rodrigo Goya, Arusha Oloumi, Gavin Ha, Yongjun Zhao, Gulisa Turashvili, et al. 2012. “The Clonal and Mutational Evolution Spectrum of Primary Triple-Negative Breast Cancers.” Nature 486 (7403): 395–99.

Shang, Shuang, Fang Hua, and Zhuo-Wei Hu. 2017. “The Regulation of β-Catenin Activity and Function in Cancer: Therapeutic Opportunities.” Oncotarget 8 (20): 33972–89.

Shan, Naing L., Yoosub Shin, Ge Yang, Philip Furmanski, and Nanjoo Suh. 2021. “Breast Cancer Stem Cells: A Review of Their Characteristics and the Agents That Affect Them.” Molecular Carcinogenesis 60 (2): 73–100.

Siegel, Peter M., and Joan Massagué. 2003. “Cytostatic and Apoptotic Actions of TGF-Beta in Homeostasis and Cancer.” Nature Reviews. Cancer 3 (11): 807–21.

Singh, Mahendra K., Disha Dadke, Emmanuelle Nicolas, Ilya G. Serebriiskii, Sinoula Apostolou, Adrian Canutescu, Brian L. Egleston, and Erica A. Golemis. 2008. “A Novel Cas Family Member, HEPL, Regulates FAK and Cell Spreading.” Molecular Biology of the Cell 19 (4): 1627–36.

Sørlie, T., C. M. Perou, R. Tibshirani, T. Aas, S. Geisler, H. Johnsen, T. Hastie, 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): 10869–74.

Spassov, Danislav S., Ana Ruiz-Saenz, Amit Piple, and Mark M. Moasser. 2018. “A Dimerization Function in the Intrinsically Disordered N-Terminal Region of Src.” Cell Reports 25 (2): 449–63.e4.

Speranza, Maria Carmela, Véronique Frattini, Federica Pisati, Dimos Kapetis, Paola Porrati, Marica Eoli, Serena Pellegatta, and Gaetano Finocchiaro. 2012. “NEDD9, a Novel Target of miR-145, Increases the Invasiveness of Glioblastoma.” Oncotarget 3 (7): 723–34.

Sun, Hao-Ran, Shun Wang, Shi-Can Yan, Yu Zhang, Peter J. Nelson, Hu-Liang Jia, Lun-Xiu Qin, and Qiong-Zhu Dong. 2019. “Therapeutic Strategies Targeting Cancer Stem Cells and Their Microenvironment.” Frontiers in Oncology 9 (October): 1104.

Swanton, Charles. 2012. “Intratumor Heterogeneity: Evolution through Space and Time.” Cancer Research 72 (19): 4875–82.

Tikhmyanova, Nadezhda, and Erica A. Golemis. 2011. “NEDD9 and BCAR1 Negatively Regulate E-Cadherin Membrane Localization, and Promote E-Cadherin Degradation.” PloS One 6 (7): e22102.

Tikhmyanova, Nadezhda, Joy L. Little, and Erica A. Golemis. 2010. “CAS Proteins in Normal and Pathological Cell Growth Control.” Cellular and Molecular Life Sciences: CMLS 67 (7): 1025–48.

Torre, L. A., R. L. Siegel, E. M. Ward, and A. Jemal. 2016. “Global Cancer Incidence and Mortality Rates and Trends—an Update.” Cancer Epidemiology. http://cebp.aacrjournals.org/content/25/1/16?utm_source=disparities17&utm_medium=collpage3&utm_campaign=150578.

Tufegdzic Vidakovic, Ana, Oscar M. Rueda, Stephin J. Vervoort, Ankita Sati Batra, Mae Akilina Goldgraben, Santiago Uribe-Lewis, Wendy Greenwood, Paul J. Coffer, Alejandra Bruna, and Carlos Caldas. 2015. “Context-Specific Effects of TGF-β/SMAD3 in Cancer Are Modulated by the Epigenome.” Cell Reports 13 (11): 2480–90.

Vander Heiden, Matthew G., Lewis C. Cantley, and Craig B. Thompson. 2009. “Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation.” Science 324 (5930): 1029–33.

Vanhaesebroeck, B., and D. R. Alessi. 2000. “The PI3K-PDK1 Connection: More than Just a Road to PKB.” Biochemical Journal 346 Pt 3 (March): 561–76.

Vermeulen, Louis, Felipe De Sousa E Melo, Maartje van der Heijden, Kate Cameron, Joan H. de Jong, Tijana Borovski, Jurriaan B. Tuynman, et al. 2010. “Wnt Activity Defines Colon Cancer Stem Cells and Is Regulated by the Microenvironment.” Nature Cell Biology 12 (5): 468–76.

Vogel, Tanja, Sandra Ahrens, Nicole Büttner, and Kerstin Krieglstein. 2010. “Transforming Growth Factor Beta Promotes Neuronal Cell Fate of Mouse Cortical and Hippocampal Progenitors in Vitro and in Vivo: Identification of Nedd9 as an Essential Signaling Component.” Cerebral Cortex 20 (3): 661–71.

Warburg, O., F. Wind, and E. Negelein. 1927. “THE METABOLISM OF TUMORS IN THE BODY.” The Journal of General Physiology 8 (6): 519–30.

Wendt, Michael K., Tressa M. Allington, and William P. Schiemann. 2009. “Mechanisms of the Epithelial–mesenchymal Transition by TGF-β.” Future Oncology 5 (8): 1145–68.

Wendt, M. K., J. A. Smith, and W. P. Schiemann. 2009. “p130Cas Is Required for Mammary Tumor Growth and Transforming Growth Factor-β-Mediated Metastasis through Regulation of Smad2/3 Activity.” The Journal of Biological Chemistry. http://www.jbc.org/content/284/49/34145.short.

Wicha, Max S. 2014. “Targeting Self-Renewal, an Achilles’ Heel of Cancer Stem Cells.” Nature Medicine.

Wu, Dianqing, and Weijun Pan. 2010. “GSK3: A Multifaceted Kinase in Wnt Signaling.” Trends in Biochemical Sciences 35 (3): 161–68.

Xia, Limin, Wenjie Huang, Dean Tian, Hongwu Zhu, Xingshun Qi, Zheng Chen, Yongguo Zhang, et al. 2013. “Overexpression of Forkhead Box C1 Promotes Tumor Metastasis and Indicates Poor Prognosis in Hepatocellular Carcinoma.” Hepatology 57 (2): 610–24.

Xia, Limin, Ping Mo, Wenjie Huang, Lin Zhang, Ying Wang, Hongwu Zhu, Dean Tian, et al. 2012. “The TNF-α/ROS/HIF-1-Induced Upregulation of FoxMI Expression Promotes HCC Proliferation and Resistance to Apoptosis.” Carcinogenesis. https://doi.org/10.1093/carcin/bgs249.

Xia, Pu, and Xiao-Yan Xu. 2015. “PI3K/Akt/mTOR Signaling Pathway in Cancer Stem Cells: From Basic Research to Clinical Application.” American Journal of Cancer Research 5 (5): 1602–9.

Xu, Jinhua, Yinghua Chen, Dezheng Huo, Andrey Khramtsov, Galina Khramtsova, Chunling Zhang, Kathleen H. Goss, and Olufunmilayo I. Olopade. 2016. “β-Catenin Regulates c-Myc and CDKN1A Expression in Breast Cancer Cells.” Molecular Carcinogenesis 55 (5): 431–39.

Xu, Jinhua, Jenifer R. Prosperi, Noura Choudhury, Olufunmilayo I. Olopade, and Kathleen H. Goss. 2015. “β-Catenin Is Required for the Tumorigenic Behavior of Triple-Negative Breast Cancer Cells.” PloS One 10 (2): e0117097.

Yang, Weiwei, Yan Xia, Haitao Ji, Yanhua Zheng, Ji Liang, Wenhua Huang, Xiang Gao, Kenneth Aldape, and Zhimin Lu. 2011. “Nuclear PKM2 Regulates β-Catenin Transactivation upon EGFR Activation.” Nature 480 (7375): 118–22.

Yang, Weiwei, Yanhua Zheng, Yan Xia, Haitao Ji, Xiaomin Chen, Fang Guo, Costas A. Lyssiotis, Kenneth Aldape, Lewis C. Cantley, and Zhimin Lu. 2012. “ERK1/2-Dependent Phosphorylation and Nuclear Translocation of PKM2 Promotes the Warburg Effect.” Nature Cell Biology 14 (12): 1295–1304.

Ye, Feng, Yan Qiu, Li Li, Libo Yang, Fei Cheng, Hongying Zhang, Bing Wei, Zhang Zhang, Linyong Sun, and Hong Bu. 2015. “The Presence of EpCAM-/CD49f Cells in Breast Cancer Is Associated with a Poor Clinical Outcome.” Journal of Breast Cancer. https://doi.org/10.4048/jbc.2015.18.3.242.

Zhang, Chun, Teng Wang, Hao Wu, Lihua Zhang, Kan Li, Fang Wang, Yun Chen, Jian Jin, and Dong Hua. 2019. “HEF1 Regulates Differentiation through the Wnt5a/β-Catenin Signaling Pathway in Human Gastric Cancer.” Biochemical and Biophysical Research Communications 509 (1): 201–8.

Zhang, Yan, Suli Tian, Jidong Xiong, Yongxu Zhou, Hongyu Song, and Chang Liu. 2018. “JQ-1 Inhibits Colon Cancer Proliferation via Suppressing Wnt/β-Catenin Signaling and miR-21.” Chemical Research in Toxicology 31 (5): 302–7.

Zhang, Ying E. 2009. “Non-Smad Pathways in TGF-Beta Signaling.” Cell Research 19 (1): 128–39.

Zhang, Ze, Xinyue Deng, Yuanda Liu, Yahui Liu, Liankun Sun, and Fangfang Chen. 2019. “PKM2, Function and Expression and Regulation.” Cell & Bioscience 9 (June): 52.

Zhan, T., N. Rindtorff, and M. Boutros. 2017. “Wnt Signaling in Cancer.” Oncogene 36 (11): 1461–73.
Zhao, Shuo, Pengxiang Min, Lei Liu, Lin Zhang, Yujie Zhang, Yueyuan Wang, Xuyang Zhao, et al. 2019. “NEDD9 Facilitates Hypoxia-Induced Gastric Cancer Cell Migration via MICAL1 Related Rac1 Activation.” Frontiers in Pharmacology 10 (April): 291.

Zheng, Mingzhe, and Paula J. McKeown-Longo. 2002. “Regulation of HEF1 Expression and Phosphorylation by TGF-β1 and Cell Adhesion.” The Journal of Biological Chemistry 277 (42): 39599–608.

Zheng, Mingzhe, and Paula J. McKeown-Longo. 2006. “Cell Adhesion Regulates Ser/Thr Phosphorylation and Proteasomal Degradation of HEF1.” Journal of Cell Science 119 (Pt 1): 96–103.

Zhong, Jessie, Jaime B. Baquiran, Navid Bonakdar, Justin Lees, Yu Wooi Ching, Elena Pugacheva, Ben Fabry, and Geraldine M. O’Neill. 2012. “NEDD9 Stabilizes Focal Adhesions, Increases Binding to the Extra-Cellular Matrix and Differentially Effects 2D versus 3D Cell Migration.” PloS One 7 (4): e35058.

Zou, Songyun, Xiangmei Wang, Po Liu, Changneng Ke, and Shi Xu. 2019. “Arginine Metabolism and Deprivation in Cancer Therapy.” Biomedecine & Pharmacotherapie 118 (October): 109210.