Expression and Methylation of *Peroxidasin* in Breast Cancer Cell Lines

by

Jemma Lilian Falkov

(1611672)

Dissertation

Submitted in fulfilment of the requirements for the degree

Master of Science

in

Molecular and Cell Biology

in the Faculty of Science, University of the Witwatersrand, Johannesburg, South Africa

Supervisor: Demetra Mavri-Damelin

August 2023

Declaration

I, Jemma Lilian Falkov (1611672), am a student registered for the degree of Master of Science in the academic year 2021.

I hereby declare the following:

• I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.

• I confirm that the research dissertation submitted for assessment for the above degree is my own unaided work except where explicitly indicated otherwise and acknowledged.

• I have followed the required conventions in referencing the thoughts and ideas of others.

• I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.

Signature  4 August 2023

Abstract

Peroxidasin (PXDN) is a haem-containing extracellular matrix peroxidase protein which forms hypohalous acids in the presence of hydrogen peroxide (H_2O_2). The predominant role of PXDN is that of a collagen IV crosslinker within the basement membrane. Increased collagen IV deposition has been linked to tissue invasion and metastasis in breast cancer and PXDN has also been shown to assist in the process of epithelial-mesenchymal transition (EMT) in cancer. Various cancer types display dysregulated levels of *PXDN* expression including breast cancer and this dysregulation has been associated with poor prognosis. This study aimed to investigate whether DNA methylation of the *PXDN* promoter may be a mechanism through which changes in *PXDN* expression observed in breast cancer are regulated. Non-invasive MCF-7 and invasive MDA-MB-231 cells were used as models for luminal A and triple negative breast cancer (TNBC) respectively. The HEK-293 cell line was used as a non-cancerous control cell line. DNA methylation levels of the *PXDN* promoter and PXDN protein expression was investigated in these cell lines through the methods of methylation sensitive PCR (MS PCR) and immunofluorescence microscopy.

Relative levels of PXDN expression were determined through immunofluorescence microscopy. Corrected total cell fluorescence (CTCF) analysis of these images revealed the highest PXDN levels to be found within the invasive MDA-MB-231 cell line, which was double that of the MCF-7 cell line. All cell lines were treated with 10 nM β-Oestradiol, which caused an increase in PXDN expression within the MCF-7 and HEK-293 cell lines and a decrease in expression within the MDA-MB-231 cell line to half its untreated value. PXDN was found to be localised in the ECM in all three cell lines.

To elucidate the role of DNA methylation, methylation sensitive PCR (MS PCR) was performed on all three cell lines, with four primer pairs spanning a region of 1305 base pairs (bp) within the *PXDN* promoter. A region of differential methylation was found between the MDA-MB-231 and HEK-293 cell lines between 524 bp and 53 bp upstream of the transcription start site (TSS). This region was unmethylated within the MDA-MB-231 cell line and methylated within the HEK-293 cell line, which correlates with expression differences between these two cell lines and suggests this region could be of regulatory significance. The four primer pairs designed to amplify the

PXDN promoter were unable to amplify this region within the MCF-7 cell line. A heterochromatic DNA conformation or a point mutation increasing CpG content creating a thermodynamically ultra-fastened (TUF) region could be the explanation behind this phenomenon, however further research is required to elucidate the mechanism responsible.

In conclusion, PXDN shows higher expression in TNBC cells than in luminal A subtype cells. The oestrogen receptor is involved in regulating *PXDN* expression, however, different mechanisms seem to be at play between the two cell lines. The contribution of CpG methylation to this change in *PXDN* expression remains unknown, as does the nature of the interaction between the oestrogen receptors and the gene. Further research is required to clarify the mechanisms involved.

Acknowledgments

I would first like thank my supervisor Professor Demetra Mavri-Damelin, this project was her idea and design. Apart from her wealth of knowledge and wisdom, some of which she has imparted on me, her patience and kindness have gently allowed me to discover and refine my own abilities as a young scientist.

I owe considerable gratitude to Dr Tebogo Marutha for all his assistance when I was starting out in the laboratory. His quiet diligence and ever-present willingness to help, created an inclusive culture where selflessness and humility were the defining attributes of his mentorship. I also would like to thank Thokozile, Mistral and Jamie for being my functional genetics laboratory family.

I am grateful to the Microscopy and Microanalysis Unit for all their assistance. In particular I would like to thank Dr Deran Reddy for his guidance with regard to the image capturing and analysis component of my research.

Thank you to the NRF for funding this research. I am grateful for having received the opportunity to further my studies within the school of Molecular and Cell Biology.

I would like to acknowledge my family, for all their love and support. Thank you to my parents for their selfless investment in my education which has given me every opportunity.

Lastly thank you to Chris for all the love, motivation and coffee.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AI	-	Aromatase inhibitor
bp	-	Base pairs
BSA	-	Bovine serum albumen
BMI	-	Body mass index
BRCA1	-	Breast cancer gene 1
BRCA2	-	Breast cancer gene 2
CTCF	-	Corrected total cell fluorescence
CpG	-	5'-Cytosine-phosphate-Guanine-3'
DAPI	-	4', 6-Diamidino-2-phenylindole dihydrochloride
ddH_20	-	double distilled water
DMEM	-	Dulbecco's Minimum Essential Medium
DNA	-	Deoxyribonucleic acid
DNMT	-	DNA methyltransferase
ECM	-	Extracellular matrix
EDTA	-	Ethylenediaminetetraacetic acid
EMT	-	Epithelial-mesenchymal transition
ER	-	Oestrogen receptor
FISH	-	Fluorescence in situ hybridisation
FITC	-	Fluorescein isothiocyanate
H_2O_2	-	Hydrogen peroxide
HER2	-	Human epidermal growth factor receptor 2
HOBr	-	Hypobromous acid
HOCl	-	Hypochlorous acid

HRP	-	Horse radish peroxidase
MWM	-	Molecular weight marker
MAPK	-	Mitogen-activated protein kinase
NADPH	-	Nicotinamide adenine dinucleotide phosphate
Nox	-	NADPH-oxidases
Nrf2	-	Nuclear factor erythroid 2-related factor 2
nTPM	-	Normalised transcript per million
PBS	-	Phosphate buffered saline
PCR	-	Polymerase chain reaction
PI3K	-	Phosphoinositide 3-kinase
PR	-	Progesterone receptor
PXDN	-	Peroxidasin
Redox	-	Reduction-oxidation
RNA	-	Ribonucleic acid
ROS	-	Reactive oxygen species
TAE	-	Tris acetate EDTA
TGF-β	-	Transforming growth factor-β
TSG	-	Tumour suppressor gene
TUF	-	Thermodynamically ultra-fastened
UV	-	Ultraviolet

1 Introduction

In 2018, 18.1 million new cancer cases and 9.6 million cancer-related deaths were reported worldwide across all age groups (Bray et al., 2018). Risk stratification and prognostic markers play a significant role in the improvement of patient survival, allowing for a better match between patient and treatment plan (Selleck et al., 2017). As a highly heterogenic disease, there are various pathways involved in the development of cancer and the molecular fingerprint of the disease differs vastly between individuals and cancer types. This variation makes the development of efficacious treatment plans with minimal toxic effects a challenging task (Semenza, 2007). The past 10 years have shown a significant growth in the field of personalised medicine, specifically within the realm of cancer treatment and patient risk stratification. Promoter DNA methylation-based biomarkers have become more prominent candidates within the field of prognostic marker research and identification (Koch et al., 2018).

Peroxidasin (PXDN) is an extracellular matrix (ECM) peroxidase protein that forms hypohalous acids for various functions (Cheng et al., 2008). This protein has been linked to cancer through various pathways and processes including the phosphoinositide 3-kinase (PI3K/AKT) pathway (Zheng and Liang, 2018); its interaction with reactive oxygen species (ROS) (Dougan et al., 2019); and, perhaps most noteworthy, tissue invasion and metastasis through the process of epithelial-mesenchymal transition (EMT) (Tauber et al., 2010). PXDN displays dysregulated levels of expression in various cancer types including ovarian, prostate, bladder, oesophageal and breast cancer (Cai et al., 2018; Di et al., 2019; Dougan et al., 2019; Sigurdardottir et al., 2021; Zheng and Liang, 2018). Furthermore, dysregulated PXDN expression has also been associated with unfavourable prognoses (Zhou et al., 2022). The mechanisms that underlie changes in PXDN expression remain unknown. This study aimed to propose DNA methylation of the *PXDN* promoter as a mechanism responsible for changes in PXDN expression observed in breast cancer by evaluating whether this form of epigenetic regulation modifies the gene and the subsequent effects this has on protein expression. Changes to DNA methylation levels of the *PXDN* promoter within breast cancer cells could potentially act as a prognostic marker for the purpose of risk stratification in breast cancer patients.

1.1 Breast cancer

Breast cancer is the most prevalent form of the disease amongst women; composing 24.5% of female cases and responsible for 15.5% of female cancer-related deaths (Sung et al., 2021). Of all cancers diagnosed worldwide across age groups, female breast cancer incidence (11.7%) is highest, followed by lung (11.4%), prostate (7.3%) and nonmelanoma of skin cancer (6.2%) (Sung et al., 2021). Female breast cancer mortality accounts for 6.6% of all cancer-related deaths, fifth after lung (18%), colorectal (9.4%), liver (8.3%) and stomach (7.7%) cancer (Sung et al., 2021). Breast cancer mortality rates are highest in developing countries (Figure 1). Shulman et al. (2010) suggest that the reason for this trend is that screening technologies such as mammography and ultrasonography require expensive equipment which is not as readily available in low-income countries. Therefore, breast cancer is generally only identified at advanced stages when prognosis is poor. Non-genetic factors which impact the risk for breast cancer development include body mass index (BMI), breast density and age (Engmann et al., 2017). Genetic risk factors include mutations in genes which repair DNA damage such as breast cancer gene 1 and 2 (*BRCA1* and *BRCA2*), checkpoint kinase 2 *(CHEK2)* and ataxia-telangiectasia mutated (*ATM*) gene*;* as well as those which regulate cell growth such as tumour protein p53 (*TP53*) and phosphatase and tensin homolog (*PTEN*) (Chavarri-Guerra et al., 2017). Breast cancer is highly heterogeneous and can present a multitude of molecular fingerprints. Histopathologic classification of breast cancer has resulted in the creation of four distinct categories (Koboldt et al., 2012): luminal A, luminal B, basal-like and human epidermal growth factor receptor 2 (HER2) enriched (Table 1). Generally,

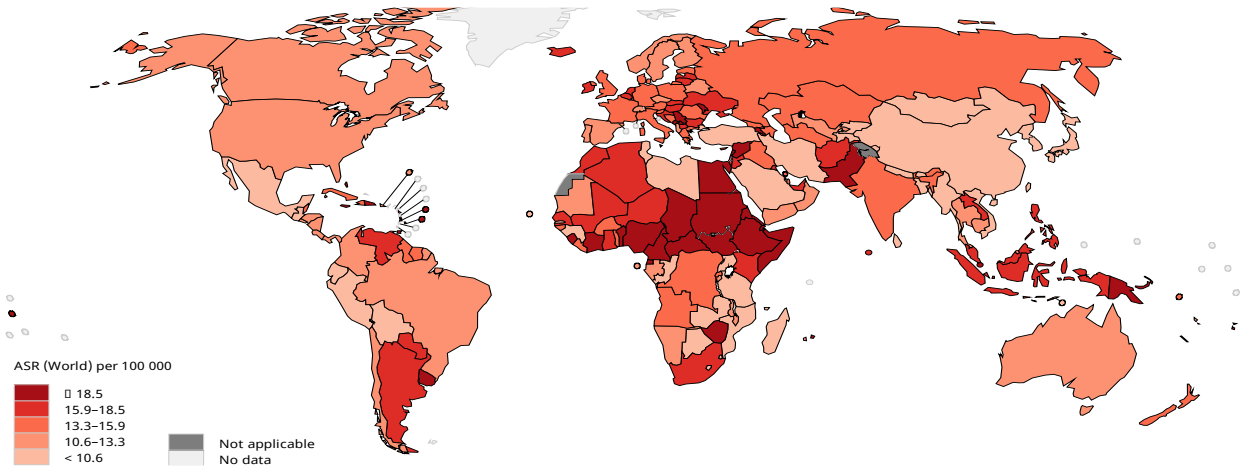


Fig. 1 Global breast cancer mortality rates. The highest breast cancer mortality rates are in less economically developed countries, particularly Central and Eastern Africa, Eastern Europe and Southern Asia. (Bray et al., 2018).

immunohistochemical assays and fluorescence in situ hybridisation (FISH) are used to determine breast cancer subtype (Harris, 2018).

Table 1: The four subtypes of breast cancer based on hormone and growth factor receptors.

Subtype	Cellular source of cancer	% of breast cancers that are this type	ERα status	PR status	HER2 status	Severity (grade, survival probability**, prognosis)	Treatment types
Luminal A	Invasive Ductal Carcinoma	30 – 40%	Positive	Positive	Negative	Low, 0.9, Favourable	Endocrine (AI) Advanced Cases: CDK4/6 or PI3K/AKT.mTOR inhibitors
Luminal B	Invasive Ductal Carcinoma	20 – 30%	Positive	Positive	Mostly negative*	Higher than Luminal A but still low, 0.5, less favourable	Endocrine (AI) and CDK4/6 or PI3K/AKT.mTOR inhibitors
Basal-like	Invasive Ductal Carcinoma	15 – 20%	Negative	Negative	Negative	High, 0.1, Poorest	PARP inhibitor chemotherapy
HER2+	Invasive Ductal Carcinoma	12 – 20%	Negative	Negative	Positive (overexpressed)	High, 0.3, Poor	HER2 inhibitor and HER2 dimerization inhibitor

* 16.4–20.8% of tumours display HER2 overexpression

** (Sørlie et al., 2001)

Luminal A and B subtypes show elevated expression of hormone-regulated pathways, particularly proliferation pathways regulated by the oestrogen receptor α (ERα), which shows elevated levels of expression in both subtypes (Prat et al., 2015). This means that these subtypes are generally responsive to endocrine therapy such as the suppression of oestrogen production with aromatase inhibitors (AIs), which prevent cell proliferation triggered by the various ER pathways. Luminal A tumours have the best prognosis and can be solely treated with endocrine therapy if they are smaller than 1 cm in diameter. Luminal B tumours can be distinguished from luminal A by lower expression of the progesterone receptor (PR), a higher percentage of *TP53* mutations (Koboldt et al., 2012), and higher expression of the proliferation marker Ki67 (Fragomeni et al., 2018). Higher Ki67 expression results in luminal B patients being more likely to develop resistance to AIs. To combat this effect AI treatment is often combined with a PI3K/AKT/mammalian target of rapamycin (mTOR) pathway inhibitor such as everolimus (Pritchard et al., 2013). About 20% of

luminal B tumours have been shown to present with *HER2* overexpression (Prat et al., 2015).

HER2 is a regulator of transcription and enhances protein synthesis and cell growth. HER2-enriched tumours show overexpression of HER2 and the highest number of mutations out of all the breast cancer subtypes (Prat et al., 2015). HER2-enriched tumours tend to have mutations in the tumour-suppressor genes *TP53* and *PI3KCA* (Koboldt et al., 2012). The HER2 receptor has both an intracellular tyrosine kinase domain and an extracellular domain. Drugs such as lapatinib bind to the intracellular tyrosine kinase domain, thereby inhibiting HER2 signalling pathways. Trastuzumab blocks HER2 functionality by binding to the extracellular domain of the receptor. Both intracellular and extracellular HER2 inhibitors have been shown to be effective in preventing disease progression (Cameron et al., 2008). Drug resistance, especially in more advanced cases of $HER2^+$ breast cancer, has been addressed by combining HER2 inhibitors with the HER2 dimerization inhibitor pertuzumab (Swain et al., 2015). Good outcomes are generally observed when trastuzumab and pertuzumab are used in conjunction for the treatment of $HER2^+$ breast cancer (Fragomeni et al., 2018).

Triple negative breast cancer (TNBC) is another term used to describe the basal-like molecular subtype which has high expression of proliferative genes associated with the skin basal layer. TNBC tumours generally exhibit low expression of *ERα*, *PR* and *HER2* and 80% of these tumours are *TP53*-mutated (Prat et al., 2015). This subtype is particularly heterogenous and has therefore been further categorised based on gene expression profiles: basal-like 1 and 2 (BL1 and BL2), mesenchymal (M), immunomodulatory (IM), luminal androgen receptor (LAR) and mesenchymal stem like (MSL) (Lehmann et al., 2011). TNBC has a higher likelihood of recurring and generally patients have shorter overall survival compared to other types of breast cancer (Blows et al., 2010). The variability that exists within the TNBC subtype makes it difficult to treat, unlike the other subtypes it is unresponsive to endocrine therapy and trastuzumab (Bettaieb et al., 2017). Although chemotherapeutic approaches can be used in the other subtypes of breast cancer, the margin of improvement between this approach and an endocrine and trastuzumab approach is minimal which allows chemotherapy to be avoided (Bettaieb et al., 2017). In TNBC, chemotherapy is the most effective form of treatment, however targeted therapies for TNBC do exist. One example is polyadenosine diphosphate-ribose polymerase (PARP) inhibitors such as Olaparib (Bettaieb et al., 2017). Olaparib or iniparib are generally used in conjunction with chemotherapeutic drugs which induce DNA cross-linking such as carboplatin (O'Shaughnessy et al., 2014). PARP is responsible

for repairing DNA damage and therefore inhibiting its action allows for DNA damage caused by chemotherapeutic drugs in cancer cells to not be reversed. The chemotherapeutic approach to TNBC involves the combination of docetaxel which inhibits microtubule dimerization and therefore prevents mitosis from occurring; doxorubicin which is an intercalating agent; and cyclophosphamide which acts as a DNA cross-linker (Isakoff, 2010).

The heterogeneous nature of breast cancer and particularly TNBC means that there is extensive variation above and beyond the four histopathological divisions that have been discussed. This variation is observable through genomic and epigenomic signatures and is worth noting when it comes to treatment decisions and selecting the correct patients for clinical trials. For example Lehmann et al. (2016) showed that the further subtyping of TNBC assisted in the prediction of neoadjuvant chemotherapy efficacy in patients. As discussed previously, DNA methylation is dysregulated in cancer. Drugs such as DNA methyltransferase inhibitors (DNMTi) which prevent the hypermethylation of tumour suppressor genes (TSGs) have been effective in improving patient survival in TNBC. Stirzaker et al. (2015) found methylation clusters with prognostic value in TNBC. One of these methylation clusters was gene body methylation within the Wilms tumour 1 (*WT1*) gene, which correlated with elevated gene expression and poor survival whereas hypermethylation of the promoter of the same gene was associated with low levels of expression and improved patient survival. Therefore, DNA methylation-based biomarkers are worth investigating within the realm of breast cancer research as they hold the potential to improve risk stratification and the analysis of the efficacy of treatment plans.

1.2 The ECM: A Master Regulator of Cancer Progression

The ECM is the non-cellular component of tissue and performs major roles in cellular support as well as various regulatory biochemical and mechanical processes including growth, proliferation, migration and differentiation (Frantz et al., 2010; Walker et al., 2018). These processes are regulated through biochemical and even mechanical signals: mechanotransduction is the initiation of signalling cascades through tensional, compression and shearing forces (Lampi and Reinhart-King, 2018). ECMs are composed of various components including collagens, elastin, proteoglycans (PGs), fibronectin, laminins, glycoproteins and hyaluronan, which are arranged in various configurations depending on tissue function (Theocharis et al., 2019). The ECM, rather than acting as a fixed structure, is a fluid environment which constantly fluctuates in response to

a multitude of signals, mechanical or chemical (Eble and Niland, 2019; Walker et al., 2018). There are two broad categories of ECM: interstitial matrices are found between and around cells whereas pericellular matrices are in close contact with cells (Theocharis et al., 2016).

As mentioned above, the ECM is composed of various fibrous proteins and PGs, which define its structural and chemical properties and differ in ratio and organisation depending on the tissue type. Of these components, collagen is the most abundant superfamily, making up roughly 30% of the protein within mammalian tissue types and performing a host of structural, organisational and mechanical functions (Eble and Niland, 2019; Theocharis et al., 2016). There are 28 different types of collagen (Gordon and Hahn, 2010), the defining feature of which is a distinct triple-helical domain formed by the combination of three α chains. The combination of these α chains results in either the formation of homo- or heterotrimers that make up the helical structure (Ricard-Blum, 2011). The sequence which defines the right-handed helical structure formed by these α chains is the Gly-X-Y sequence where X and Y are predominantly taken up by proline and hydroxyproline amino acids respectively, preceded by a glycine amino acid (Ricard-Blum, 2011; Theocharis et al., 2016). There are seven categories making up the collagen protein superfamily; two of these are the fibrillar and network-forming categories. Fibrillar collagens, such as collagen I, are found in connective tissues such as skin, bone, cartilage and cornea (Muiznieks and Keeley, 2013). Collagen I is the most abundant fibrillar collagen and is involved in processes such as wound repair and organ development (Walker et al., 2018). Network forming collagens form the basal lamina of basement membranes with collagen IV as the most common example (Ricard-Blum, 2011).

Basement membranes are a type of pericellular ECM that form a thin barrier between parenchyma and stroma (Eble and Niland, 2019; Theocharis et al., 2016). The main component of basement membranes is a stabilising collagen IV network (Theocharis et al., 2016). Collagen IV molecules form hexamers by binding to one another's NC1 domain at the C-terminal, this network is further strengthened by other proteins such as perlecan and laminin (Ricard-Blum, 2011). The basement membrane is responsible for anchoring cells and enforcing cellular adhesion, impairment of this structure is associated with metastasis in cancer (Chang et al., 2017; Eble and Niland, 2019). Metastasis is responsible for most cases of cancer mortality. The movement of metastatic cancer cells from the primary tumour site through the barriers of basement membranes and connective tissue to other parts of the body, requires motility of malignant cells (Theocharis et al., 2016). This movement also requires the alteration in the structure of the ECM itself in a tumour-permissive

way (Kalluri, 2016). Cancer cells must attain a motile phenotype in order to engage in the process of metastasis, this entails the growth of actin extracellular obtrusions which are termed invadosomes (Chang et al., 2017). Invadosomes infiltrate the ECM and degrade it through the recruitment of lytic enzymes such as matrix metalloproteinases (MMPs), this allows infiltration of the basement membrane and results in metastasis (Chang et al., 2017). MMP-2 and -9 are specifically active during tumour invasion through the degradation of collagen IV and therefore of the basement membrane (Deryugina and Quigley, 2006).

Paradoxically, although collagen acts as a barrier to cancer invasion (as seen through the necessity of MMPs to break down the basement membrane during metastasis), increased collagen deposition and cross-linking is actually associated with an increase in metastasis (Lampi and Reinhart-King, 2018; Levental et al., 2009; Walker et al., 2018). Levental et al. (2009) induced collagen crosslinking which caused ECM stiffening and promoted malignancy in breast cancer (Figure 2). As the most abundant ECM protein, collagen is responsible for changes in ECM stiffness: Lo et al. (2000) showed that an increase in ECM stiffness stimulates cell growth, survival, and migration. This increase in the deposition of collagen (as well as other fibrous ECM proteins) interferes with cellular adhesion, polarity and increases growth factor signalling (Walker et al., 2018). Collagen cross-linking is performed by enzymes such as lysyl-6-oxidase (LOX) which are produced by cancer-associated fibroblasts (CAFs) and increase collagen deposition thereby resulting in an increase in ECM density and interstitial pressure (Karagiannis et al., 2012). LOX expression is induced by TGF-β signalling or by hypoxia, both of which are associated with the progression of cancer (Postovit et al., 2008).

It has been established that the ECM and its dysregulation are key components in the processes of tumorigenesis, invasion and metastasis. The stiffening of the ECM is caused by a multitude of aberrant cell signalling and structural processes. One of these is the increase in collagen deposition and crosslinking which 1) triggers signal transduction cascades resulting in cell proliferation and survival; and 2) allows for the breaching of the basement membrane and assists in cell motility. The combination of changes in cell morphology through processes such as EMT, and the characteristics of the basement membrane are both firmly linked to the composition and nature of the ECM. This review will now discuss how PXDN, a collagen cross-linker, could be involved in promoting tumour metastasis just as LOX, another collagen cross-linker, has been shown to promote metastasis by stimulating the above processes.

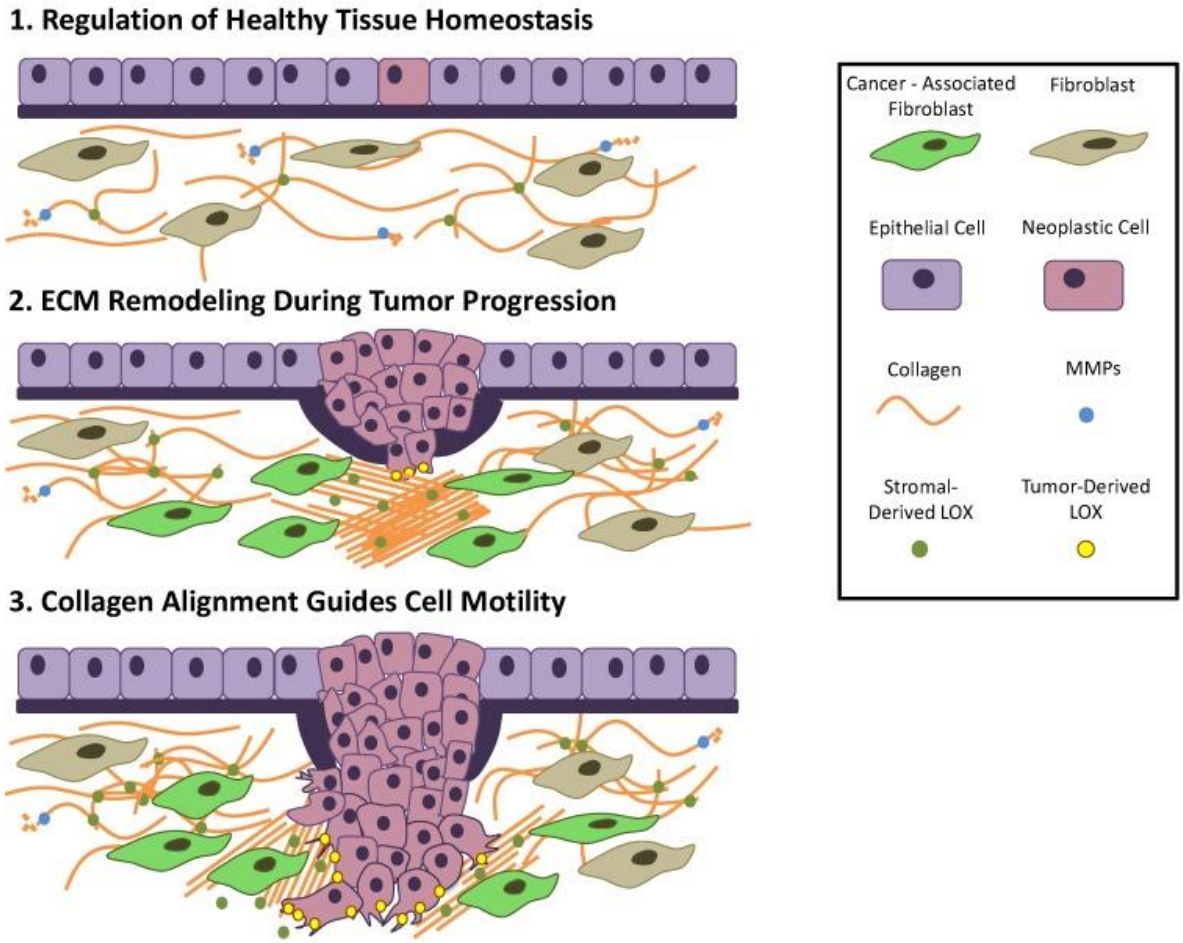


Fig. 2 The remodelling of the ECM takes place during cancer progression. 1) Cancer cells (in this case epithelial neoplastic cells) replicate and put the basement membrane under strain. 2) The basement membrane bulges under this strain. Enzymes such as LOX cause the linearisation of collagen (i.e. collagen cross-linking) which is produced by cancer-associated fibroblasts. 3) The integrity of the basement membrane is compromised (by MMPs produced by invading cells) and cancer cells migrate, aided in their migration along the cross-linked collagen. (Walker, Mojares and Del Río Hernández, 2018). MMPs – matrix metalloproteinases; LOX – lysyl-6-oxidase

1.3 PXDN: a Regulator of Cellular Adhesion

PXDN is a haem-containing peroxidase which has been shown to form hypohalous acids in the presence of hydrogen peroxide (H_2O_2) produced by nicotinamide adenine dinucleotide phosphate (NADPH) oxidases (NOX enzymes) in the electron transport chain (Cheng et al., 2008). PXDN is distinct from other peroxidase proteins in that it contains ECM domains; these include a leucine rich repeat domain, multiple immunoglobin I-set domains and a von Willebrand factor type C domain (El-Gebali et al., 2019). Leucine rich repeat domains have both polar and non-polar

components and have been shown to facilitate interactions between proteins and lipids as well as protein-protein interactions (Karaulanov et al., 2006). Immunoglobin domains are associated with the major histocompatibility complex recognised by T-cells for the adaptive immune system, indicating protein-ligand and protein-protein interactions (Chan et al., 2000). Limited research has been performed on the von Willebrand factor type C protein domains, however Zhang et al. (2007) found them to be associated with cellular processes such as migration, adhesion, and signalling.

Bhave et al. (2012) showed that PXDN catalyses sulfilimine bond formation between lysine and methionine amino acids of the C-terminals of two collagen IV protomers, resulting in the formation of the collagen IV hexamer (Figure 3). Hypohalous acids such as hypobromous acid (HOBr) and hypochlorous acid (HOCl) are produced by PXDN as reaction intermediates during H_2O_2-mediated oxidation. These intermediates react with collagen IV molecules ultimately resulting in sulfilimine bond formation which stabilises basement membranes (Bhave et al., 2012). The NC1 collagen IV hexamer is composed of two collagen IV monomers and two collagen IV dimers. The protomers of these dimers are linked to one another by a sulfilimine bond (Figure 3). Mutations in the *Drosophila PXDN* orthologue and the *C. elegans* orthologue *PXN-2* show similar effects on the basement membrane to mutations in collagen IV (Bhave et al., 2012; Gotenstein et al., 2010). This indicates that the crosslinks formed in collagen IV by PXDN are essential to stabilising and ensuring the integrity of the basement membrane (Bhave et al., 2012). As discussed earlier, although basement membrane integrity is reinforced through the cross-linking of collagen IV, during tumorigenesis increased deposition and crosslinking is counterintuitively associated with tumour progression (Walker et al., 2018).

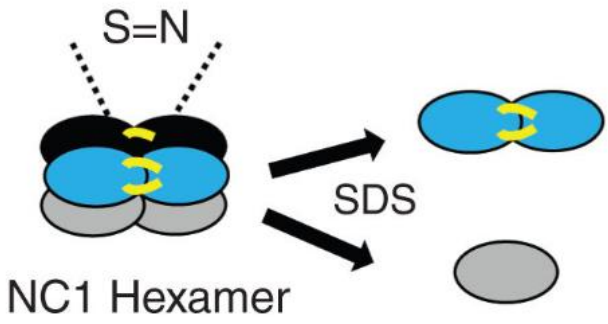


Fig. 3 The NC1 collagen IV hexamer composed of collagen monomers and dimers. The collagen IV NC1 hexamer found in basement membranes dissociates into sulfilimine cross-linked dimers and uncross-linked monomers when exposed to SDS. (Bhave et al., 2012).

The role of PXDN in basement membrane stabilisation raises the question of whether its dysregulation could contribute to the increased mobility observed in cancer. EMT is a characteristic of cancer cells which allows for increased cell motility and therefore invasion and metastasis (Hay, 1995). This is because mesenchymal cells can migrate through the ECM, which epithelial cells are unable to do (Hay, 1995). Higher ROS levels have been shown to be present in cancer cells (Cruz-Bermúdez et al., 2019) and this increase in ROS has been linked to EMT. For example, high ROS levels activate the TGF-β signalling pathway (González-Ramos et al., 2012) which stimulates the activation of the mitogen-activated protein kinase (MAPK) pathway in epithelial cells, resulting in EMT (Rhyu et al., 2005). PXDN's formation of the sulfilimine bond is facilitated by the ROS H_2O_2 and therefore the increase in ROS levels associated with EMT could also lead to increased PXDN crosslinking activity. Further, *PXDN* has been shown to be regulated by the master redox regulator nuclear factor erythroid 2-related factor 2 (Nrf2) (Hanmer and Mavri-Damelin, 2018).

Changes in regular expression levels of PXDN have also been directly associated with EMT (Tauber et al., 2010). Cano et al. (2000) showed that the transcription factor Snai1 initiates EMT by 1) suppressing the expression of E-cadherin which is essential for the maintenance of cellular adhesion and 2) increases the expression of mesenchymal markers such as vimentin, fibronectin, and N-cadherin. Sitole and Mavri-Damelin (2018) showed that the Snai1 transcription factor regulates *PXDN* which suggests its involvement in EMT. PXDN is involved in the EMT process in various contexts during development including formation of the neural tube and muscle-epidermal attachment (Gotenstein et al., 2010; Sitole and Mavri-Damelin, 2018; Tindall et al., 2005). Therefore, PXDN is involved in EMT and increased expression in cancer can be correlated with increased invasive potential and metastasis.

1.4 PXDN, Oxidative Stress and Signalling Pathways Associated with Cancer

High ROS levels cause damage to nucleic acids and cellular structures which triggers apoptosis (Redza-Dutordoir and Averill-Bates, 2016). Dougan et al. (2019) showed that PXDN regulated oxidative stress and ROS levels and linked these phenomena to a decrease in apoptosis in prostate cancer, proposing PXDN as a biomarker for the disease. They found increased phospho-p53 and Bcl-2 associated x-protein (Bax) expression as well as increased fluorescence in the apoptosis-detecting TUNEL assay following PXDN knockdown. Therefore, PXDN regulates apoptosis

through the metabolism of ROS, preventing the activation of apoptosis that an excess of ROS would usually induce.

Other than the regulation of oxidative stress and the stimulation of EMT, PXDN has been linked to cancer through certain signalling pathways such as the PI3K/AKT pathway. The PI3K/AKT pathway results in the activation of various transcription factors including the mammalian target of rapamycin (mTOR) which promotes cell growth (LoRusso, 2016). mTOR and forkhead box transcription factor class O (FOXO) are both transcription factors which are activated by PI3K/AKT signalling and are responsible for transcribing genes associated with cell proliferation and survival respectively (LoRusso, 2016). Zheng and Liang (2018) found that *PXDN* knockdown resulted in a significant decrease in the expression of proteins related to the PI3K/AKT pathway in HEY ovarian human carcinoma cells. Examples of these proteins are phosphorylated AKT and PI3K which showed decreased levels of expression when siRNA was used to knockout *PXDN*. Therefore, through the PI3K/AKT pathway, *PXDN* knockdown resulted in a decrease in cell proliferation, migration and invasion in cervical cancer cells (Zheng and Liang, 2018). Levental et al. (2009) investigated the impact of increased collagen cross-linking on the process of breast tumorigenesis. They found that inducing collagen cross-linking increased the number of focal adhesions, increased PI3K signalling, caused ECM stiffening and initiated the process of invasion of an epithelium, and therefore confirms the link between the PI3K pathway and collagen cross-linking.

1.5 Dysregulated *PXDN* Expression in Cancer

The *PXDN* gene is found on the reverse strand of chromosome 2 at position 2p25.3 and is approximately 100 kbp in length (Hunt et al., 2018) with a promoter of about 10100 bp (Kent et al., 2002). As previously discussed, Snai1 and Nrf2 are transcription factors of *PXDN* and have been shown to regulate EMT and redox processes respectively (Hanmer and Mavri-Damelin, 2018; Sitole and Mavri-Damelin, 2018).

According to data compiled by the Human Protein Altas, *PXDN* shows particularly high expression in female tissues, of these tissues breast tissue shows the highest expression of *PXDN* (Figure 4). Sigurdardottir et al. (2021) conducted immunohistochemistry and reverse transcriptase polymerase chain reaction (RT-PCR) experiments and found that *PXDN* is expressed in epithelial cells, fibroblasts, and endothelial cells within the mammary gland. Therefore, *PXDN* shows

particularly high expression in breast tissue and dysregulation could contribute to cancer progression. *PXDN* shows altered expression levels in various types of cancer as shown in Figure 4, which was compiled from results of The Cancer Genome Atlas (TCGA) project.

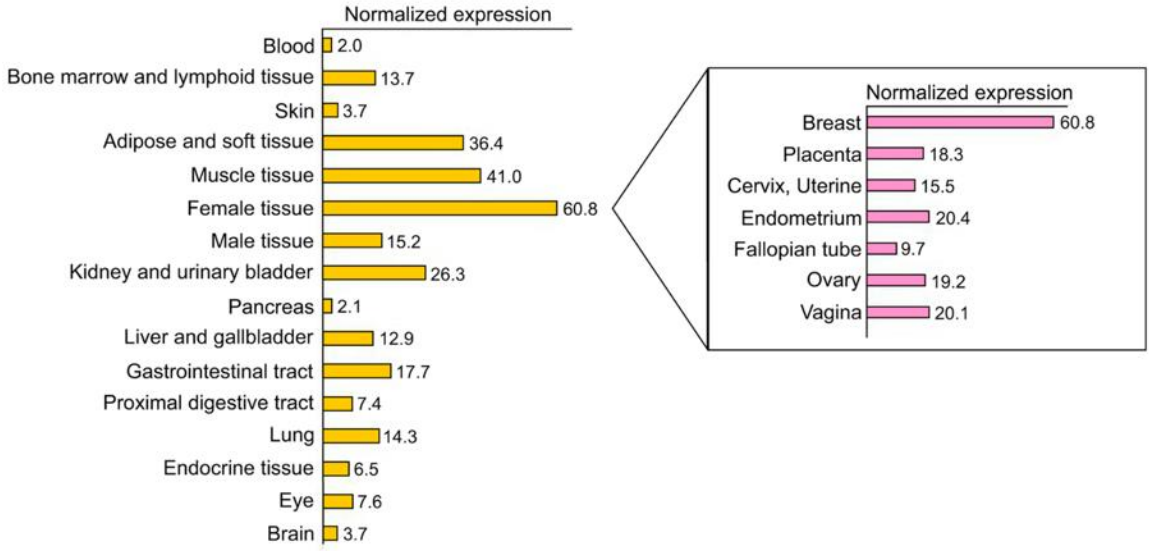


Fig. 4 Normalised *PXDN* expression data from the Human Protein Atlas. *PXDN* shows highest expression in female tissues and of these breast tissue expresses the highest level of *PXDN* (Sigurdardottir et al., 2021).

Figure 5 indicates fragments per kilobase of exon per million reads (FPKM) from RNA-seq data from the cancer genome atlas (TCGA) project (Uhlen et al., 2017). The highest level of *PXDN* expression in Figure 5 is within testis, breast and ovarian cancer which supports the hypothesis that it could be involved in disease progression. Indeed various studies have found elevated levels of *PXDN* expression in cancers including ovarian, prostate, bladder, breast and oesophageal cancer (Cai et al., 2018; Di et al., 2019; Dougan et al., 2019; Sigurdardottir et al., 2021; Zheng and Liang, 2018). A pan-cancer analysis of PXDN was conducted by Zhou et al. (2022); gene expression and clinical data from 33 tumours were downloaded from TCGA and compared to their normal constituents. *PXDN* expression was closely correlated to overall survival and acted as a risk factor in several cancer types including bladder urothelial carcinoma (BLCA), squamous cell carcinoma and stomach adenocarcinoma (STAD).

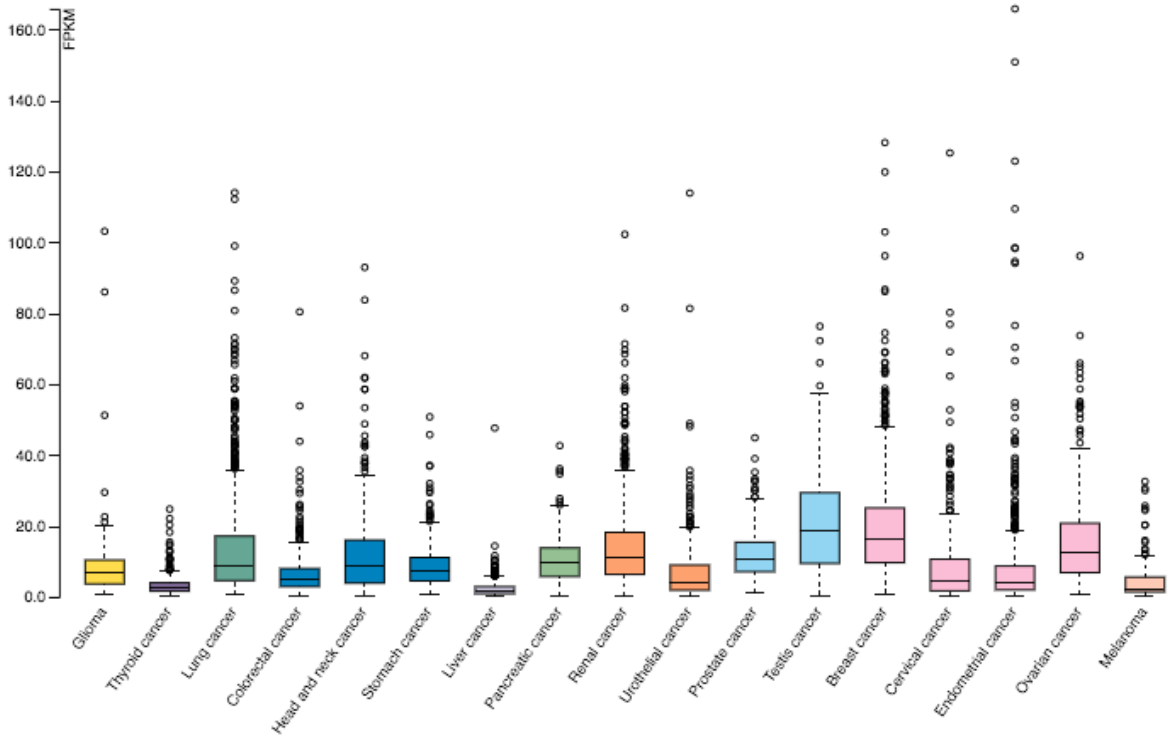


Fig. 5 *PXDN* expression in 17 cancer types determined by RNA-Sep data from TCGA project. Highest expression levels are observed in testis cancer (median: 18.7 FPKM), breast cancer (median: 16.4 FPKM) and ovarian cancer (median: 12.6 FPKM). Expression levels are determined by analysing RNA-Seq data from TCGA project, this graph was compiled by the protein atlas site (proteinatlas.org).

Figure 6 shows PXDN as an unfavourable prognostic marker in endometrial, cervical and stomach cancer. Generally a gene is classified as an unfavourable prognostic gene if it has an FPKM value higher than one and patient survival is low. These Kaplan-Meier plots show high *PXDN* expression correlated with low levels of patient survival (Uhlen et al., 2017). Zhou et al. (2022) also found a significant correlation between later cancer stages and *PXDN* expression in various tumours including BLCA, colon adenocarcinoma (COAD), testicular germ cell tumours (TGCT), uterine corpus endometrial carcinoma (UCEC) and uveal melanoma (UVM). Supporting these findings, Dougan et al. (2019) and Zheng and Liang et al. (2018) found that increased *PXDN* expression correlated with a more advanced stage of cancer in prostate and ovarian cancer respectively.

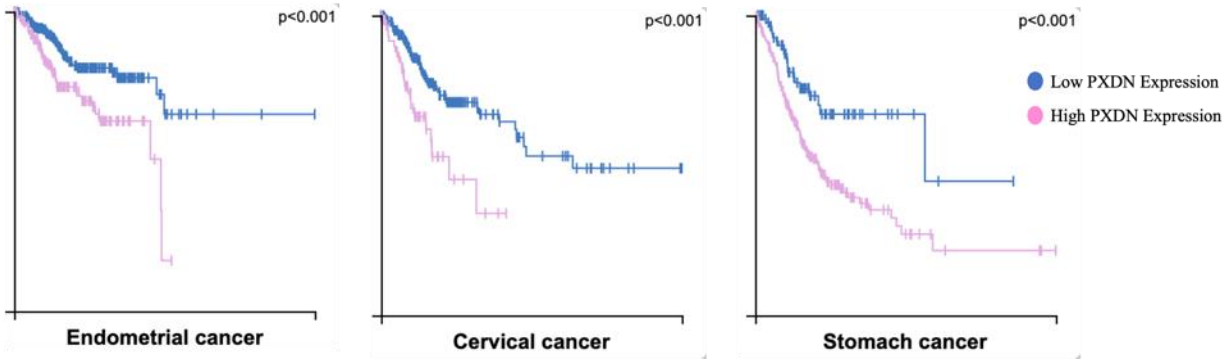


Fig. 6 High PXDN expression is associated with unfavourable prognosis in endometrial, cervical and stomach cancer. These Kaplan-Meier survival plots were determined using RNA-seq PXDN expression data and patient survival rates. An unfavourable prognostic marker is defined by an FPKM value greater than 1 correlated with decreased patient survival. This data was compiled by the Human Protein Atlas (proteinatlas.org).

1.6 DNA Methylation and Cancer

Epigenetics describes the inheritance of phenotypes caused by chemical modifications to DNA and histones which impact gene expression without altering DNA sequence. One of the ways in which epigenetic regulation takes place is through the modification of histones, for example, methylation, acetylation and phosphorylation (Klug et al., 2015). Histone methylation occurs on the lysine and arginine amino acid residues in histone proteins. The function and type of histone methylation depends on the location and context, for example, H3K9me3 modifications are seen in active genes but also have been associated with repressive effects within constitutive heterochromatin (Michalak et al., 2019). RNA molecules such as microRNAs (miRNAs) and long noncoding RNAs (lncRNAs) are involved in gene expression and silencing, and these RNA molecules play key roles in gene regulation within the epigenome (Klug et al., 2015). One of the most well-studied forms of epigenetic modification is DNA methylation.

Cytosine bases which precede guanine nucleotides (CpG sites) are targeted for methylation by DNA methyltransferases (DNMTs), which incorporate a methyl group to the fifth carbon of the cytosine base forming 5-methylcytosine (5mC) (Moore et al., 2013). Ten-eleven translocation (TET) methylcytosine dioxygenases execute demethylation by oxidising 5mC to 5-hydroxymethyl cytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Ito et al., 2011). DNA methylation plays a crucial role in gene regulation and is involved in various processes including X-chromosome inactivation (Mohandas et al., 1981), imprinting (Reik et al., 1987) and the

maintenance of tissue-specific expression patterns. Within gene promoters, CpG methylation inhibits transcription by sterically hindering transcription factor binding and recruiting methyl-CpG-binding proteins which bind to methyl groups. These methyl-CpG-binding proteins interact with histone deacetylase to bring about a less open chromatin configuration and therefore further suppress expression (Zhu et al., 2016). However, where DNA methylation occurs outside gene promoters, it has become clear that the mechanisms behind this epigenetic alteration are more complex than was once believed and its effect can vary depending on location and context.

DNA methylation originated in bacteria and is present in many eukaryotes (Zemach et al., 2010). Despite being an integral component of the genomes of many organisms, the 5mC modification is not found in some eukaryotes such as the model organism *Drosophila melanogaster* (Greenberg and Bourc'his, 2019). 5mC is spontaneously deaminated to thymine which causes G • T mismatches, these mutations are frequently seen in tumours and this could explain why the 5mC epigenetic regulatory mechanism has been lost from certain evolutionary lineages (Holliday and Grigg, 1993). Indeed, mammals which do exhibit DNA methylation have a significantly lower CpG content than expected. This may be the reason why most CpG sites within the human genome are methylated, however between 20% and 30% are not methylated and form clusters generally around 1kb in size referred to as CpG islands (CGIs) (Gardiner-Garden and Frommer, 1987). The majority of CGIs can be found within gene regulatory elements (Li and Zhang, 2014) and are located in about two thirds of the promoters within the human genome, often unmethylated (Bird et al., 1985; Gardiner-Garden and Frommer, 1987). However, where 5mCs occur within gene bodies, they are mostly methylated and assist with transcriptional elongation and even influence splicing (Baylin and Jones, 2016; Jones, 2012; Kulis et al., 2012; Zafon et al., 2019). Therefore DNA methylation is not always associated with transcriptional repression; de Almeida et al. (2019) investigated a cohort of CpG sites and found that 206 CpG sites showed differences in methylation which correlated to changes in expression within 169 genes between normal and breast tumour tissue. Furthermore, there was a negative correlation between CpG methylation levels and gene expression for CpG sites within gene promoters whilst they identified a positive correlation between hypomethylated CpG sites and gene expression within gene bodies.

There are certain patterns of cytosine methylation which are strongly associated with the process of tumorigenesis and the methylation of these cells can be distinguished from that of non-tumorigenic cells in three different ways (Baylin and Jones, 2016). The first of these is what has

been referred to as global hypomethylation; the majority of this hypomethylation associated with cancer occurs within transposable elements (TEs) which make up ~50% of the human genome. TEs generally have a high CpG content and are most commonly methylated; demethylation of TEs is associated with genomic instability and cancer (Bae et al., 2012; Kim et al., 2012; Wolff et al., 2010). The hypomethylation of the promoters of oncogenes constitutes a much smaller component of the global hypomethylation observed within cancer cells. For example the hypomethylation of *SMYD3* (a gene encoding histone lysine methyltransferase) and the *Erb-A1* oncogenes has been linked to colorectal cancer and lymphatic leukaemia respectively (Li et al., 2018; Lipsanen et al., 1988) and the hypomethylation of the thiosulfate sulfurtransferase-like domain containing 1 (*TSTD1*) gene was associated with poor overall survival and poor treatment response in breast cancer (Ansar et al., 2022). Another observation regarding DNA hypomethylation which is relevant to cancer is that CpGs within gene bodies seem to be hypomethylated in tumorigenic cells (Jones, 2012; Kulis et al., 2012). For example Kulis et al. (2012) found that hypomethylation of gene bodies and enhancers was a characteristic in the differentiation of B-cells in chronic lymphocytic leukaemia which means that DNA methylation plays a role beyond that of regulating transcription factor binding sites within gene promoters.

Secondly, the hypermethylation of the CGIs within the promoters of TSGs has long been known to be associated with cancer progression. As discussed earlier, roughly 60% of gene promoters contain CGIs which are largely unmethylated. The cytosine methylation of the promoters of genes with tumour suppressor functions allow for carcinogenesis to take place uninhibited (Herman et al., 1995; Melki et al., 1999). The hypermethylation of the TSG *BRCA1* is known to be involved in the progression of breast cancer (Esteller et al., 2000). Paydar et al. (2019) found that *BRCA1* promoter hypermethylation is strongly associated with poor patient prognosis. Other studies have confirmed this (Gupta et al., 2014; Iwamoto et al., 2011) and the drug resveratrol was successfully used to prevent the hypermethylation of the *BRCA1* promoter in MCF-7 cells (luminal A subtype breast cancer cells) (Papoutsis et al., 2012). Other TSGs which display promoter hypermethylation and a decrease in expression in breast cancer are *ATM* (Brennan et al., 2012; Flanagan et al., 2009), *TMS1* (target for methylation-mediated silencing) (Parsons and Vertino, 2006), and the gene coding for E-cadherin (*CDH1*) (Shargh et al., 2014). Moelans et al. (2011) found that the promoters of paired box 6 (*PAX6*), *BRCA2*, paired box 5 (*PAX5*), *WT1*, Cadherin-13 (*CDH13*) and MutS homolog 6 (*MSH6*) were hypermethylated in both ductal carcinoma *in situ* (DCIS) and invasive ductal cancer (IDC) samples which correlate to early breast cancer. Therefore, the

hypermethylation of CGIs is highly relevant to breast cancer research, particularly where this phenomenon occurs within genes that carry out tumour suppressor functions.

Thirdly, 5mCs are highly mutagenic and, where they occur within gene bodies, CpG sites are highly methylated and particularly concentrated within exons (Baylin and Jones, 2016; Jones, 2012). Studies have shown that within gene bodies, DNA methylation actually has a positive correlation with transcription and is highly conserved (Varley et al., 2013; Yang et al., 2014). The deamination of 5mC to thymine is prevalent in causing loss-of-function mutations within TSGs. For example, Rideout et al. (1990) showed that this phenomenon occurs within the tumour suppressor gene *TP53* and Greenblatt et al. (1994) confirmed that more than 50% of the mutations within the same gene in colorectal cancer occurred at 5mC sites. In addition to 5mC deamination, 5mC has been shown to form carcinogenic adducts when exposed to benzo(a)pyrene (BaP) (Bukowska and Sicińska, 2021; Greenblatt et al., 1994). BaP is one of the carcinogenic compounds found within cigarette smoke and these adducts formed between the metabolites of BaP and 5mC have been linked to increased lung cancer mutations in smokers (Baylin and Jones, 2016; Bukowska and Sicińska, 2021; Greenblatt et al., 1994).

The integral role that these changes in DNA methylation patterns play in the process of tumorigenesis have naturally lead to the investigation into drugs which have the ability to alter DNA methylation levels. DNMT inhibitors (DMTNi) such as azacytidine and decitabine cause DNA hypomethylation and have been approved for use during the treatment of acute myeloid leukaemia (AML) and myelodysplastic syndrome (MDS) (Wouters and Delwel, 2016). These same drugs have been investigated for the treatment of TNBC (Singh et al., 2021; Wong, 2021; Yu et al., 2018). DNMTi drugs have been most efficacious when used in conjunction with other forms of treatment, for example Mirza et al. (2010) found that the proliferation of TNBC cells MDA-MB-231 was inhibited using a DNMTi and paclitaxel, Adriamycin or 5-fluorouracil (common anticancer drugs used for the treatment of breast cancer). Another form of epigenetic treatment which is under investigation is a group of drugs which are known as histone deacetylase inhibitors (HDACi). Histone deacetylation is a histone modification which is often found in hypermethylated gene promoters whilst histone acetylation promotes a more open chromatin configuration and therefore increased gene expression. Therefore this group of drugs is worth noting when investigating the effect of TSG repression in cancer development. However, Connolly et al. (2021) studied the use of HDACi entinostat in endocrine-resistant breast cancer and found

that it did not improve patient survival in hormone receptor positive, HER2-negative advanced breast cancer.

It is clear that changes in DNA methylation patterns are integral to the process of tumorigenesis. However, depending on the position of the CpG site, alterations in methylation level can have varying impacts on expression. Apart from opening up a realm of potential targets for cancer treatment, the study of DNA methylation has also raised the concept of methylation-based biomarkers. N-Myc downstream-regulated gene 4 (*NDRG4*), bone morphogenic protein (*BMP*) and Septin-9 (*SEPT9*) are all methylation-based biomarkers implemented clinically for the early detection of colorectal cancer (Koch et al., 2018). Short stature homeobox 2 (*SHOX2*) and Glutathione S-Transferase Pi 1 (*GSTP1*) are methylation-based biomarkers currently used for the early detection of lung cancer and the diagnosis and prognosis prediction of prostate cancer respectively (Koch et al., 2018). Similarly, methylation patterns have been used to screen patients for breast cancer and calculate risk of breast cancer development (Tang et al., 2016).

Methylation-based markers for breast cancer have been investigated for the purpose of potentially improving patient prognosis by early disease detection, particularly in detecting recurrence of breast cancer for patients in remission (Tang et al., 2016). Case-control studies to investigate gene-specific hyper- or hypomethylation of promoter regions in breast cancer cases have been performed for *BRCA1* and *ATM*. Therefore, particularly in the screening process, analysis of promoter methylation levels could help with risk stratification (Tang et al., 2016). de Almeida et al. (2019) performed a bioinformatic analysis on TCGA data and investigated 780 breast tumour samples and 83 normal tissue samples. They investigated CpG sites which were differentially methylated in normal and breast tumour tissue and found sites within zinc finger protein 154 (*ZNF154*) and homeobox D9 (*HOXD9*) which were hypermethylated within breast tumour tissue and correlated to poor prognosis. Stirzaker et al. (2015) analysed the prognostic value of CpG clusters in TNBC and identified 36 TNBC-specific hypermethylated regions occurring within gene bodies and promoters. They found that many of these hypermethylated sites specific to TNBC occurred within genes which code for transcription factors and zinc fingers, including *ZNF154* and *ZNF671*. Both Stirzaker et al. (2015) and de Almeida et al. (2019) found that the majority of the genes that were hypermethylated within breast cancer tumours were transcription factors and homeobox genes and de Almeida et al. showed that hypomethylated regions coded for transmembrane proteins and immunoglobulins.

Apart from the three above-mentioned examples of histone and DNA methylation-regulated ECM genes which have been connected to the progression of various forms of cancer, Koch et al. (2018) identified 1800 unique DNA methylation-based cancer biomarkers. It is clear that DNA methylation in particular has been considered a promising field of research within the disciplines of cancer prognosis and diagnosis biomarkers. The question this study aims to address is whether the collagen cross-linker PXDN holds the potential to act as a clinical biomarker for breast cancer.

1.7 *PXDN* Promoter Methylation: A Potential Prognostic Marker?

As discussed previously, *PXDN* expression is higher in various types of cancer and is an unfavourable prognostic marker. The collagen cross-linking role of PXDN is associated with matrix stiffening and EMT, therefore supporting the observed association of increased expression and poor prognosis. The mechanism behind increased *PXDN* expression levels remains unknown, however it is possible that increased expression is caused by hypomethylation of the CpG island within the *PXDN* promoter or hypermethylation of CpG sites within the *PXDN* gene body. If this is the case, it is possible that *PXDN* methylation levels could act as a biomarker. Whether or not demethylation of the *PXDN* promoter can account for increased expression levels of *PXDN* observed in breast cancer remains to be seen. This study aims to answer these questions and proposes *PXDN* promoter hypermethylation as a potential prognostic biomarker for breast cancer if DNA methylation is indeed the one of the causes of changes to *PXDN* expression levels observed in breast cancer in particular.

As discussed in the previous section, breast cancer patient prognosis improves significantly in cases of early detection, therefore the continual discovery of new biomarkers which improve the screening process and risk determination, is crucial (Tang et al., 2016). If PXDN does prove to be a successful prognostic marker for breast cancer patients, it stands to act as a significant tool within the realm of breast cancer treatment programs. Koch et al. (2018) make a noteworthy point in their paper on DNA methylation biomarkers: the location of CpG binding sites within the promoter of genes is of great importance because some sites are more significant in the silencing of expression than others. This is an important point to keep in mind whilst considering this research question.

1.8 Laboratory Methods for Analysing DNA Methylation

There are various methods which are used for the investigation of CpG methylation patterns in DNA. Factors such as cost, accuracy, sensitivity and time are important points of consideration when choosing a method for DNA methylation analysis. Another key consideration is the nature of the analysis to be performed; genome-wide verses site-specific DNA methylation patterns require distinct approaches.

The gold standard of DNA methylation analysis is the method developed by Frommer et al. (1992) who discovered that sodium bisulfite treatment revealed a way to differentiate between C and 5mC at CpG sites. Taq DNA polymerase is unable to make this distinction and cannot incorporate methyl groups into the extending strand. Therefore the amplicons of standard PCR reactions lack the epigenetic markers of the original sequence. Sodium bisulfite treatment of DNA results in the deamination of unmethylated cytosine nucleotides. This allows for a distinction to be made between methylated and unmethylated cytosine bases, as the former will remain cytosine whereas the latter will have been converted to uracil by the deamination process. The bisulfite-modified DNA is then amplified through PCR and sequenced. Comparison of this sequence to the original pre-treated template sequence reveals the positions of methylated cytosines.

Since the development of the bisulfite treatment approach for DNA methylation analysis, various PCR-based methods have been conceived which utilise this same principal. Two distinct approaches have emerged, the first group of methods involves primers that amplify the template bisulfite-modified DNA regardless of methylation status. An example of this approach is the bisulfite sequencing method designed used by Frommer et al (1992), as discussed previously, and high-resolution melting (Worm et al., 2001).

High-resolution melting (HRM) can determine the methylation level of amplicons by analysing the melting curve of a PCR reaction with a bisulfite-converted DNA template (Worm et al., 2001). After bisulfite treatment and subsequent PCR, a methylated template will have a higher GC content than an unmethylated one. Therefore, these differences in base composition allow for a distinction to be made between unmethylated and methylated templates due to the stronger nature of the guanine/cytosine bond when compared to the adenine/thymine bond. This difference allows the methylated CpG content to be determined by the amount of energy required to denature the double stranded DNA (dsDNA). During the initial PCR reaction, a dsDNA-intercalating fluorescent dye

is incorporated and fluoresces when bound to dsDNA. The degree of fluorescence can be measured which allows the amplification process to be followed using a real-time PCR machine. PCR bias results in the preferential amplification of the unmethylated template due to differences in base composition and needs to be taken into consideration when performing this method (Warnecke et al., 1997). However, this hurdle can be addressed through primer design; Wojdacz and Hansen (2006) show that including a limited number of CpG sites within the primer allows for primer bias compensation, increasing amplification of the methylated sequence.

The second group of methods involves the use of different sets of primers for methylated and unmethylated converted sequences. This primer design approach is used in one of the more basic forms of DNA methylation analysis: methylation-specific PCR (Herman et al., 1996). In methylation-specific PCR, DNA is amplified by two sets of primers following bisulfite treatment, with a distinct primer set for methylated DNA and one for unmethylated DNA. Depending on which primer set produces a band, as visualised by agarose gel electrophoresis, indicates the methylation state of the investigated region. Primer design can be a challenging component of PCR-based bisulfite methods for DNA methylation analysis. In methods where a single primer set needs to bind to both methylated and unmethylated bisulfite-modified DNA, high CpG content is problematic. This is because CpG dinucleotides will cause different sequences to be present in unmethylated and methylated DNA after bisulfite treatment. Therefore, it is recommended that no CpG sites are included in the primers where they can be avoided (Clark et al., 1994). This is highly problematic for regions of the genome such as gene promoter regions which are typically GC rich especially near transcription start sites. For example, the *PXDN* promoter contains a CpG island, which overlaps the transcription start site, and as such designing CpG-free primers for this region is not possible (see Figures 7 and 8). Therefore, methods such as methylation-specific PCR which have different sets of primers for methylated and unmethylated converted sequences, can be helpful because they make it possible to analyse these GC rich regions of the genome. However, it is important to note that a major problem with this approach is incomplete bisulfite conversion. Another difficulty with this method is recognising sequences which are partially methylated.

Next generation sequencing-based methods exist for the analysis of DNA methylation. One example is bisulfite pyrosequencing, which involves the sequencing of bisulfite-modified DNA by measuring the luminescence caused by a luciferase reaction (York et al., 2012). The luciferase reaction is triggered by pyrophosphate release when a deoxynucleotide is incorporated during a

DNA synthesis reaction by DNA polymerase. Nanopore sequencing is a sequencing method which does not require prior bisulfite treatment. In this method, ssDNA is passed through a membrane-bound protein and changes in charge indicate which base is present. This technology is able to differentiate between cytosine and 5mC without any treatment of DNA prior to sequencing (Simpson et al., 2017). These sequencing methods are highly accurate and efficient, however they are very expensive.

The PCR bisulfite-based methods for DNA methylation analysis are cost effective and relatively simple, however there are a few drawbacks to the sodium bisulfite treatment approach. Bisulfite treatment is harsh and results in DNA fragmentation, which means that a large amount of DNA is required for the analysis since much of it is damaged during the treatment (Miura et al., 2012). Ultimate conversion of unmethylated cytosines to thymine means that the complexity of the region of interest is significantly reduced because only three bases are present instead of four. This phenomenon, combined with the fragmentation of DNA caused by bisulfite treatment means that it is often not possible to amplify long fragments (Kurdyukov and Bullock, 2016). Incomplete bisulfite conversion is another problem which is encountered with bisulfite-based methods since false positive results of CpG methylation can occur. If a sequencing method is being used for the analysis then this can be mitigated by identifying cytosine nucleotides which are not within CpG sites, therefore possibly indicating incomplete bisulfite conversion. However, this approach is not possible in methods such as methylation-specific PCR or HRM that do not give results at nucleotide or CpG resolution (Hernández et al., 2013).

The first method developed for the study of DNA methylation was based on the use of restriction endonucleases with selective digestion patterns (Cedar et al., 1979). Methylation sensitive restriction enzymes such as HpaII do not cleave their recognition site (5'-CCGG-3') in the presence of 5mC, whereas MspI is not sensitive to methylation, has the same recognition site as HpaII and will cleave irrespective of cytosine methylation. Methylation-sensitive PCR (MS PCR), not to be confused with the sodium bisulfite treatment-based method methylation specific PCR, is simple and cost-effective. This method avoids the complications which accompany bisulfite conversion and can be performed in a regular thermocycler. Other early research on DNA methylation of specific regions of the genome using the MS PCR method was performed by Singer et al. (1979), who digested mouse liver DNA with HpaII and MspI; and Singer-Sam et al. (1990) who performed PCR to quantify methylation levels of mouse spleen DNA after digestion with

HpaII to amplify the 5' end of the X-linked phosphoglycerate kinase gene. Both these studies made use of agarose gel electrophoresis to assess PCR amplification which indicated methylation levels of the regions they were investigating. These methods of studying methylation are only able to investigate methylation levels at the specific enzyme recognition sites and false positives are a problem due to incomplete digestion by restriction enzymes (Yegnasubramanian et al., 2006).

The high GC content of the *PXDN* promoter limits the choice of methods for DNA methylation analysis. This is because it is virtually impossible to design primers that do not contain CpG sites, especially around the TSS which is a region of interest (see Figure 7 and 8). It is possible the HRM method could be used for this purpose due to the fact that primers containing CpG sites can be used. Methylation-specific PCR is also an option for such a GC-rich region due to the fact that different primers can be used for the analysis of methylated and unmethylated templates. However, both of these methods require bisulfite conversion which adds a significant layer of complexity to the analysis and troubleshooting process as issues such as PCR bias, incomplete conversion and false positives need to be taken into account. It is clear that, in terms of efficiency and accuracy, next-generation sequencing-based methods such as nanopore technology are powerful tools when it comes to methylation analysis, however their cost makes them relatively inaccessible. The simplicity and cost-effective nature of the MS PCR method, as well as the ability of this method to analyse GC-rich regions, make it a good choice for initial investigations into the DNA methylation levels of the *PXDN* promoter.

1.9 Aims and Objectives

This study aimed to investigate whether methylation within the *PXDN* promoter could contribute to differences in expression levels of *PXDN* observed in breast cancer cell lines. All objectives were performed on the luminal A breast cancer cell line MCF-7, the TNBC cell line MDA-MB-231 and the human embryonic kidney cell line HEK-293. This last cell line was included because previous studies in our laboratory had shown *PXDN* expression within this cell line and at the time of the commencement of this study there was no data available for PXDN expression levels within the breast cancer cell lines. Therefore the HEK-293 cell line was used as a positive control for PXDN expression. The objectives were to:

1. Examine PXDN protein expression levels in invasive MDA-MB-231 and non-invasive MCF-7 breast cancer cell lines, using immunofluorescence microscopy.

2. Analyse and compare *PXDN* promoter CpG methylation levels of the invasive and non-invasive breast cancer cell lines through MS PCR.

2 Methods and Materials

2.1 Immunofluorescence Microscopy

Fluorescently tagged antibodies allow for the study of protein localisation, expression and activity through the tagging of specific targets (Giepmans et al., 2006). When light of the correct wave length is provided to the tagged antibody-bound proteins, the atoms within the fluorescent tag are excited and (if the protein is present and antibody-bound) will fluoresce to indicate the presence of the protein. Therefore the protein can be quantified and located in this way. For each cell line there were two control samples: one treated only with the secondary antibody to indicate non-specific binding of secondary antibody and the other with no antibodies at all to indicate background fluorescence of the cells. The experimental samples were treated with both the primary and the secondary antibodies.

2.1.1 Cell Culture

All cell culture reagents are Gibco (ThermoFisher Scientific, Waltham, Massachusetts, USA), unless otherwise specified. MCF-7 (ATCC-CRL-3435) ductal breast cancer cells were kindly provided by Dr Kutlwano Xulu and Prof Tanya Augustine who had recently acquired them from the American Type Culture Collection (ATCC). The MDA-MB-231 (ATCC-HTB-26) adenocarcinoma breast cancer cells were kindly provided by Prof. Mandeep Kaur, University of the Witwatersrand. The human embryonic kidney HEK-293 (ATCC-CRL-1573) cells were a gift from Dr Clement Penny, University of the Witwatersrand. All cells were cultured in a medium composed of 1:1 Dulbecco's Minimum Essential Medium (DMEM) to Ham's F12 supplemented with 10% foetal bovine serum (FBS), 1% penicillin streptomycin and 10 mM N-2-hydroxyethylpiperazine-N-2-ethane sulfonic acid (HEPES) buffer and incubated at 37°C with 5% CO_2. Cells were used at a passage number between 7 and 50 in line with guidelines from the European Collection of Authenticated Cell Cultures (ECACC) (https://www.culturecollections.org.uk/). Cells were split when a confluency of 70% to 90% was reached by discarding media, rinsing three times with 1 x phosphate buffered saline (PBS) then detaching cells by adding 900 µL of trypsin-EDTA and 300 µL 1 x PBS and incubating for 9 minutes (MCF-7 cells); 300 µL trypsin-EDTA and 700 µL 1 x PBS and incubating for three minutes (HEK-293 cells); or 1.5 mL TrypLE™ Express, which is more gentle on cells, and

incubating for three minutes (MDA-MB-231). After incubation trypsin-EDTA/TrypLE™ Express was deactivated with an equal volume of cell culture media and the appropriate volume (depending on confluency) of cell suspension was added back into culture dish with 8 mL fresh media. The approximate seeding density used was 1 x 10^4 cells/cm^2.

2.1.2 Fixing and Permeabilization of cells

Depending on the cell line, 80 000 to 120 000 cells were seeded onto four separate autoclaved glass cover slips in a six well plate and incubated at 37°C for 48 hours. For cell treatment, cells were treated with 10 nM β-Oestradiol and incubated for 24 hours. All wash steps were performed with 2 mL of 1X PBS. Culture medium was removed and cells washed once followed by fixation in 2 mL of 3% formaldehyde (Merck, Munich, Germany) which was diluted with 1X PBS for 10 minutes at room temperature followed by three washes. Cells were then permeabilised with 0.1% triton-X 100 (Merck, Munich, Germany) for seven minutes at room temperature and then washed three times.

2.1.3 Immunostaining: Primary Antibody

Cells were blocked with 2 mL of 1% bovine serum albumen (BSA) (Glentham Life Sciences, Corsham, United Kingdom) blocking solution for one hour to prevent non-specific binding of the primary antibody. The primary monoclonal antibody, mouse anti-PXDN (Santa Cruz Biotechnology, Dallas, Texas, USA) was added to two of the four coverslips in a 1:50 dilution within 1% BSA. The datasheet for this antibody suggested a primary antibody concentration of 1:200, however, the fluorescence at this concentration was very faint. Therefore, we increased the concentration of the primary antibody to 1:50. To the other coverslips 1% BSA blocking solution without any antibody was added only. Cells were then incubated for one hour at 37°C.

2.1.4 Immunostaining: Secondary Antibody and DAPI

The primary antibody solution was removed and cells washed five times (from this moment on all steps were performed in the dark). A 1:200 dilution of the goat anti-mouse Kappa-FITC seconday antibody (SouthernBiotech, Birmingham, Alabama, USA) was added in 1% BSA to the relevant coverslips. Initially we tried a secondary antibody concentration of 1:500 as suggested by the

antibody product sheet, however when we increased the concentration to 1:200 the images were better and we weren't seeing any non-specific binding in the controls. Only 1% BSA was added to the no antibody control coverslip. Cells were incubated in the dark for one hour at 37 ℃ followed by five washes. Cell nuclei were counterstained with 0.1 µg/ml 4, 6-diamidino-2-phenylindole (DAPI) (Merck, Munich, Germany) for five minutes which binds to adenine and thymine and therefore stains the nucleus of cells, which aids during visualisation. After staining with DAPI cells were washed three times.

2.1.5 Mounting and Visualisation

Coverslips were mounted face down onto microscope slides with Fluoromount solution (Merck, Munich, Germany) and allowed to dry in the dark at room temperature for two hours. After this cells can be stored in the dark at 4 ℃. Visualisation of cells was performed with the Olympus BX63 OFM microscope (Olympus, Tokyo, Japan). The FITC was excited at 490 nm and DAPI at 350 nm. The images were all taken using the 60X oil-immersion objective lens.

2.1.6 Image Analysis and Expression Quantification

One image from each treatment group was chosen per cell line. The Fiji Image J software version 9.2.0 was used to take fluorescence intensity measurements from three randomly chosen cells within each image. For each cell, three corresponding background readings were taken. The corrected total cell fluorescence (CTCF) values were calculated using the following equation:

CTCF = integrated density (cell) – (area of selected cell x mean fluorescence of background readings)

2.1.7 Statistical Analysis

Statistical analysis of the CTCF values was performed using the GraphPad Prism software version 9.5.1. An unpaired, parametric, Welch's correction t-test was performed to identify significance between mean CTCF values. Mean CTCF values were plotted on a graph and error bars calculated using standard deviation from the mean.

2.2 Methylation Sensitive PCR

This protocol involves DNA extraction followed by restriction with either a methylation insensitive (MspI) or methylation sensitive (HpaII) restriction enzyme. PCR was used to amplify across sites to determine if they had been restricted or not, indicated by absence or presence of a band in an agarose electrophoresis gel. This allowed for inference on the methylation status of that region to be made. The promoter of *PXDN* contains a high concentration of CpG sites within a CGI, and therefore the MS PCR protocol was considered the best-suited method to achieve this aim. The promoter of *PXDN* falls within a region of high GC%; a CGI. Figure 7 and Table 6 show the GC% values for a region of 1305 bp within the *PXDN* promoter which was examined for DNA methylation from 1037 bp upstream to 268 bp downstream of the transcription start site (TSS). Four sets of primers were used to investigate the methylation levels of the *PXDN* promoter region, covering a combined total of 13 MspI/HpaII recognition sites (Figure 7 and 8). All the products for section 2.2 are from New England Biolabs, Ipswich, United Kingdom unless otherwise stated.

2.2.1 Analysing the CG% of the *PXDN* Promoter and Restriction Enzyme Recognition Sites

Figure 7 shows the GC% within the promoter of *PXDN* which is highest between about 200 bp upstream and 100 bp downstream of the TSS. These CG% values informed the primer design process for the MS PCR protocol. Figure 8 shows the positioning of the CCGG MspI/HpaII recognition sites which could potentially be methylated and their position relative to the binding sites for the primer pairs designed. The control amplicon was designed to not contain any CCGG recognition sites.

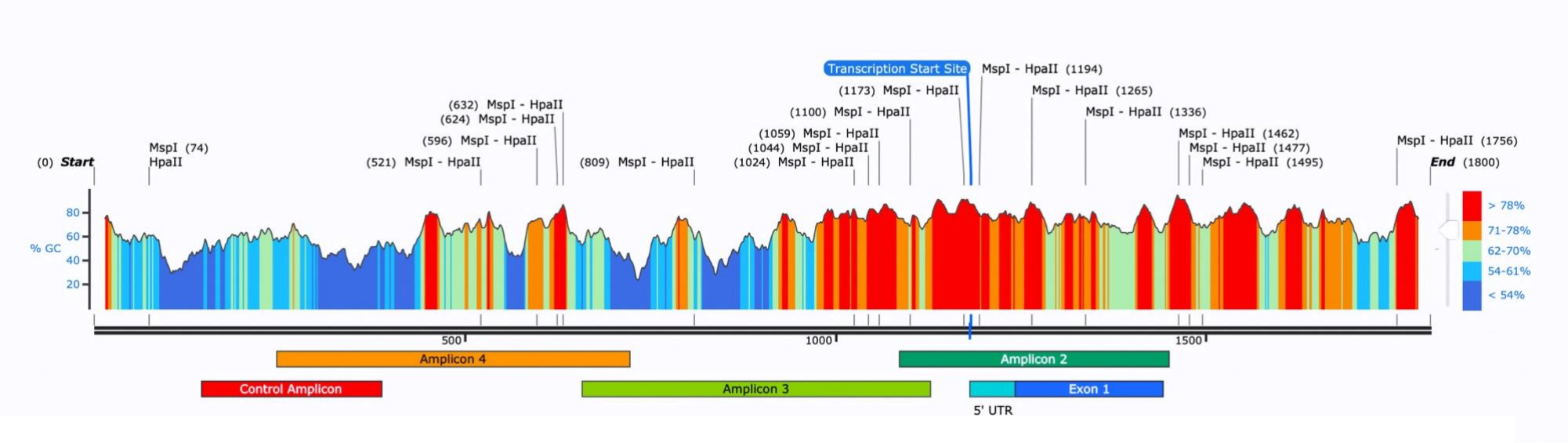


Fig. 7 The GC% of the *PXDN* promoter. The CG content of this region is highest between 208 bp upstream and 98 bp downstream of the TSS. CG% is lowest further upstream of the TSS, from about 750 bp upstream of the TSS. Four amplicons cover a region of 1305 bp. Amplicon 2 begins 95 bp upstream from the TSS and is 370 bp long containing five MspI/HpaII recognition sites. Amplicon 3 is 471 bp in length, begins 524 bp upstream of the TSS and contains five MspI/HpaII restriction sites. Amplicon 4 begins 935 bp upstream of the TSS, is 476 bp long and contains four MspI/HpaII restriction sites. The control amplicon contains no MspI/HpaII restriction sites; it begins 1037 bp upstream of the TSS and is 243 bp long.

Fig. 8 The locations of CCGG sites with respect to primer pairs and the TSS of the *PXDN* promoter. MspI/HpaII CCGG recognition sites can be seen in red. The P2A primer pair overlaps with exon 1 and is therefore downstream of the transcription start site. All the other primers are upstream of the 5' UTR (untranslated region). The control amplicon does not contain any CCGG sites.

2.2.2 DNA Extraction

2.2.2.1 Phenol Chloroform genomic DNA Extraction Method

The phenol chloroform DNA extraction method was used because of the high concentration and purity of high molecular weight genomic DNA the technique consistently produces (Ghaheri et al., 2016; Liu et al., 2022; Torii et al., 2021). All centrifugation steps were performed in a MiniSpin centrifuge (Eppendorf, Hamburg, Germany). Cells were grown to between 80% and 90% confluence, then media was removed and cells washed three times with 1 x PBS. Using a cell scraper, cells were detached in 1 mL of 1 x PBS and the suspension centrifuged at 12 000 x g. Supernatant was discarded and cells were resuspended in 50 µL of 1 x PBS. RNA was degraded by adding 1 µL of 10 mg/mL RNase A (20 µg/mL final concentration) and then 450 µL of cell lysis buffer (10 mM Tris-HCl pH 8; 100 mM EDTA pH 8; 2% SDS) was added followed by an hour incubation at 37 ℃. To remove protein, 2.5 µL of 18 mg/mL stock proteinase K was added (100 µg/mL final concentration). This was followed by a further two hours of incubation at 50 ℃ with vortexing every 20 minutes.

Phenol:chloroform:isoamyl alcohol (Merck, Munich, Germany) was added in a 25:24:1 ratio so that the total volume in each tube was 1 mL. Samples were centrifuged at 16 000 x g for 25 minutes at room temperature and the upper aqueous phase removed and placed in a new tube. This step was repeated. Chloroform:isoamyl alcohol (24:1) was added to the second aqueous phase followed by vortexing for 25 minutes at 4 ℃. The upper aqueous phase was removed and placed in a new tube. This step was repeated and then to the fourth aqueous phase 0.1 volumes of 3 M sodium acetate and 2 volumes of 100% ethanol was added. Tube was stored overnight at -20 ℃.

DNA was collected by centrifugation at 16 000 x g for 15 minutes at 4 ℃. Pellet was washed with 750 µL of 70% ethanol three times by centrifugation at 16 000 x g for five minutes each at 4 ℃. The pellet was air-dried at room temperature and then resuspended in ddH_2O preheated to 60 ℃ to assist with the dissolving of the pellet. DNA concentration was assessed with the NanoDrop spectrophotometer (ThermoFisher Scientific, Waltham, Massachusetts, USA) and quality observed using a 1% agarose electrophoresis gel for 30 minutes at 100 V (method described under 2.4.4). See appendix for an example of DNA extraction results.

2.2.2.2 Genomic DNA Extraction Kit Method

When troubleshooting problems with the MS PCR protocol, it was suspected that inhibitors may have been present within gDNA samples which were carried over from the phenol chloroform DNA extraction method. Therefore, the GeneJet genomic DNA purification kit (ThermoFisher Scientific, Waltham, Massachusetts, USA) was used to compare PCR results from samples extracted using both methods. All the reagents in section 2.2.2.2 are from ThermoFisher Scientific, Waltham, Massachusetts, USA unless otherwise specified. Media was removed and cells washed three times with 1 x PBS, then using a cell scraper cells were detached in 1 mL of 1 x PBS and the suspension centrifuged at 250 x g for five minutes. Supernatant was discarded and cells were resuspended in 200 µL of 1 x PBS. A volume of 200 µL of Lysis solution and 20 µL of proteinase K solution were added to cell suspension which was then gently vortexed and incubated at 56 ℃ for 10 minutes. To digest RNA, 20 µL of RNase A solution was added to the tube followed by vortexing and a further 10 minutes incubation period at room temperature. An amount of 400 µL of 50% ethanol was added, the solution was gently mixed by pipetting and then transferred to a GeneJet Genomic DNA Purification Column within a collection tube. The column was centrifuged for one minute at 6000 x g then placed in a new collection tube (the flow-through was discarded). This was followed by the addition of 500 µL of wash buffer I to the column, which was centrifuged at 8000 x g for one minute before discarding the flow through. For the final wash step, 500 µL of wash buffer II was added, followed by a three minute centrifugation at 16 000 x g. The flowthrough was discarded and the column placed in a 1.5 mL nuclease-free microcentrifuge tube. Lastly, 150 µL of ddH_2O was added to the column, incubated at room temperature for two minutes and then centrifuged at 8000 x g for one minute to elute the DNA. DNA concentration was assessed with the NanoDrop spectrophotometer (ThermoFisher Scientific, Waltham, Massachusetts, United States) and quality observed using a 1% agarose electrophoresis gel for 30 minutes at 100 V (method described under 2.2.5). See appendix for example of DNA extraction results.

2.2.3 Restriction Digest

HpaII and MspI are restriction enzymes which are methylation sensitive and insensitive respectively. The restriction digest is performed before PCR because the methyl groups bound to the template will not be replicated along with the template during PCR, therefore the differential digestion by the two different enzymes will only be informative with regard to DNA methylation

before amplification. The reaction components are listed below in Table 2. For each cell line, three digest reactions were set up: 1) gDNA restricted with the MspI methylation insensitive restriction enzyme. 2) gDNA restricted by the HpaII methylation sensitive restriction enzyme, and 3) a no enzyme control where gDNA has been incubated at the same concentration with the same buffer, however an extra 1 µL ddH_2O was added instead of a restriction enzyme. The digest was incubated at 37 ℃ for one hour and the HpaII enzyme was inactivated by incubation at 80 ℃ for 20 minutes. The MspI restriction enzyme does not require an inactivation step. Agarose gel electrophoresis on a 1% gel for 30 minutes at 100 V was used to check the quality of digest (method described under 2.4.4).

Table 2: Components and volumes of MspI/HpaII restriction digest reactions

Component	Volume (µL)
gDNA (100 ng/µL)	7
rCutSmart Buffer	5
Restriction Enzyme (MspI or HpaII)*	1
ddH2O	Up to 50

*20 000 units/mL

2.2.4 PCR

After the restriction digest, PCR was used to amplify the DNA, which revealed the presence or absence of CpG methylation within regions of the *PXDN* promoter due to differential digestion by the two restriction enzymes depending on the methylation status of the amplified region. Kapa Taq ReadyMix PCR Kit (Merck, Munich, Germany) was the DNA polymerase mastermix used (0.5 U Taq DNA Polymerase, 0.2 mM of each dNTP and 1.5 mM $MgCl_2$ at 1X within a 25 µL reaction). The components of the PCR reaction are detailed in Table 3. For each primer pair and cell line, four PCR reactions were set up: 1) MspI-digested gDNA, 2) HpaII-digested gDNA, 3) no enzyme control gDNA, 4) no template control (NTC).

Table 3: PCR reaction components and volumes for amplification of regions within the *PXDN* promoter

Component	Volume (µL)
Restricted gDNA (14 ng/µL)	7.1
2X KAPA Taq ReadyMix with dye	12.5
Forward Primer (10 µM)	1
Reverse Primer (10 µM)	1
ddH2O	Up to 25

The four primer sets used to amplify part of the CpG island found in the promoter of *PXDN* are shown below (Tables 4 and 5) and were purchased from Inqaba Biotec, Pretoria, Gauteng. P2A and the control primers were designed in NCBI Primer Blast, P3 and P4 were designed previously in our laboratory (Hanmer and Mavri-Damelin, 2018). The sequences of the primer pairs used in this study, as well as the details of their amplicons such as GC% and the number of MspI/HpaII recognition sites are detailed in Tables 4 and 5 respectively. The PCR reaction conditions are detailed in Table 6. The results of the PCR reaction were analysed by agarose gel electrophoresis on a 1.7% gel at 100 V for 45 to 50 minutes.

Table 4: Primers designed for the amplification of regions within the *PXDN* promoter

Primer	Forward Primer (5' to 3')	Reverse Primer (5' to 3')
P2A	CCTCGGGGATTCAGAGGGG	GCACTCACAGGATGGAGGTC
P3	CAGACTCCCTTGCTGTGCGCTTTG	AGCTGTGCACATGCGCGAGGCT
P4	TCTGAATCTGGCACCGTCACCGTC ACCCTG	AGCTGTGCACATGCGCGAGGCT
Control	TCCCATTCCAGGCTGCTTTC	ATACGCACAAAGGTGGCGTT

Table 5: The amplicons within the *PXDN* promoter and their primer parameters

	Position (with reference to transcription start site)	Length (bp)	Number of MspI/HpaII recognition sites	GC%	Annealing Temperature of primers (℃)
P2A	-95 bp to 268 bp	370	5	76.21	58
P3	-524 bp to -53 bp	471	5	62.63	58
P4	-935 bp to -459 bp	476	4	59.03	55
Control	-1037 to -794	243	0	53.91	58

Table 6: PCR reaction conditions

Step	Temperature (℃)	Duration (minutes)	Number of Cycles
Initial denaturation	95	5	1
Denaturation	95	0.5	
Annealing	See table 5	0.5	35
Elongation	72	1	
Final extension	72	5	1

It became necessary to amplify a region of the genome separate to that of *PXDN* (see section 3.2). Therefore, the following primer pair shown in Table 7 was used to amplify a region of the *TP53* promoter on chromosome 17. The amplicon is 805 bp and the primers have an annealing temperature of 59.3 ℃.

Table 7: Primer pair designed to amplify the *TP53* promoter

Primer	Forward Primer (5' to 3')	Reverse Primer (5' to 3')
TP53	TGGGAGTTGTAGTCTGAACGCTTC	GAGAAGCTCAAAACTTTTAGCGCC

2.2.5 Agarose Gel Electrophoresis

Agarose gels were prepared for the assessment of gDNA quality after extraction, and the analysis of all MS PCR results. The appropriate mass of agarose powder was added to 50 ml 1 X Tris acetate EDTA (TAE) Buffer. Agarose powder (Cleaver Scientific, Rugby, United Kingdom) and buffer solution was heated using a microwave until the powder completely dissolved. A volume of 5 µL of DNA intercalating agent ethidium bromide (10 mg/ml) (Merck, Munich, Germany) was added after the agarose had cooled sufficiently and then the solution was poured into a casting tray and left to set for roughly 30 minutes. The KapaTaq DNA polymerase master mix already contains a dye and therefore no loading dye was needed for samples which had undergone PCR. However, when analysing quality of extracted DNA, samples were mixed with a 6X DNA loading dye in a ratio of 5:1. Then 10 µL of each sample and 5 µL of the NEB 100 bp molecular weight marker or the NEB 1 kb molecular weight marker, were added to the wells and the gel electrophoresed at 100 V within 1X TAE buffer for the appropriate amount of time depending on product size. Gel

images were captured by the Image Lab 6.0 program linked to the Bio-Rad Gel Doc XR system; this system uses ultraviolet light to show the position of DNA within the gel.

2.2.5.1 Interpreting the MS PCR Results

The MS PCR agarose electrophoresis gels were loaded in a consistent pattern of samples, for ease of interpretation (see Figure 9). Lane 1 contains the molecular weight marker indicating band size. Lane 2 was loaded with the amplified template digested by methylation insensitive MspI; irrespective of template methylation status a band should not appear if the template contains enzyme recognition sites. Lane 3 indicates the methylation status of the template: a fully methylated template will not be cut at CCGG sites by the methylation sensitive HpaII enzyme and will therefore show a band on the gel. A template where at least one of the HpaII recognition sites is unmethylated should be cut by the enzyme and will not show a band on the gel. The no enzyme control sample in lane 4 should show a band regardless of methylation status. The control amplicon, which contained no MspI/HpaII recognition sites, should always have appeared with three bands in lanes 2, 3 and 4. It is important to note that due to the nature of the method, a 'methylated' result indicates that all of the CCGG sites within the amplicon are methylated. An 'unmethylated' result indicates that at least one of the CCGG sites present within the amplicon are unmethylated. It is not clear how many of the restriction sites or which of them are unmethylated as all the regions examined, except the control amplicon, contained more than one CCGG site (see Table 5).

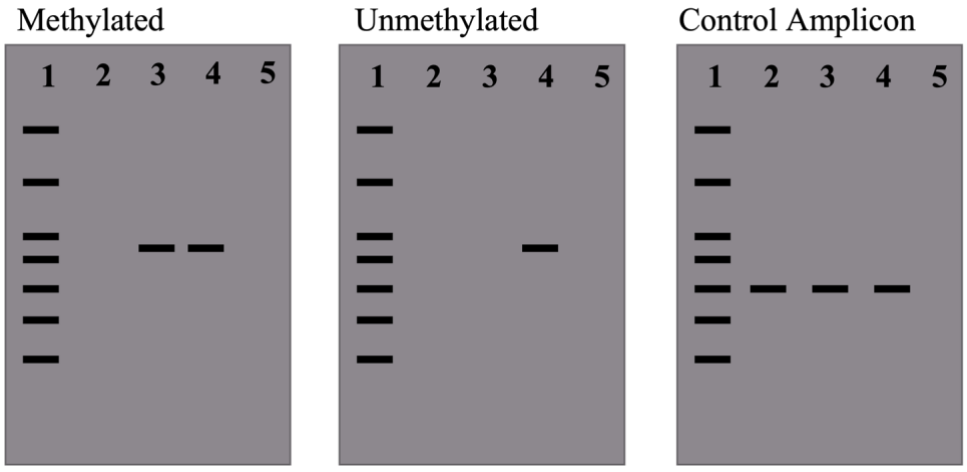


Fig. 9 Methylation sensitive PCR agarose gel electrophoresis layout of amplified methylated and unmethylated templates. Lane 1 contained the 100 bp molecular weight marker; lane 2 the MspI-digested template; in lane 3 the methylation-sensitive HpaII-digested template amplicon was

loaded. Lane 4 was an unrestricted amplified template. Lane 5 was a no template control for the PCR reaction.

3 Results

3.1 Immunofluorescence Microscopy

The immunofluorescence microscopy images were used to analyse PXDN expression levels. Blue light in merged images indicates the position of the nucleus through fluorescence of the nuclear counterstain DAPI. Green light is indicating FITC fluorescence which is conjugated to the secondary antibody, showing PXDN protein expression. The no antibody control (NAC) and no primary control (NPC) show autofluorescence of cells and non-specific binding of the secondary antibody respectively under green light (490 nm).

A difference in expression of PXDN is clearly observable between the treated and untreated images of the MDA-MB-231 cell line: the latter appears brighter, indicating higher endogenous PXDN expression levels (Figure 10). The treated and untreated images show more fluorescence than the controls in spindle-like projections between cells (see red arrows in figure 10). This suggests PXDN is localising in the ECM as suggested by the literature. However, there could be some localisation within the nucleus as well because green fluorescence is visible in these regions but at a lower intensity. For the MDA-MB-231 cell line, the treated and untreated samples show higher green fluorescence than the controls (Figure 10).

In the MCF-7 cell line the treated sample appears brighter than the untreated one (Figure 11). The treated and untreated MCF-7 images showed more green fluorescence than both controls, particularly between and around the cells. The red arrows in Figure 11 indicate regions that appear to be extracellular and suggest localisation of PXDN to the ECM. The treated HEK-293 sample is slightly brighter than the untreated sample and, similar to the other two cell lines, the regions of most intense fluorescence seem to be those outside of the cell rather than within the nucleus or cytoplasm (highlighted by red arrows in Figure 12). This suggests that PXDN is localising within the ECM (Figure 12). Protein quantification was performed by acquiring CTCF values from the images in Figures 10 to 12 using Fiji Image J.

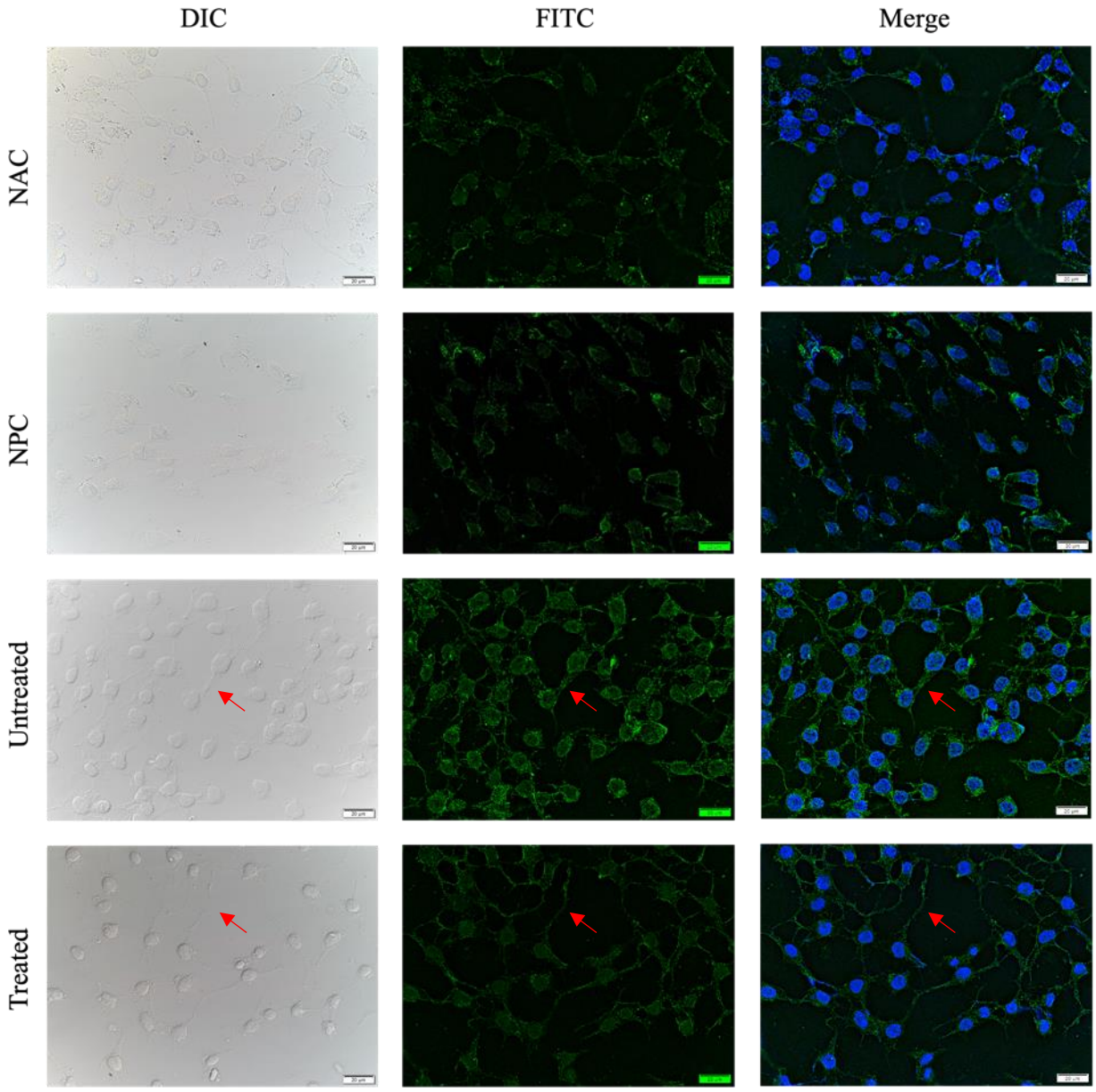


Fig. 10 PXDN expression in MDA-MB-231 cells. Highest PXDN expression can be observed in the untreated MDA-MB-231 cells. In the treated and untreated samples extracellular matrix structures between cells, highlighted by red arrows, are more visible than in the controls, suggesting PXDN's localisation in the ECM.

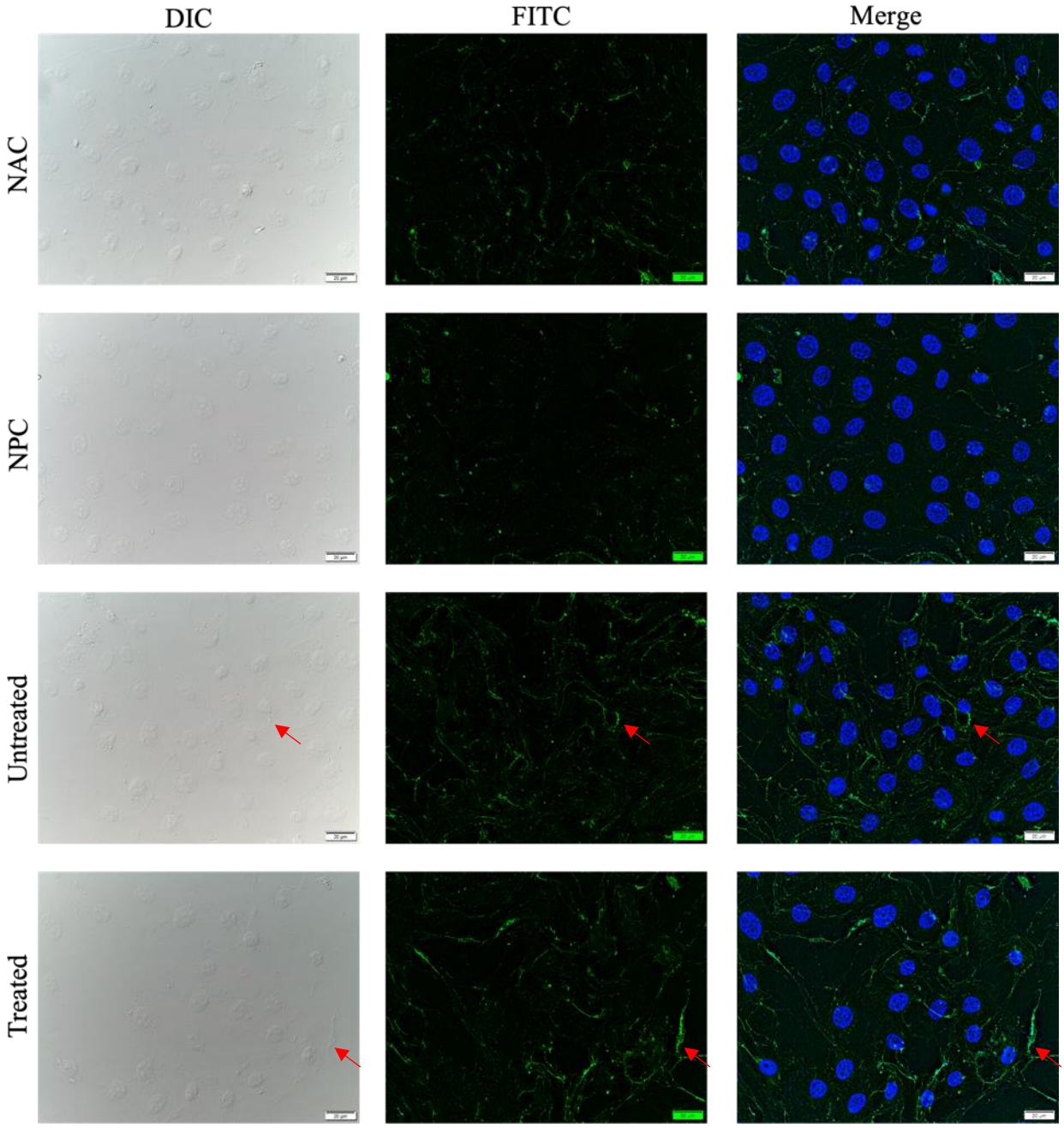


Fig. 11 PXDN expression in MCF-7 cells. There is higher fluorescence in the treated and untreated samples than in the control which indicates there is PXDN expression, however it is not clear which of these samples is showing higher PXDN expression by merely looking at the images. CTCF analysis was carried out to determine this (see Figure 12). Red arrows indicate regions which show high levels of green fluorescence and therefore suggest PXDN localisation.

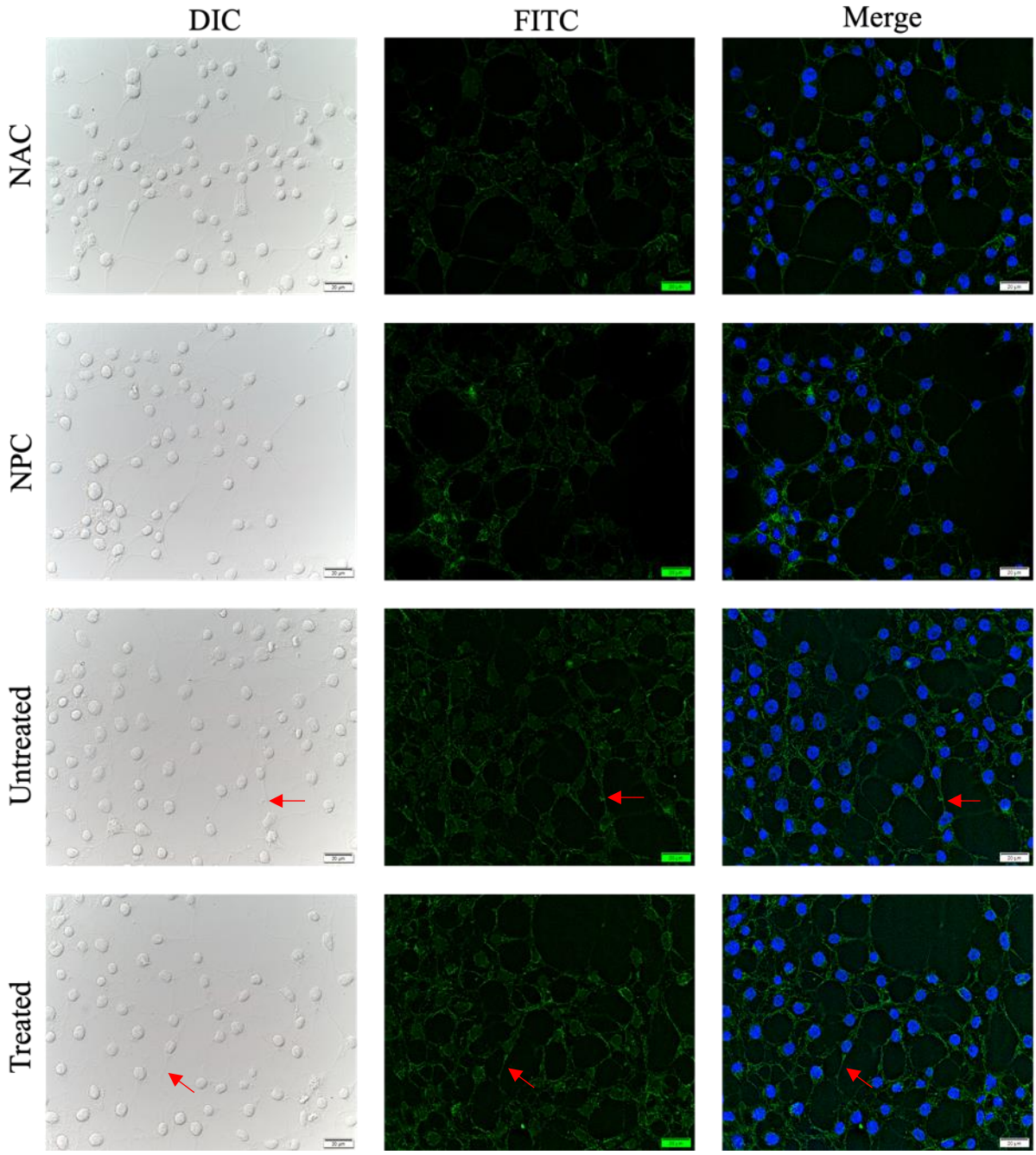


Fig. 12 PXDN expression in HEK_293 cells. It is difficult to assess differences in PXDN expression between the treated and untreated samples, semi-quantitative analyses of the images are required to clarify these differences (see Figure 12). The treated and untreated samples of HEK-293 cells are brighter than the control samples. The ECM regions between cells show brighter fluorescence in the treated and untreated samples than in the controls, examples of these regions are indicated by red arrows.

A graph depicting the average CTCF values for each sample and cell line is shown in Figure 13. The control samples for each cell line indicate the contribution of non-specific secondary antibody binding and autofluorescence to CTCF values. A t-test between the NAC and NPC CTCF values for each cell line revealed P-values of 0.26, 0.76 and 0.14 for the MDA-MB-231, MCF-7 and HEK-293 cell lines respectively. Therefore, expression differences between controls were insignificant for all cell lines. This establishes that there was no significant contribution of non-specific binding by the secondary antibody to fluorescence over and above that of cellular autofluorescence at the excitation wavelength.

The average CTCF values for treated and untreated MDA-MB-231 green light channel images were $1.26 \times 10^5 \pm 0.03 \times 10^5$ and $2.66 \times 10^5 \pm 0.24 \times 10^5$ arbitrary units respectively and a Welch's t-test gave a P value of 0.0089 indicating significance. Therefore, the untreated MDA-MB-231 samples showed average CTCF readings more than double that of the treated samples, indicating that β-oestradiol treatment decreased PXDN expression within the MDA-MB-231 cells (see Figure 13). The NPC and NAC MDA-MB-231 samples showed average CTCF values of $0.66 \times 10^5 \pm 0.05 \times 10^5$ and $0.57 \times 10^5 \pm 0.09 \times 10^5$ arbitrary units respectively which were significantly lower than both the treated and untreated samples within the same cell line. This confirms that the PXDN expression values from the images were due to differences in protein expression between the two samples and not autofluorescence of cells.

In the MCF-7 images the treated and untreated average CTCF values were $1.33 \times 10^5 \pm 0.08 \times 10^5$ and $0.76 \times 10^5 \pm 0.12 \times 10^5$ respectively, a Welch's t-test between these values showed significance with a P value of 0.0037. These results show that the average CTCF value of the untreated sample was a little over half that of the treated one, indicating that β-oestradiol treatment increased MCF-7 PXDN expression (Figure 13). The NPC and NAC values were $0.43 \times 10^5 \pm 0.05 \times 10^5$ and $0.41 \times 10^5 \pm 0.05 \times 10^5$ respectively, and similar to the MDA-MB-231 cell line, the control values were both significantly lower than the treated and untreated values.

β-oestradiol treatment increased PXDN expression within the HEK-293 cell line. The untreated HEK-293 average CTCF value was $0.56 \times 10^5 \pm 0.09 \times 10^5$ and the treated value was $1.09 \times 10^5 \pm 0.18 \times 10^5$. The P-value for the t-test between the untreated and treated samples indicated significance at 0.0201. The NPC and NAC average CTCF values were $0.49 \times 10^5 \pm 0.08 \times 10^5$ and $0.37 \times 10^5 \pm 0.04 \times 10^5$ arbitrary units respectively, and were not significantly different from one another (P = 0.1414). Both of the control samples were

significantly different from the treated sample, however only the NAC sample significantly differed from the untreated sample. The t-test for NAC vs untreated gave a P value of 0.0535, which indicates that the background fluorescence CTCF value is significantly different to the PXDN-bound FITC fluorescence CTCF value. The P value for the NPC vs untreated t-test was 0.3667, this suggests that there is no significant difference between the fluorescence of non-specifically bound secondary antibody fluorescence and fluorescence indicating PXDN expression. This could indicate that the HEK-293 endogenous PXDN expression is lower than what was indicated by the CTCF value analysis of the images.

β-Oestradiol treatment seems to have a contrasting effect on the PXDN expression within MDA-MB-231 cells compared to that observed within the HEK-293 and MCF-7 cell lines (Figure 13). The MDA-MB-231 β-Oestradiol-treated sample showed roughly half the expression of the untreated sample (P = 0.0089) whereas the treatment had the opposite effect on the MCF-7 and HEK-293 cell lines, increasing PXDN expression significantly (P = 0.0037 and 0.0201 respectively). Treated samples between cell lines did not show significant differences in expression, however the untreated MDA-MB-231 and HEK-293 cells and the untreated MDA-MB-231 and MCF-7 cells showed significant differences in PXDN expression (P = 0.0018 and P = 0.0014 respectively). The highest overall PXDN expression was recorded in the untreated MDA-MB-231 cells, followed by the β-Oestradiol-treated samples of all three cell lines (which all showed similar *PXDN* expression).

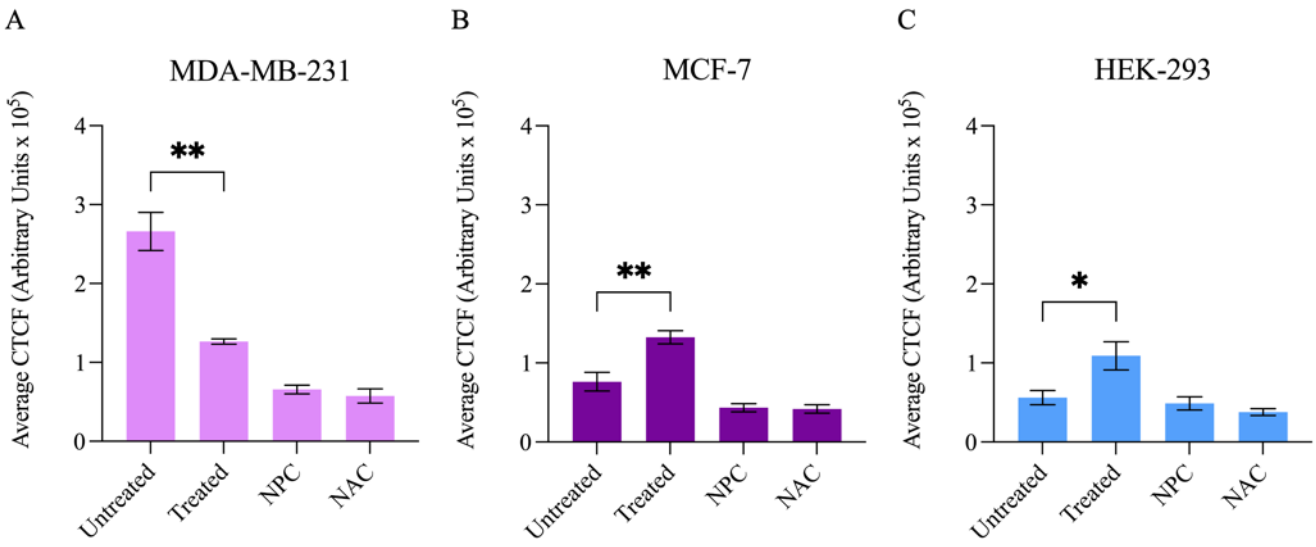


Fig. 13 PXDN expression levels in MCF-7, MDA-MB-231 and HEK-293 cell lines. Expression levels were evaluated from CTCF values of the cells within immunofluorescence microscopy images (figures 9-11). Highest overall PXDN expression appears in the untreated MDA-MB-231 sample, which is significantly greater than the treated sample of the same cell line (P = 0.0089). The treated samples of both the MCF-7 and the HEK-293 cells are significantly greater than the untreated samples (P = 0.0037 and P = 0.0201). There is no significant difference between the NPC and NAC samples in all three cell lines, indicating that minimal non-specific binding of the secondary antibody took place. Results were gathered from a single cell culture repeat.

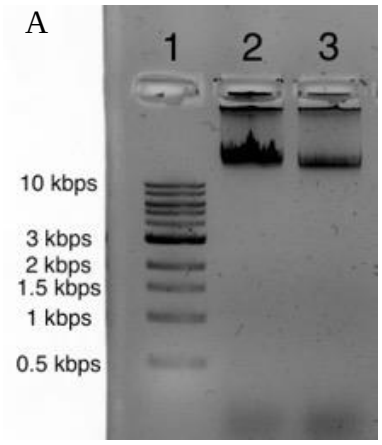
3.2 Methylation Sensitive PCR

Multiple DNA extraction protocols were performed through the duration of this study. After extraction, spectrophotometry and agarose gel electrophoresis were used to assess concentration, purity and quality of extracted gDNA. An example of an agarose gel depicting gDNA extracted by the phenol chloroform protocol is shown in Figure 14A. The corresponding absorbance ratios show the DNA is relatively pure and does not contain contaminants such as phenol or salts (Figure 14B). The A260/A280 ratios are good for gDNA, ratios less than 1.8 are problematic and could indicate phenol contamination (Lucena-Aguilar et al., 2016), however this is not the case. A260/A230 ratios are roughly at 2.3 which is also good for gDNA. Overall, gDNA appears to be of good quality, intact, high molecular weight and pure.

Once the purity and concentration of the extracted gDNA had been established, a restriction digest reaction was set up and followed by PCR to amplify the restricted DNA. The numbers of the wells (one through five) are constant i.e. well 1 is the molecular weight marker; 2 is MspI-digested gDNA; 3 contains HpaII-digested gDNA; 4 is loaded with undigested gDNA; and 5 contains the no template control (see Figure 9).

The four primer pairs presented in the next few figures which amplify the *PXDN* promoter are as follows: amplicon 2 begins 95 bp upstream from the TSS and is 370 bp long containing five MspI/HpaII recognition sites. Amplicon 3 is 471 bp in length, begins 524 bp upstream of the TSS and contains five MspI/HpaII restriction sites. Amplicon 4 begins 935 bp upstream of the TSS, is 476 bp long and contains four MspI/HpaII restriction sites. The control amplicon contains no

A



B

Lane	Sample	Nucleic Acid Concentration (ng/µL)	A260/280	A260/230
2	MCF-7 phenol chloroform-extracted gDNA	388.392	1.907	2.358
3	MCF-7 phenol chloroform-extracted gDNA	491.034	1.922	2.332

Fig. 14 Agarose electrophoresis gel and spectrophotometry readings for MCF-7 phenol chloroform-extracted gDNA. A Extracted gDNA samples were electrophoresed on a 1% agarose gel for 30 minutes. Clear bands greater than 10 kilobase pairs in size and the absence of smears indicate intact, undegraded high molecular weight gDNA. **B** DNA samples were resuspended in 200 µL of ddH_2O, readings indicate relatively high concentrations greater than 200 ng/µL with concentrations indicated as ng/µL. A260/A280 and A260/A230 ratios were also captured.

MspI/HpaII restriction sites; it begins 1037 bp upstream of the TSS and is 243 bp long (see Table 5).

For assistance with interpreting the MS PCR results, see figure 9. Briefly, lane 1 contains the ladder, lane 2 is the MspI-digested template. Lane 3 is the HpaII-digested template, followed by lane 4 showing the unrestricted template which should always show a band. Lane 5 was loaded with a no template control and should always show no band. In the MDA-MB-231 cell line, the P2 and P3 amplicons were unmethylated (Figure 15A and 15B), whilst the P4 amplicon was methylated (Figure 15C). This shows that in the MDA-MB-231 cell line, within the region of the promoter that was investigated, the promoter of *PXDN* is methylated further upstream from and unmethylated further downstream and closer to the TSS. Figure 15D shows the control amplicon designed to span across a region with no MspI and HpaII sites and as such should always give a positive PCR product as seen in Figure 15D.

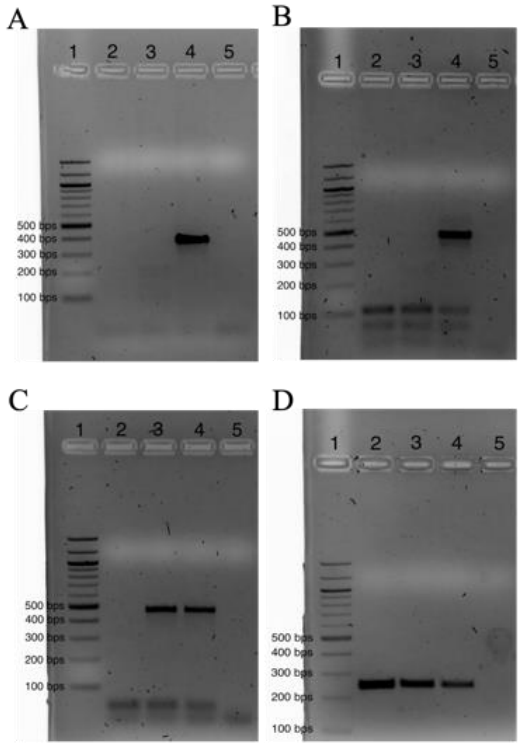


Fig. 15 Methylation sensitive PCR in the MDA-MB-231 cell line. The lanes 1 to 5 are constant for the above images: (1) 100bp NEB MWM (2) MspI restricted gDNA (3) HpaII restricted gDNA (4) unrestricted gDNA (5) no template control **A** Amplified with the P2A primer pair which produces an amplicon of 370 bp. **B** Amplified with the P3 primer pair which produces an amplicon of 471 bp. **C** Amplified with the P4 primer pair which produces an amplicon of 476 bp. **D** Amplified with the CNTRL primer pair which produces an amplicon of 243 bp. These results were produced in triplicate.

Figure 16 shows the methylation status of the *PXDN* promoter within the HEK-293 cell line. Figure 16A contains some faint bands in lane 3, however the band in lane 4 is darker and therefore indicates that the *PXDN* promoter in the HEK-293 cell line is most probably unmethylated at the P2 amplicon. Amplicons P3 (Figure 16B) and P4 were methylated, however there is a faint band in lane 2 of Figure 16C which suggests an incomplete digest of amplicon P4 by the MspI restriction enzyme. It is still clear that the enzyme did digest the template, even if this digestion was incomplete, and therefore the methylation status of this amplicon remains relatively unambiguous (Figures 16B and 16C). Figure 16D shows successful amplification of the control amplicon in unrestricted and restricted samples. This positive control confirms that the absence of bands in the gels of the other amplicons, which contain MspI and HpaII recognition sites, is not due to failure of the PCR protocol.

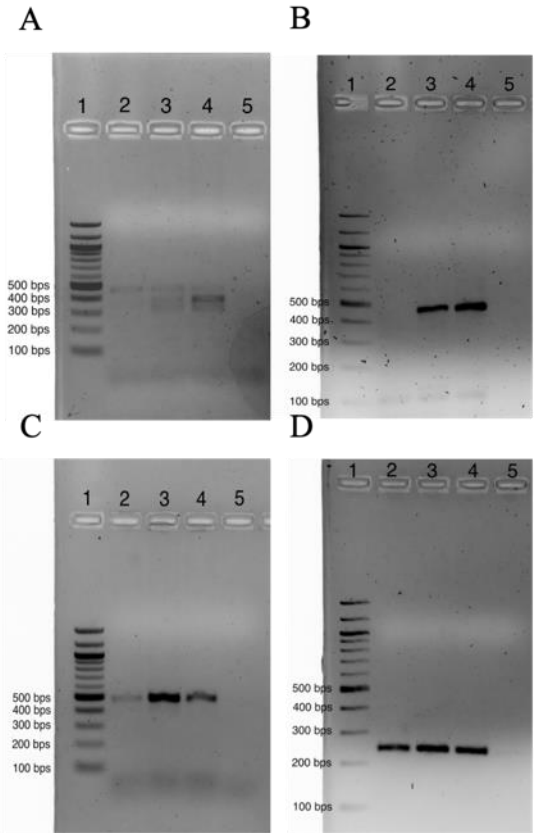
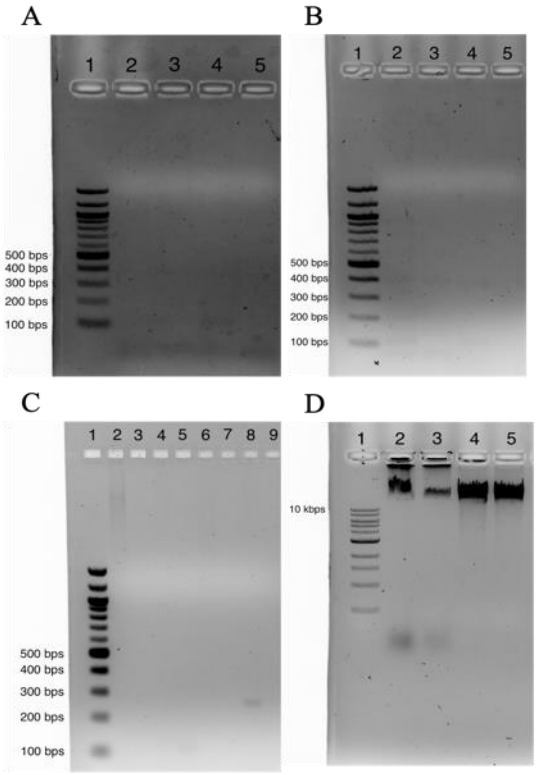


Fig. 16 Methylation sensitive PCR in the HEK-293 cell line. The lanes 1 to 5 are constant for the above images: (1) 100bp NEB MWM (2) MspI restricted gDNA (3) HpaII restricted gDNA (4) unrestricted gDNA (5) no template control **A** Amplified with the P2A primer pair which produces an amplicon of 370 bp. **B** Amplified with the P3 primer pair which produces an amplicon of 471 bp. **C** Amplified with the P4 primer pair which produces an amplicon of 476 bp. **D** Amplified with the CNTRL primer pair which produces an amplicon of 243 bp. The results shown in the above gels were carried out in triplicate for all images except A.

Figure 17 shows the beginning of the troubleshooting process which was adopted when the region of interest within the *PXDN* promoter proved difficult to amplify in MCF-7 cells. Figure 17A and 17B show that no bands were produced in the agarose gel for the P2 or P3 amplicons after restriction digest and PCR. Figure 17C shows the results of PCR performed with all primer pairs on the unrestricted MCF-7 template gDNA. None of the primer pairs were able to amplify this region of the *PXDN* promoter within the MCF-7 cell line (Figure 17C). Therefore, to ensure that there were no inhibitors present within the MCF-7 DNA which would have prevented amplification, the phenol chloroform extraction protocol was repeated.

During this repetition, in addition to the extraction of MCF-7 gDNA by the phenol chloroform protocol, MCF-7 DNA was extracted using the GeneJet genomic DNA purification kit (ThermoFisher Scientific) to check if a different DNA extraction method would assist in amplifying the *PXDN* promoter. The GeneJet kit was also used to extract DNA from the HEK-293 cells so that a positive control template which had successfully amplified the *PXDN* promoter could assist in the troubleshooting process. An agarose gel of extracted DNA and the corresponding absorbance ratios are shown in Figures 17D and 17E. The quality of the extracted MCF-7 DNA was pure and of sufficiently high concentration.

After the successful extraction of DNA from the relevant cell lines as indicated in Figure 17D and 17E, PCR was carried out using the phenol chloroform-extracted MCF-7 DNA, the kit-extracted MCF-7 DNA and the HEK-293 kit-extracted DNA. Figure 18A shows that the PCR reaction was successful for all the primer pairs when the HEK-293 DNA was used as a template. However, when the MCF-7 DNA (extracted by either protocol) was used, amplification was unsuccessful. Figure 18B shows a repeat of the experiment carried out in Figure 18A, with the same result. Therefore it was decided that there could be a biological difference in *PXDN* promoter region between the two cell lines inhibiting amplification of the region within the MCF-7 cell line.



E

Lane	Sample	Nucleic Acid Concentration (ng/µL)	A260/280	A260/230
2	MCF-7 phenol chloroform-extracted gDNA	253.093	1.881	2.235
3	MCF-7 phenol chloroform-extracted gDNA	109.623	1.763	2.237
4	MCF-7 kit extracted gDNA	110.241	1.847	2.216
5	HEK kit extracted gDNA	113.217	1.822	2.191

Fig. 17 Troubleshooting difficulties amplifying the *PXDN* promoter in DNA extracted from MCF-7 cells. The lanes 1 to 5 are constant for the above images: (1) 100bp NEB MWM (2) MspI restricted gDNA (3) HpaII restricted gDNA (4) unrestricted gDNA (5) no template control. **A** Amplified with the P2A primer pair **B** Amplified with the P3 primer pair **C** PCR of unrestricted MCF DNA using all primer pairs. The contents of the lanes were as follows: (1) 100 bp NEB MWM (2) P2A reaction (3) P2A no template control (4) P3 reaction (5) P3 no template control (6) P4 reaction (7) P4 no template control (8) CNTRL reaction (9) CNTRL no template control. No bands were produced in any of the above reactions, except for a very faint band at the correct position (243 bp) within the CNTRL primer pair reaction. **D** Repeat of gDNA extraction using the phenol chloroform and kit extraction methods in MCF-7 and HEK-293 cells. Clear bands greater than 10 kilobase pairs in size with no smearing indicates undegraded gDNA. **E** DNA samples were resuspended in 200 µL of ddH_2O, readings indicate relatively high concentrations greater than 200 ng/µL with concentrations indicated as ng/µL. A260/A280 and A260/A230 ratios were also captured.

The *TP53* gene is located in a different region of the genome to *PXDN*, a set of primers amplifying a region of *TP53* were used next to check if there was something occurring within the MCF-7 *PXDN* promoter that was inhibiting PCR (the details of the *TP53* primer pair can be found in Table 7 in section 2). Figure 18C shows clear bands in lanes 2 to 4, indicating that the *TP53* promoter was successfully amplified by all DNA samples. It is important to note that exactly the same DNA aliquots were used in this set of reactions as in the PCR reactions set up for Figure 18B and that the reactions were run concurrently. The possible mechanisms for this discrepancy between cell lines will now be addressed.

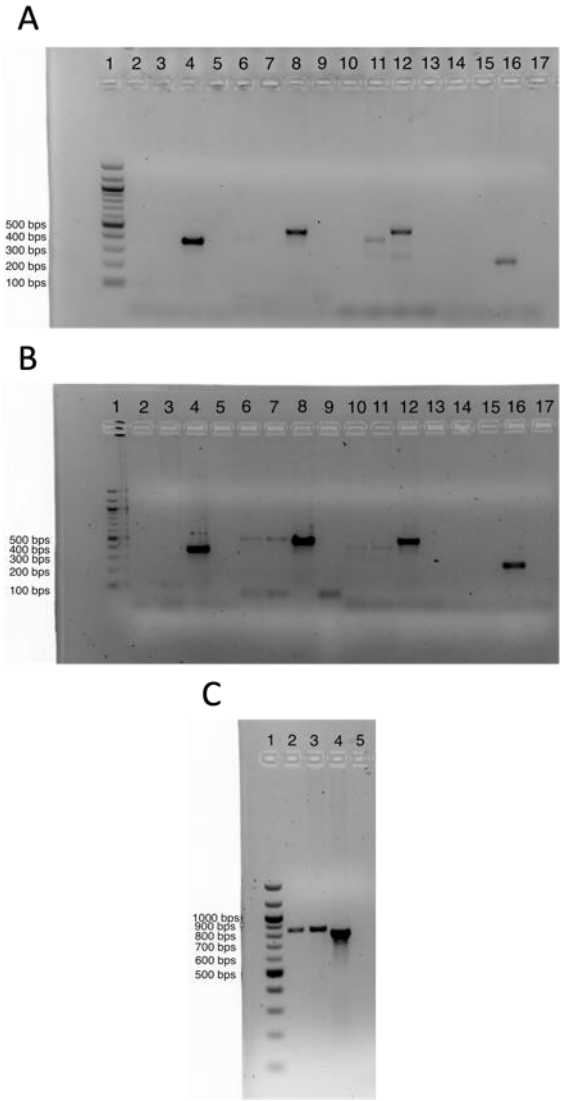


Fig. 18 Further troubleshooting *PXDN* promoter amplification difficulties in the MCF-7 cell line. PCR with all primer pairs and MCF-7 gDNA extracted by the phenol chloroform and MCF-7 and HEK-293 gDNA extracted using the kit. **A** HEK-293 gDNA was used as a positive control in the PCR reactions as previously these primers have successfully amplified HEK-293 gDNA. The contents of the wells were as follows: (1) NEB 100 bp MWM (2) P2A MCF-7 phenol chloroform (3) P2A MCF-7 extraction kit (4) P2A HEK-293 extraction kit (5) P2A no template control (6) P3 MCF-7 phenol chloroform (7) P3 MCF-7 extraction kit (8) P3 HEK-293 extraction kit (9) P3 no template control (10) P4 MCF-7 phenol chloroform (11) P4 MCF-7 extraction kit (12) P4 HEK-293 extraction kit (13) P4 no template control (14) CNTRL MCF-7 phenol chloroform (15) CNTRL MCF-7 extraction kit (16) CNTRL HEK-293 extraction kit (17) CNTRL no template control. Clear bands are only visible within the lanes where the PCR reaction was amplifying a HEK-293 gDNA template. Amplification of the MCF-7 templates, extracted by both the phenol chloroform and the kit protocol failed. **B** A repeat of the conditions and samples in A after gDNA quality and purity was checked. **C** Amplification of a region within the *TP53* promoter using MCF-7 phenol chloroform-extracted gDNA and HEK-293 and MCF-7 kit-extracted DNA. The contents of the wells were as follows (1) NEB 100 bp MWM (2) MCF-7 phenol chloroform (3) MCF-7 extraction kit (4) HEK-293 extraction kit (5) TP53 no template control.

4 Discussion

PXDN is a crosslinker of collagen IV in the ECM which has been linked to the process of EMT in cancer. According to data generated by the Human Protein Atlas, breast tissue has the highest PXDN expression when compared to other tissue types. An increase in *PXDN* expression has been observed in breast cancer, and linked to poor prognosis (Sigurdardottir et al., 2021; Zhou et al., 2022). The transcription factors Nrf2 and Snai1 have been identified as regulators of *PXDN* expression (Hanmer and Mavri-Damelin, 2018; Sitole and Mavri-Damelin, 2018), and some research has been done on histone methylation within the promoter of the gene as a down-regulator of gene expression (Zhou et al., 2022). Low levels of histone methylation within the *PXDN* promoter have been linked to poor prognosis in various cancer types (Zhou et al., 2022). However, minimal research on CpG methylation of the *PXDN* promoter has been performed. It is unclear whether changes in this form of gene regulation play a part in the increase in *PXDN* expression that is seen in various types of cancer and associated with poor prognosis. This study aimed to answer this question by examining the protein expression levels and CpG methylation levels of the *PXDN* promoter within TNBC and luminal A breast cancer cell lines.

The immunofluorescence microscopy images revealed that both the MDA-MB-231 and MCF-7 cell lines express PXDN. This particular method however, is not ideal for expression comparison between cell lines because it is challenging to include an internal control. The absence of the internal control makes it difficult to know exactly how much protein expression is indicated by a given degree of fluorescence. For example, in the western blot protocol a β-actin loading control shows the density of a band for a known amount of protein. However, with this in mind, the comparative expression data will be discussed as it has relevance to this study and suggests an avenue of research which could be worth pursuing. Furthermore, the protein expression data generated in this study mirrors mRNA expression data generated in the Human Genome Project. This supporting data indicates that the patterns observed may be somewhat reliable despite the unsuitability of this method for comparison of protein expression levels between cell lines.

The MDA-MB-231 cells showed the highest PXDN expression: approximately three and a half times greater than that observed within the MCF-7 cells. There was no significant difference between MCF-7 and HEK-293 PXDN protein expression. The Human Protein Atlas compiled PXDN messenger RNA (mRNA) expression data, revealing similar results to the PXDN protein

expression data generated in this study. The mRNA values were 59.5 nTPM, 24.5 nTPM and 2.4 nTPM for the MDA-MB-231, MCF-7 and HEK-293 cells respectively, where nTPM stands for normalised transcript per million. This RNA expression data from the Human Protein Atlas indicates that the *PXDN* expression for the MDA-MB-231 cells is 2.4 times greater than that observed within the MCF-7 cells. The HEK-293 mRNA expression is roughly tenfold lower than the MCF-7 mRNA expression, and almost 25 times lower than that of the MDA-MB-231 cells. This data strongly correlates to the protein expression data generated from the CTCF values of the immunofluorescence microscopy images. There is no corresponding protein expression data for PXDN from the Human Protein Atlas. The images of all three cell lines indicated that PXDN was predominantly localising extracellularly. This localisation is supported by studies such as those performed by Sitole and Mavri-Damelin (2018), Hanmer and Mavri-Damelin (2018) and Péterfi et al. (2009). This suggests that PXDN is localising in the extracellular matrix which was expected because of the protein's collagen IV crosslinking function (Bhave et al., 2012).

Treatment with β-Oestradiol increased PXDN expression within the ERα positive cell line MCF-7 and also within the HEK-293 cells, which have very low *ER*α expression levels according to mRNA expression data from the Human Protein Atlas. The MDA-MB-231 cells do not express the ERα and treatment had the opposite effect on these cells, decreasing PXDN expression by 50%. These changes strongly suggest the involvement of the ER pathways in the regulation of PXDN expression. Studies have shown that treatment of MCF-7 cells with β-Oestradiol results in an increase in cellular proliferation by binding and activating the transcriptional activity of the ERs (Chen et al., 2001; Katzenellenbogen et al., 1987; Vrtačnik et al., 2014). Although MDA-MB-231 TNBC cells do not express ERα, they do express ERβ and G-protein coupled oestrogen receptor 1 (GPER-1). Shanle et al. (2013) showed that ERβ bound by the β-oestradiol ligand inhibited cell growth and proliferation by regulating the expression of *CDKN1A, DKK1, WNT4* and *CDH1* which are involved in the G1/S checkpoint of the cell cycle and in the Wnt/β-catenin pathway. Schüler-Toprak et al. (2016) performed ERβ knockdown by siRNA transfection in MDA-MB-231 cells and found a three-fold increase in invasiveness related to activation of the TGFβ signalling pathway as well as an increase in other genes involved in the ECM, tumour invasion and the metabolism of vitamin D_3. These results show that ERβ, which would have been activated through β-oestradiol treatment, has an impact on genes involved in cell invasion and metastasis. Therefore ERβ activation could have a repressive impact on PXDN expression.

The decrease in PXDN expression observed in this study could be caused by activation of ERβ through treatment with β-oestradiol, however further investigation would be required to confirm this. The role of the ERs in PXDN is an intriguing one and these pathways seem to offer potential insight into the role and regulation of the gene, not only in breast cancer progression, but also in the development of an invasive phenotype such as that seen in TNBC. Further study is required to elucidate the mechanisms at play.

The MS PCR protocol was used to identify methylation in the *PXDN* promoter. We used four sets of primers to amplify across the promoter. From downstream to upstream of the TSS: Section P2 closest to the TSS at 268 bp to -95 bp; P3 from -53 bp to -524 bp; P4 from -459 bp to -935 bp; and the control primers from -794 to -1037. The control set was designed to have no MspI/HpaII restriction sites. We found the MDA-MB-231 cell line to be unmethylated, as indicated by the lack of bands in both MspI- and HpaII-restricted DNA. However, the P4 region, which lies furthest from the TSS contains DNA methylation. For a band to be absent in the HpaII-restricted DNA lane, at least one CCGG site within the region being amplified must have been restricted, and therefore must have been unmethylated. Therefore at least one of the CCGG sites were unmethylated in the P2 and P3 regions, and all of the CCGG sites must have been methylated within the P4 region in the MDA-MB-231 cell line.

Within the HEK-293 cell line, the *PXDN* promoter appears to be unmethylated within the region of the P2 amplicon. However, further upstream in the regions spanned by the P3 and P4 amplicons, the gene appears to be methylated (Figure 16). It is important to note that the agarose gel in Figure 16A (indicating the P2 region methylation status in HEK-293 cells) shows some faint bands in lanes 2 and 3. These bands are fainter than the band in lane 4, indicating that the region is most likely methylated. These faint bands are possibly due to incomplete restriction digest of the template. If digestion by HpaII were being prevented by the presence of 5mC nucleotide due to methylation of the template, it is likely that a dark band and not a faint one would be visible in lane 3. It is clear that due to the presence of a band for the P3 amplicon, all of the five CCGG recognition sites within this P3 region are methylated within the HEK-293 cell line.

For all amplicons in the HEK-293 and MDA-MB-231 cell lines, results were carried out in triplicate (except for the P2 region in the HEK-293 *PXDN* promoter, only one gel was successful for this region). Therefore, we can be confident of the methylation status of these regions. It is also important to note that the control region always amplified as expected and therefore validated the

method, indicating that absences of bands within gels were not due to the failure of PCR reactions. The P3 region was differentially methylated between the MDA-MB-231 and the HEK-293 cell lines. The nature of the MS PCR method means that the presence of a band in the HpaII-digested DNA PCR reaction indicates all CCGG sites within that region are methylated. However, if the band is absent, then anywhere from one to all the CCGG sites could be unmethylated. To get information regarding which CpG sites (if any) within the amplicon are methylated, a sequencing-based method such as nanopore sequencing, bisulfite sequencing or bisulfite pyrosequencing would need to be carried out (Liu et al., 2021). However, such sequencing technologies are hugely expensive and such an undertaking was beyond the means of this study. These differences in methylation correspond with differences in protein expression between the two cell lines as we observed that the MDA-MB-231 cell line shows many times more protein and mRNA expression than the HEK-293 cell line does. Therefore, it is possible that DNA methylation plays a part in the regulation of *PXDN*.

The advantages of the MS PCR protocol are its simplicity and cost-effectiveness. However, there are a few drawbacks to the protocol such as incomplete digestion of DNA. In an attempt to address this issue, longer digests and increased enzyme to gDNA ratios were used to avoid the appearance of bands in MspI-digested DNA, as can be seen in lane 2 of Figure 16C. The use of control gDNA samples, which were completely methylated or unmethylated, would have made the optimisation process of this protocol slightly easier by revealing false positive results. These control samples would have allowed for an indication of restriction enzyme effectiveness, streamlining the troubleshooting process. As discussed earlier, DNA which is 100% methylated at CpG sites can be purchased and a PCR reaction can be used to strip DNA of methyl groups (Wojdacz et al., 2008). Another interesting method which could be explored is 5-azacytidine treatment, which removes CpG methylation through DNMT inhibition, thereby resulting in unmethylated DNA samples which can be used as unmethylated controls in the methylation analysis protocols and protein expression protocols.

Despite multiple attempts, we were unable to amplify any region of the *PXDN* promoter consistently within the MCF-7 cell line. This occurred despite using both the same procedures and additional methods such as alternative extraction protocols that would have served to remove known restriction digest and PCR inhibitors such as phenol (Schrader et al., 2012). The addition of dimethylsulfoxide (DMSO) to the PCR reaction to assist with primer denaturing was also tried,

however this alteration did not result in successful amplification of the region. Therefore, within the MCF-7 cell line it seems there is another mechanism involved. Veal et al. (2012) conducted an investigation into why some regions of the genome prove particularly troublesome to amplify. They investigated regions of the genome which gave weak signals within Illumina Infinium single nucleotide polymorphism (SNP) array datasets by comparing them with control regions which in turn produced normal intensity signals within the same datasets. First, they produced single copy DNA probes for the corresponding "weak" and "normal" Illumina signal regions. A Southern blot was then performed with these samples after various DNA denaturing conditions were used (no denaturing; heating to 100 ℃ for one minute; heating to 100 ℃ for one minute followed by snap freezing) and followed by restriction enzyme digestion. Analysis of Southern blot results revealed that the "normal" Illumina signal regions were successfully denatured, however the "weak" signal regions were not. Clear bands are visible within the "weak" signal regions which show the resistance of these regions to denaturing. Relative efficiencies of amplification were then assessed between the "normal" and "weak" Illumina signal regions, PCR efficiencies were reduced by 10% to 50% in the "weak" Illumina signal regions samples. Different methods were tested to enhance denaturing conditions such as adding DMSO; increasing the initial denaturation temperature to 98 ℃; and heating samples to 130 ℃ for one minute. The latter of these methods proved the most effective. These experiments show that for these regions of the genome which are difficult to amplify the most likely explanation is that they resist the initial denaturing step of the reaction.

Further, Veal et al. (2012) proposed a mechanism whereby relatively small regions which have very high GC content do not denature when exposed to the regular denaturing temperatures of standard PCR. These regions then keep the complementary strands together, even though adjacent regions may be denatured, allowing for the template to quickly renature once the temperature is lowered again. Veal et al. refer to these regions as thermodynamically ultra-fastened (TUF). They tested this hypothesis through a series of digestions with different restriction enzymes. Amplification was improved when restriction enzyme digestion separated the amplicon from the region which was high in GC content.

The *PXDN* promoter has a high GC%, particularly within the regions covered by amplicon 3 and amplicon 2 (Figure 7). In fact, one of the regions under observation within the study conducted by Veal et al. included the region within chromosome 2 near where *PXDN* is located. The trouble with the TUF hypothesis is that the gene was successfully amplified with no trouble within the

MDA-MB-231 and HEK-293 cell lines. There may, however, be mutations within the gene which change a few bases in the *PXDN* promoter in the MCF-7 cell line, resulting in the observed differences between this cell line and the others. Therefore, this phenomenon is something to keep in mind, it may be worth attempting to heat the DNA sample to 130 ℃ and snap freeze it on ice before adding the template to the PCR reaction as suggested by the study performed by Veal et al. (2012).

Another possible explanation for the difficulty experienced in the attempt to amplify this region of the MCF-7 genome is that there is a chromosomal aberration within this region. Hampton et al. (2009) identified 157 somatic breakpoints within the genome of the MCF-7 cell line, one of the breakpoints they identified occurred within sulfatase 2 (*SULF2*). Knockdown of *SULF2* was linked to cell proliferation and survival as well as increased levels of anchorage independent growth. Jönsson et al. (2007) compared DNA copy number alterations (CNA) within the genomic profiles of 10 different breast cancer cell lines including the MDA-MB-231 and MCF-7 cell lines. The MCF-7 genome showed CNAs which caused changes to the structure and sequence of large regions of chromosomes. Some of these caused duplications of sequences, and others extensive deletions of gene-containing regions. These aberrations were then linked to changes in gene expression, for example within the 20q13 region CNAs were directly linked to increased expression of *SULF2* and oncogene *ZNF217*. The combined changes in expression within the MCF-7 cell line caused by CNAs correlated with the luminal A subtype. There were more deleterious effects of the CNAs within the MDA-MB-231 cell line than within the MCF-7 cell line, for example homozygous deletions within the 9p21.3 region affected the tumour suppressor genes *CDKN2A* and *CDKN2B*. The studies conducted by Jönsson et al. (2007) and Hampton et al. (2009) highlight the many changes in gene expression caused by chromosomal abnormalities observed within these cell lines that we use as breast cancer models. However, neither study, though both extensive, found any changes within chromosome 2. It is possible that with repeated passaging of cells an accumulation of mutations and alterations could result in the deletion of the *PXDN* promoter region, however the passage number was low for the MCF-7 cell line (less than 25) and the cell line had been recently acquired from the ATCC. Therefore, it is unlikely that such a significant divergence from the documented genotype would have occurred within these few passage events.

A study conducted by Liu et al. (1997) found that a T → C point mutation exerted an inhibitory effect on PCR amplification when present up to 84 bp downstream of a forward primer. Varying Taq DNA polymerases; reaction conditions such as annealing temperatures and extension time; and concentrations of reaction components did not alleviate the observed inhibition. No mechanism was proposed as responsible for this phenomenon. This study was conducted in 1997, which is a long time ago and no more recent studies which mirror the findings observed could be found. This point mutation is from T to C, and therefore increases the CG content. It is possible, that within the correct conditions this change could create a TUF as proposed by Veal et al (2012). There is a region within amplicon 2, a few base pairs before the TSS of the *PXDN* promoter, where the GC content exceeds 85% for about 65 bp. It is possible that a point mutation here could trigger a TUF region, resulting in rapid re-annealing of the adjacent promoter regions once the initial denaturing step at the beginning of the PCR reaction has been completed, and the subsequent failure of amplification. If this point mutation were present within the MCF-7 cell line and not within the other cell lines used as models within this study, then this could be a possible explanation as to why amplification was successful within the MDA-MB-231 and HEK-293 cell lines but not the MCF-7 cell line.

Unfortunately, the DNA methylation status of the *PXDN* promoter remains unknown for the non-invasive, luminal A subtype MCF-7 breast cancer cell line. The MS PCR method yielded reliable results for the MDA-MB-231 and HEK293 cell lines. However, there are limitations to this approach for studying DNA methylation. There are many CpG sites within the amplicons used to analyse the *PXDN* promoter which fall outside of CCGG restriction enzyme recognition sites (see Figure 8). From the results it is clear that at the 3' side of the *PXDN* promoter, the region covered by the P2 amplicon is methylated in both the HEK-293 and MDA-MB-231 cell lines. The region covered by P3 amplicon appears unmethylated. This change occurs at different points within these two cell lines: it appears that the HEK-293 cell line is methylated further downstream than the MDA-MB-231 cell line. The large difference in PXDN expression levels between the two cell lines could be a result of this region within the promoter of differential 5mC levels. This method is limited because it does not give information on 5mC levels at a CpG-level resolution and the results are qualitative rather than quantitative. This method does, however highlight the fact that it is likely that DNA methylation of the *PXDN* promoter is involved in gene regulation, and suggests that higher resolution sequencing methods such as bisulfite pyrosequencing would be worth considering.

5 Conclusion and Future Prospects

The PXDN protein was shown to be most highly expressed within the TNBC cell line MDA-MB-231, three and a half times the expression seen within the luminal A MCF-7 cell line. This correlation between invasive potential and PXDN expression in breast cancer is an avenue worth exploring. Invasion and migration assays would be good methods to test whether PXDN is directly involved in increasing the invasive potential of breast cancer cells. β-Oestradiol treatment increased ERα-positive MCF-7 PXDN expression, but decreased ERα-negative MDA-MB-231 PXDN expression. Therefore, it is highly likely that the ERα and ERβ are regulators of *PXDN* expression. Chromatin immunoprecipitation would need to be performed to investigate the nature of the interaction between ERs and the *PXDN* promoter. A differential region of methylation between the MDA-MB-231 and HEK-293 cell lines reveals a region between 524 bp and 53 bp upstream of the *PXDN* TSS which is methylated in the HEK-293 cell line and unmethylated in the MDA-MB-231 cell line, these differences correlate with the expression data gathered for these respective cell lines.

To truly understand the role of DNA methylation within this region, a DNA methylation investigative method which yields quantitative results at a CpG-level resolution such as sequencing-based techniques like bisulfite-based pyrosequencing or nanopore sequencing technologies would be useful. Methods such as HRM, which are quantitative could reveal if methylation signatures are consistent throughout the cell population.

Treating the MCF-7 cell line with β-oestradiol increases *PXDN* expression, therefore it would be interesting to see if this has an impact on the methylation levels of the gene promoter. If expression increases with β-oestradiol treatment then it may suggest that the DNA is in a closed, inaccessible conformation due to the presence of histone modifications and DNA methylation, which upon treatment are reversed to allow for an open conformation of DNA amenable to transcription, and investigation by MS PCR.

It is clear that *PXDN* is involved in breast cancer, however the mechanisms whereby the expression of this gene are regulated remain somewhat unclear. This study has established that DNA methylation is involved in regulating PXDN expression in breast cancer cell lines, and that further study into this is a worthy pursuit. Of particular interest to future studies is the relationship between

PXDN and ERs and the nature of their interactions between ERα-positive and -negative breast cancer molecular subtypes.

6 References

Ansar, M., Thu, L.T.A., Hung, C.-S., Su, C.-M., Huang, M.-H., Liao, L.-M., Chung, Y.-M., Lin, R.-K., 2022. Promoter hypomethylation and overexpression of TSTD1 mediate poor treatment response in breast cancer. Front. Oncol. 12.

Bae, J.M., Shin, S.-H., Kwon, H.-J., Park, S.-Y., Kook, M.C., Kim, Y.-W., Cho, N.-Y., Kim, N., Kim, T.-Y., Kim, D., Kang, G.H., 2012. ALU and LINE-1 hypomethylations in multistep gastric carcinogenesis and their prognostic implications. Int. J. Cancer 131, 1323–1331. https://doi.org/10.1002/ijc.27369

Baylin, S.B., Jones, P.A., 2016. Epigenetic Determinants of Cancer. Cold Spring Harb. Perspect. Biol. 8, a019505. https://doi.org/10.1101/cshperspect.a019505

Bettaieb, A., Paul, C., Plenchette, S., Shan, J., Chouchane, L., Ghiringhelli, F., 2017. Precision medicine in breast cancer: reality or utopia? J. Transl. Med. 15, 139. https://doi.org/10.1186/s12967-017-1239-z

Bhave, G., Cummings, C.F., Vanacore, R.M., Kumagai-Cresse, C., Ero-Tolliver, I.A., Rafi, M., Kang, J.-S., Pedchenko, V., Fessler, L.I., Fessler, J.H., Hudson, B.G., 2012. Peroxidasin forms sulfilimine chemical bonds using hypohalous acids in tissue genesis. Nat. Chem. Biol. 8, 784–790. https://doi.org/10.1038/nchembio.1038

Bird, A., Taggart, M., Frommer, M., Miller, O.J., Macleod, D., 1985. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell 40, 91–99. https://doi.org/10.1016/0092-8674(85)90312-5

Blows, F.M., Driver, K.E., Schmidt, M.K., Broeks, A., Leeuwen, F.E. van, Wesseling, J., Cheang, M.C., Gelmon, K., Nielsen, T.O., Blomqvist, C., Heikkilä, P., Heikkinen, T., Nevanlinna, H., Akslen, L.A., Bégin, L.R., Foulkes, W.D., Couch, F.J., Wang, X., Cafourek, V., Olson, J.E., Baglietto, L., Giles, G.G., Severi, G., McLean, C.A., Southey, M.C., Rakha, E., Green, A.R., Ellis, I.O., Sherman, M.E., Lissowska, J., Anderson, W.F., Cox, A., Cross, S.S., Reed, M.W.R., Provenzano, E., Dawson, S.-J., Dunning, A.M., Humphreys, M., Easton, D.F., García-Closas, M., Caldas, C., Pharoah, P.D., Huntsman, D., 2010. Subtyping of Breast Cancer by Immunohistochemistry to Investigate a Relationship between Subtype and Short and Long Term Survival: A Collaborative Analysis of Data for 10,159 Cases from 12 Studies. PLOS Med. 7, e1000279. https://doi.org/10.1371/journal.pmed.1000279

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

Brennan, K., Garcia-Closas, M., Orr, N., Fletcher, O., Jones, M., Ashworth, A., Swerdlow, A., Thorne, H., KConFab Investigators, Riboli, E., Vineis, P., Dorronsoro, M., Clavel-Chapelon, F., Panico, S., Onland-Moret, N.C., Trichopoulos, D., Kaaks, R., Khaw, K.-T., Brown, R., Flanagan, J.M., 2012. Intragenic ATM methylation in peripheral blood DNA

as a biomarker of breast cancer risk. Cancer Res. 72, 2304–2313. https://doi.org/10.1158/0008-5472.CAN-11-3157

Bukowska, B., Sicińska, P., 2021. Influence of Benzo(a)pyrene on Different Epigenetic Processes. Int. J. Mol. Sci. 22, 13453. https://doi.org/10.3390/ijms222413453

Cai, X., Yang, Xiwen, Jin, C., Li, L., Cui, Q., Guo, Y., Dong, Y., Yang, Xiaohua, Guo, L., Zhang, M., 2018. Identification and verification of differentially expressed microRNAs and their target genes for the diagnosis of esophageal cancer. Oncol. Lett. 16, 3642–3650. https://doi.org/10.3892/ol.2018.9066

Cameron, D., Casey, M., Press, M., Lindquist, D., Pienkowski, T., Romieu, C.G., Chan, S., Jagiello-Gruszfeld, A., Kaufman, B., Crown, J., Chan, A., Campone, M., Viens, P., Davidson, N., Gorbounova, V., Raats, J.I., Skarlos, D., Newstat, B., Roychowdhury, D., Paoletti, P., Oliva, C., Rubin, S., Stein, S., Geyer, C.E., 2008. A phase III randomized comparison of lapatinib plus capecitabine versus capecitabine alone in women with advanced breast cancer that has progressed on trastuzumab: updated efficacy and biomarker analyses. Breast Cancer Res. Treat. 112, 533–543. https://doi.org/10.1007/s10549-007-9885-0

Cano, A., Pérez-Moreno, M.A., Rodrigo, I., Locascio, A., Blanco, M.J., del Barrio, M.G., Portillo, F., Nieto, M.A., 2000. The transcription factor Snail controls epithelial–mesenchymal transitions by repressing E-cadherin expression. Nat. Cell Biol. 2, 76–83. https://doi.org/10.1038/35000025

Cedar, H., Solage, A., Glaser, G., Razin, A., 1979. Direct detection of methylated cytosine in DNA by use of the restriction enzyme Mspl. Nucleic Acids Res. 6, 2125–2132. https://doi.org/10.1093/nar/6.6.2125

Chan, J.R., Hyduk, S.J., Cybulsky, M.I., 2000. α4β1 Integrin/VCAM-1 Interaction Activates αLβ2 Integrin-Mediated Adhesion to ICAM-1 in Human T Cells. J. Immunol. 164, 746–753. https://doi.org/10.4049/jimmunol.164.2.746

Chang, T.T., Thakar, D., Weaver, V.M., 2017. Force-dependent breaching of the basement membrane. Matrix Biol., Basement Membranes in Health and Disease 57–58, 178–189. https://doi.org/10.1016/j.matbio.2016.12.005

Chavarri-Guerra, Y., Blazer, K.R., Weitzel, J.N., 2017. Genetic Cancer Risk Assessment for Breast Cancer in Latin America. Rev. Investig. Clin. Organo Hosp. Enfermedades Nutr. 69, 94–102.

Chen, I., Hsieh, T., Thomas, T., Safe, S., 2001. Identification of estrogen-induced genes downregulated by AhR agonists in MCF-7 breast cancer cells using suppression subtractive hybridization. Gene 262, 207–214. https://doi.org/10.1016/S0378-1119(00)00530-8

Cheng, G., Salerno, J.C., Cao, Z., Pagano, P.J., Lambeth, J.D., 2008. Identification and characterization of VPO1, a new animal heme-containing peroxidase. Free Radic. Biol. Med. 45, 1682–1694. https://doi.org/10.1016/j.freeradbiomed.2008.09.009

Clark, S.J., Harrison, J., Paul, C.L., Frommer, M., 1994. High sensitivity mapping of methylated cytosines. Nucleic Acids Res. 22, 2990–2997.

Connolly, R.M., Zhao, F., Miller, K.D., Lee, M.-J., Piekarz, R.L., Smith, K.L., Brown-Glaberman, U.A., Winn, J.S., Faller, B.A., Onitilo, A.A., Burkard, M.E., Budd, G.T., Levine, E.G., Royce, M.E., Kaufman, P.A., Thomas, A., Trepel, J.B., Wolff, A.C., Sparano, J.A., 2021. E2112: Randomized Phase III Trial of Endocrine Therapy Plus Entinostat or Placebo in Hormone Receptor–Positive Advanced Breast Cancer. A Trial of the ECOG-ACRIN Cancer Research Group. J. Clin. Oncol. 39, 3171–3181. https://doi.org/10.1200/JCO.21.00944

Cruz-Bermúdez, A., Laza-Briviesca, R., Vicente-Blanco, R.J., García-Grande, A., Coronado, M.J., Laine-Menéndez, S., Alfaro, C., Sanchez, J.C., Franco, F., Calvo, V., Romero, A., Martin-Acosta, P., Salas, C., Garcia, J.M., Provencio, M., 2019. Cancer-associated fibroblasts modify lung cancer metabolism involving ROS and TGF-β signaling. Free Radic. Biol. Med. 130, 163–173. https://doi.org/10.1016/j.freeradbiomed.2018.10.450

de Almeida, B.P., Apolónio, J.D., Binnie, A., Castelo-Branco, P., 2019. Roadmap of DNA methylation in breast cancer identifies novel prognostic biomarkers. BMC Cancer 19, 219. https://doi.org/10.1186/s12885-019-5403-0

Deryugina, E.I., Quigley, J.P., 2006. Matrix metalloproteinases and tumor metastasis. Cancer Metastasis Rev. 25, 9–34. https://doi.org/10.1007/s10555-006-7886-9

Di, Y., Chen, D., Yu, W., Yan, L., 2019. Bladder cancer stage-associated hub genes revealed by WGCNA co-expression network analysis. Hereditas 156. https://doi.org/10.1186/s41065-019-0083-y

Dougan, J., Hawsawi, O., Burton, L.J., Edwards, G., Jones, K., Zou, J., Nagappan, P., Wang, G., Zhang, Q., Danaher, A., Bowen, N., Hinton, C., Odero-Marah, V.A., 2019. Proteomics-Metabolomics Combined Approach Identifies Peroxidasin as a Protector against Metabolic and Oxidative Stress in Prostate Cancer. Int. J. Mol. Sci. 20. https://doi.org/10.3390/ijms20123046

Eble, J.A., Niland, S., 2019. The extracellular matrix in tumor progression and metastasis. Clin. Exp. Metastasis 36, 171–198. https://doi.org/10.1007/s10585-019-09966-1

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., Sonnhammer, E.L.L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S.C.E., Finn, R.D., 2019. The Pfam protein families database in 2019. Nucleic Acids Res. 47, D427–D432. https://doi.org/10.1093/nar/gky995

Engmann, N.J., Golmakani, M.K., Miglioretti, D.L., Sprague, B.L., Kerlikowske, K., 2017. Population-Attributable Risk Proportion of Clinical Risk Factors for Breast Cancer. JAMA Oncol. 3, 1228–1236. https://doi.org/10.1001/jamaoncol.2016.6326

Esteller, M., Silva, J.M., Dominguez, G., Bonilla, F., Matias-Guiu, X., Lerma, E., Bussaglia, E., Prat, J., Harkes, I.C., Repasky, E.A., Gabrielson, E., Schutte, M., Baylin, S.B., Herman, J.G., 2000. Promoter Hypermethylation and BRCA1 Inactivation in Sporadic Breast and

Ovarian Tumors. JNCI J. Natl. Cancer Inst. 92, 564–569. https://doi.org/10.1093/jnci/92.7.564

Flanagan, J.M., Munoz-Alegre, M., Henderson, S., Tang, T., Sun, P., Johnson, N., Fletcher, O., Dos Santos Silva, I., Peto, J., Boshoff, C., Narod, S., Petronis, A., 2009. Gene-body hypermethylation of ATM in peripheral blood DNA of bilateral breast cancer patients. Hum. Mol. Genet. 18, 1332–1342. https://doi.org/10.1093/hmg/ddp033

Fragomeni, S.M., Sciallis, A., Jeruss, J.S., 2018. Molecular subtypes and local-regional control of breast cancer. Surg. Oncol. Clin. N. Am. 27, 95–120. https://doi.org/10.1016/j.soc.2017.08.005

Frantz, C., Stewart, K.M., Weaver, V.M., 2010. The extracellular matrix at a glance. J. Cell Sci. 123, 4195–4200. https://doi.org/10.1242/jcs.023820

Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L., Paul, C.L., 1992. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc. Natl. Acad. Sci. U. S. A. 89, 1827–1831. https://doi.org/10.1073/pnas.89.5.1827

Gardiner-Garden, M., Frommer, M., 1987. CpG Islands in vertebrate genomes. J. Mol. Biol. 196, 261–282. https://doi.org/10.1016/0022-2836(87)90689-9

Ghaheri, M., Kahrizi, D., Yari, K., Babaie, A., Suthar, R.S., Kazemi, E., 2016. A comparative evaluation of four DNA extraction protocols from whole blood sample. Cell. Mol. Biol. Noisy--Gd. Fr. 62, 120–124.

Giepmans, B.N.G., Adams, S.R., Ellisman, M.H., Tsien, R.Y., 2006. The fluorescent toolbox for assessing protein location and function. Science 312, 217–224. https://doi.org/10.1126/science.1124618

González-Ramos, M., Mora, I., de Frutos, S., Garesse, R., Rodríguez-Puyol, M., Olmos, G., Rodríguez-Puyol, D., 2012. Intracellular redox equilibrium is essential for the constitutive expression of AP-1 dependent genes in resting cells: Studies on TGF-β1 regulation. Int. J. Biochem. Cell Biol. 44, 963–971. https://doi.org/10.1016/j.biocel.2012.03.003

Gordon, M.K., Hahn, R.A., 2010. Collagens. Cell Tissue Res. 339, 247–257. https://doi.org/10.1007/s00441-009-0844-4

Gotenstein, J.R., Swale, R.E., Fukuda, T., Wu, Z., Giurumescu, C.A., Goncharov, A., Jin, Y., Chisholm, A.D., 2010. The C. elegans peroxidasin PXN-2 is essential for embryonic morphogenesis and inhibits adult axon regeneration. Development 137, 3603–3613. https://doi.org/10.1242/dev.049189

Greenberg, M.V.C., Bourc'his, D., 2019. The diverse roles of DNA methylation in mammalian development and disease. Nat. Rev. Mol. Cell Biol. 20, 590–607. https://doi.org/10.1038/s41580-019-0159-6

Greenblatt, M.S., Bennett, W.P., Hollstein, M., Harris, C.C., 1994. Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis. Cancer Res. 54, 4855–4878.

Gupta, S., Jaworska-Bieniek, K., Narod, S.A., Lubinski, J., Wojdacz, T.K., Jakubowska, A., 2014. Methylation of the BRCA1 promoter in peripheral blood DNA is associated with triple-negative and medullary breast cancer. Breast Cancer Res. Treat. 148, 615–622. https://doi.org/10.1007/s10549-014-3179-0

Hampton, O.A., Den Hollander, P., Miller, C.A., Delgado, D.A., Li, J., Coarfa, C., Harris, R.A., Richards, S., Scherer, S.E., Muzny, D.M., Gibbs, R.A., Lee, A.V., Milosavljevic, A., 2009. A sequence-level map of chromosomal breakpoints in the MCF-7 breast cancer cell line yields insights into the evolution of a cancer genome. Genome Res. 19, 167–177. https://doi.org/10.1101/gr.080259.108

Hanmer, K.L., Mavri-Damelin, D., 2018. Peroxidasin is a novel target of the redox-sensitive transcription factor Nrf2. Gene 674, 104–114. https://doi.org/10.1016/j.gene.2018.06.076

Harris, E.E.R., 2018. Precision Medicine for Breast Cancer: The Paths to Truly Individualized Diagnosis and Treatment [WWW Document]. Int. J. Breast Cancer. https://doi.org/10.1155/2018/4809183

Hay, E.D., 1995. An Overview of Epithelio-Mesenchymal Transformation. Cells Tissues Organs 154, 8–20. https://doi.org/10.1159/000147748

Herman, J.G., Graff, J.R., Myöhänen, S., Nelkin, B.D., Baylin, S.B., 1996. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. Proc. Natl. Acad. Sci. U. S. A. 93, 9821–9826.

Herman, J.G., Merlo, A., Mao, L., Lapidus, R.G., Issa, J.P., Davidson, N.E., Sidransky, D., Baylin, S.B., 1995. Inactivation of the CDKN2/p16/MTS1 gene is frequently associated with aberrant DNA methylation in all common human cancers. Cancer Res. 55, 4525–4530.

Hernández, H.G., Tse, M.Y., Pang, S.C., Arboleda, H., Forero, D.A., 2013. Optimizing methodologies for PCR-based DNA methylation analysis. BioTechniques 55, 181–197. https://doi.org/10.2144/000114087

Holliday, R., Grigg, G.W., 1993. DNA methylation and mutation. Mutat. Res. Mol. Mech. Mutagen., Special Issue In Memory of Max Clark, a Pioneer in Fundamental Mutation Reserach 285, 61–67. https://doi.org/10.1016/0027-5107(93)90052-H

Hunt, S.E., McLaren, W., Gil, L., Thormann, A., Schuilenburg, H., Sheppard, D., Parton, A., Armean, I.M., Trevanion, S.J., Flicek, P., Cunningham, F., 2018. Ensembl variation resources. Database 2018. https://doi.org/10.1093/database/bay119

Isakoff, S.J., 2010. Triple Negative Breast Cancer: Role of Specific Chemotherapy Agents. Cancer J. Sudbury Mass 16, 53–61. https://doi.org/10.1097/PPO.0b013e3181d24ff7

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., Zhang, Y., 2011. Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. Science 333, 1300–1303. https://doi.org/10.1126/science.1210597

Iwamoto, T., Yamamoto, N., Taguchi, T., Tamaki, Y., Noguchi, S., 2011. BRCA1 promoter methylation in peripheral blood cells is associated with increased risk of breast cancer with BRCA1 promoter methylation. Breast Cancer Res. Treat. 129, 69–77. https://doi.org/10.1007/s10549-010-1188-1

Jones, P.A., 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. 13, 484–492. https://doi.org/10.1038/nrg3230

Jönsson, G., Staaf, J., Olsson, E., Heidenblad, M., Vallon-Christersson, J., Osoegawa, K., de Jong, P., Oredsson, S., Ringnér, M., Höglund, M., Borg, Å., 2007. High-resolution genomic profiles of breast cancer cell lines assessed by tiling BAC array comparative genomic hybridization. Genes. Chromosomes Cancer 46, 543–558. https://doi.org/10.1002/gcc.20438

Kalluri, R., 2016. The biology and function of fibroblasts in cancer. Nat. Rev. Cancer 16, 582–598. https://doi.org/10.1038/nrc.2016.73

Karagiannis, G.S., Poutahidis, T., Erdman, S.E., Kirsch, R., Riddell, R.H., Diamandis, E.P., 2012. Cancer-Associated Fibroblasts Drive the Progression of Metastasis through both Paracrine and Mechanical Pressure on Cancer Tissue. Mol. Cancer Res. 10, 1403–1418. https://doi.org/10.1158/1541-7786.MCR-12-0307

Karaulanov, E.E., Böttcher, R.T., Niehrs, C., 2006. A role for fibronectin-leucine-rich transmembrane cell-surface proteins in homotypic cell adhesion. EMBO Rep. 7, 283–290. https://doi.org/10.1038/sj.embor.7400614

Katzenellenbogen, B.S., Kendra, K.L., Norman, M.J., Berthois, Y., 1987. Proliferation, hormonal responsiveness, and estrogen receptor content of MCF-7 human breast cancer cells grown in the short-term and long-term absence of estrogens. Cancer Res. 47, 4355–4360.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., Haussler, and D., 2002. The Human Genome Browser at UCSC. Genome Res. 12, 996–1006. https://doi.org/10.1101/gr.229102

Kim, Y.-J., Lee, J., Han, K., 2012. Transposable Elements: No More "Junk DNA." Genomics Inform. 10, 226–233. https://doi.org/10.5808/GI.2012.10.4.226

Klug, W.S., Cummings, M.R., Spencer, C.A., Palladino, M.A., 2015. Concepts of Genetics, 11th ed. Pearson, Edinburgh.

Koboldt, D.C., Fulton, R.S., McLellan, M.D., Schmidt, H., Kalicki-Veizer, J., McMichael, J.F., Fulton, L.L., Dooling, D.J., Ding, L., Mardis, E.R., Wilson, R.K., Ally, A., Balasundaram, M., Butterfield, Y.S.N., Carlsen, R., Carter, C., Chu, A., Chuah, E., Chun, H.-J.E., Coope, R.J.N., Dhalla, N., Guin, R., Hirst, C., Hirst, M., Holt, R.A., Lee, D., Li, H.I., Mayo, M., Moore, R.A., Mungall, A.J., Pleasance, E., Gordon Robertson, A., Schein, J.E., Shafiei, A., Sipahimalani, P., Slobodan, J.R., Stoll, D., Tam, A., Thiessen, N., Varhol, R.J., Wye,

N., Zeng, T., Zhao, Y., Birol, I., Jones, S.J.M., Marra, M.A., Cherniack, A.D., Saksena, G., Onofrio, R.C., Pho, N.H., Carter, S.L., Schumacher, S.E., Tabak, B., Hernandez, B., Gentry, J., Nguyen, H., Crenshaw, A., Ardlie, K., Beroukhim, R., Winckler, W., Getz, G., Gabriel, S.B., Meyerson, M., Chin, L., Park, P.J., Kucherlapati, R., Hoadley, K.A., Todd Auman, J., Fan, C., Turman, Y.J., Shi, Y., Li, L., Topal, M.D., He, X., Chao, H.-H., Prat, A., Silva, G.O., Iglesia, M.D., Zhao, W., Usary, J., Berg, J.S., Adams, M., Booker, J., Wu, J., Gulabani, A., Bodenheimer, T., Hoyle, A.P., Simons, J.V., Soloway, M.G., Mose, L.E., Jefferys, S.R., Balu, S., Parker, J.S., Neil Hayes, D., Perou, C.M., Malik, S., Mahurkar, S., Shen, H., Weisenberger, D.J., Triche Jr, T., Lai, P.H., Bootwalla, M.S., Maglinte, D.T., Berman, B.P., Van Den Berg, D.J., Baylin, S.B., Laird, P.W., Creighton, C.J., Donehower, L.A., Getz, G., Noble, M., Voet, D., Saksena, G., Gehlenborg, N., DiCara, D., Zhang, J., Zhang, H., Wu, C.-J., Yingchun Liu, S., Lawrence, M.S., Zou, L., Sivachenko, A., Lin, P., Stojanov, P., Jing, R., Cho, J., Sinha, Raktim, Park, R.W., Nazaire, M.-D., Robinson, J., Thorvaldsdottir, H., Mesirov, J., Park, P.J., Chin, L., Reynolds, S., Kreisberg, R.B., Bernard, B., Bressler, R., Erkkila, T., Lin, J., Thorsson, V., Zhang, W., Shmulevich, I., Ciriello, G., Weinhold, N., Schultz, N., Gao, J., Cerami, E., Gross, B., Jacobsen, A., Sinha, Rileen, Arman Aksoy, B., Antipin, Y., Reva, B., Shen, R., Taylor, B.S., Ladanyi, M., Sander, C., Anur, P., Spellman, P.T., Lu, Y., Liu, W., Verhaak, R.R.G., Mills, G.B., Akbani, R., Zhang, N., Broom, B.M., Casasent, T.D., Wakefield, C., Unruh, A.K., Baggerly, K., Coombes, K., Weinstein, J.N., Haussler, D., Benz, C.C., Stuart, J.M., Benz, S.C., Zhu, J., Szeto, C.C., Scott, G.K., Yau, C., Paull, E.O., Carlin, D., Wong, C., Sokolov, A., Thusberg, J., Mooney, S., Ng, S., Goldstein, T.C., Ellrott, K., Grifford, M., Wilks, C., Ma, S., Craft, B., Yan, C., Hu, Y., Meerzaman, D., Gastier-Foster, J.M., Bowen, J., Ramirez, N.C., Black, A.D., Pyatt, R.E., White, P., Zmuda, E.J., Frick, J., Lichtenberg, T.M., Brookens, R., George, M.M., Gerken, M.A., Harper, H.A., Leraas, K.M., Wise, L.J., Tabler, T.R., McAllister, C., Barr, T., Hart-Kothari, M., Tarvin, K., Saller, C., Sandusky, G., Mitchell, C., Iacocca, M.V., Brown, J., Rabeno, B., Czerwinski, C., Petrelli, N., Dolzhansky, O., Abramov, M., Voronina, O., Potapova, O., Marks, J.R., Suchorska, W.M., Murawa, D., Kycler, W., Ibbs, M., Korski, K., Spychała, A., Murawa, P., Brzeziński, J.J., Perz, H., Łaźniak, R., Teresiak, M., Tatka, H., Leporowska, E., Bogusz-Czerniewicz, M., Malicki, J., Mackiewicz, A., Wiznerowicz, M., Van Le, X., Kohl, B., Viet Tien, N., Thorp, R., Van Bang, N., Sussman, H., Duc Phu, B., Hajek, R., Phi Hung, N., Viet The Phuong, T., Quyet Thang, H., Zaki Khan, K., Penny, R., Mallery, D., Curley, E., Shelton, C., Yena, P., Ingle, J.N., Couch, F.J., Lingle, W.L., King, T.A., Maria Gonzalez-Angulo, A., Mills, G.B., Dyer, M.D., Liu, S., Meng, X., Patangan, M., The Cancer Genome Atlas Network, Genome sequencing centres: Washington University in St Louis, Genome characterization centres: BC Cancer Agency, Broad Institute, Brigham & Women's Hospital & Harvard Medical School, University of North Carolina, C.H., University of Southern California/Johns Hopkins, Genome data analysis: Baylor College of Medicine, Institute for Systems Biology, Memorial Sloan-Kettering Cancer Center, Oregon Health & Science University, The University of Texas MD Anderson Cancer Center, University of California, S.C.I., NCI, Biospecimen core resource: Nationwide Children's Hospital Biospecimen Core Resource, Tissue source sites: ABS-IUPUI, Christiana, Cureline, Duke University Medical Center, The Greater Poland Cancer Centre, ILSBio, International Genomics Consortium, Mayo Clinic, MSKCC, MD Anderson Cancer Center, 2012. Comprehensive molecular portraits of human breast tumours. Nature 490, 61–70. https://doi.org/10.1038/nature11412

Koch, A., Joosten, S.C., Feng, Z., de Ruijter, T.C., Draht, M.X., Melotte, V., Smits, K.M., Veeck, J., Herman, J.G., Van Neste, L., Van Criekinge, W., De Meyer, T., van Engeland, M., 2018. Analysis of DNA methylation in cancer: location revisited. Nat. Rev. Clin. Oncol. 15, 459–466. https://doi.org/10.1038/s41571-018-0004-4

Kulis, M., Heath, S., Bibikova, M., Queirós, A.C., Navarro, A., Clot, G., Martínez-Trillos, A., Castellano, G., Brun-Heath, I., Pinyol, M., Barberán-Soler, S., Papasaikas, P., Jares, P., Beà, S., Rico, D., Ecker, S., Rubio, M., Royo, R., Ho, V., Klotzle, B., Hernández, L., Conde, L., López-Guerra, M., Colomer, D., Villamor, N., Aymerich, M., Rozman, M., Bayes, M., Gut, M., Gelpí, J.L., Orozco, M., Fan, J.-B., Quesada, V., Puente, X.S., Pisano, D.G., Valencia, A., López-Guillermo, A., Gut, I., López-Otín, C., Campo, E., Martín-Subero, J.I., 2012. Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. Nat. Genet. 44, 1236–1242. https://doi.org/10.1038/ng.2443

Kurdyukov, S., Bullock, M., 2016. DNA Methylation Analysis: Choosing the Right Method. Biology 5, 3. https://doi.org/10.3390/biology5010003

Lampi, M.C., Reinhart-King, C.A., 2018. Targeting extracellular matrix stiffness to attenuate disease: From molecular mechanisms to clinical trials. Sci. Transl. Med. 10. https://doi.org/10.1126/scitranslmed.aao0475

Lehmann, B.D., Bauer, J.A., Chen, X., Sanders, M.E., Chakravarthy, A.B., Shyr, Y., Pietenpol, J.A., 2011. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J. Clin. Invest. 121, 2750–2767. https://doi.org/10.1172/JCI45014

Lehmann, B.D., Jovanović, B., Chen, X., Estrada, M.V., Johnson, K.N., Shyr, Y., Moses, H.L., Sanders, M.E., Pietenpol, J.A., 2016. Refinement of Triple-Negative Breast Cancer Molecular Subtypes: Implications for Neoadjuvant Chemotherapy Selection. PLOS ONE 11, e0157368. https://doi.org/10.1371/journal.pone.0157368

Levental, K.R., Yu, H., Kass, L., Lakins, J.N., Egeblad, M., Erler, J.T., Fong, S.F.T., Csiszar, K., Giaccia, A., Weninger, W., Yamauchi, M., Gasser, D.L., Weaver, V.M., 2009. Matrix Crosslinking Forces Tumor Progression by Enhancing Integrin Signaling. Cell 139, 891–906. https://doi.org/10.1016/j.cell.2009.10.027

Li, B., Pan, R., Zhou, C., Dai, J., Mao, Y., Chen, M., Huang, T., Ying, X., Hu, H., Zhao, J., Zhang, W., Duan, S., 2018. SMYD3 promoter hypomethylation is associated with the risk of colorectal cancer. Future Oncol. 14, 1825–1834. https://doi.org/10.2217/fon-2017-0682

Li, E., Zhang, Y., 2014. DNA Methylation in Mammals. Cold Spring Harb. Perspect. Biol. 6, a019133. https://doi.org/10.1101/cshperspect.a019133

Lipsanen, V., Leinonen, P., Alhonen, L., Jänne, J., 1988. Hypomethylation of Ornithine Decarboxylase Gene and Erb-A1 Oncogene in Human Chronic Lymphatic Leukemia. Blood 72, 2042–2044. https://doi.org/10.1182/blood.V72.6.2042.2042

Liu, A.W., Villar-Briones, A., Luscombe, N.M., Plessy, C., 2022. Automated phenol-chloroform extraction of high molecular weight genomic DNA for use in long-read single-molecule sequencing. F1000Research 11, 240. https://doi.org/10.12688/f1000research.109251.1

Liu, Q., Thorland, E.C., Sommer, S.S., 1997. Inhibition of PCR Amplification by a Point Mutation Downstream of a Primer. BioTechniques 22, 292–300. https://doi.org/10.2144/97222st01

Liu, Y., Rosikiewicz, W., Pan, Z., Jillette, N., Wang, P., Taghbalout, A., Foox, J., Mason, C., Carroll, M., Cheng, A., Li, S., 2021. DNA methylation-calling tools for Oxford Nanopore sequencing: a survey and human epigenome-wide evaluation. Genome Biol. 22, 295. https://doi.org/10.1186/s13059-021-02510-z

Lo, C.M., Wang, H.B., Dembo, M., Wang, Y.L., 2000. Cell movement is guided by the rigidity of the substrate. Biophys. J. 79, 144–152.

LoRusso, P.M., 2016. Inhibition of the PI3K/AKT/mTOR Pathway in Solid Tumors. J. Clin. Oncol. https://doi.org/10.1200/JCO.2014.59.0018

Lucena-Aguilar, G., Sánchez-López, A.M., Barberán-Aceituno, C., Carrillo-Ávila, J.A., López-Guerrero, J.A., Aguilar-Quesada, R., 2016. DNA Source Selection for Downstream Applications Based on DNA Quality Indicators Analysis. Biopreservation Biobanking 14, 264–270. https://doi.org/10.1089/bio.2015.0064

Melki, J.R., Vincent, P.C., Clark, S.J., 1999. Concurrent DNA hypermethylation of multiple genes in acute myeloid leukemia. Cancer Res. 59, 3730–3740.

Michalak, E.M., Burr, M.L., Bannister, A.J., Dawson, M.A., 2019. The roles of DNA, RNA and histone methylation in ageing and cancer. Nat. Rev. Mol. Cell Biol. 20, 573–589. https://doi.org/10.1038/s41580-019-0143-1

Mirza, S., Sharma, G., Pandya, P., Ralhan, R., 2010. Demethylating agent 5-aza-2-deoxycytidine enhances susceptibility of breast cancer cells to anticancer agents. Mol. Cell. Biochem. 342, 101–109. https://doi.org/10.1007/s11010-010-0473-y

Miura, F., Enomoto, Y., Dairiki, R., Ito, T., 2012. Amplification-free whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. Nucleic Acids Res. 40, e136. https://doi.org/10.1093/nar/gks454

Moelans, C.B., Verschuur-Maes, A.H.J., van Diest, P.J., 2011. Frequent promoter hypermethylation of BRCA2, CDH13, MSH6, PAX5, PAX6 and WT1 in ductal carcinoma in situ and invasive breast cancer. J. Pathol. 225, 222–231. https://doi.org/10.1002/path.2930

Mohandas, T., Sparkes, R.S., Shapiro, L.J., 1981. Reactivation of an Inactive Human X Chromosome: Evidence for X Inactivation by DNA Methylation. Science 211, 393–396. https://doi.org/10.1126/science.6164095

Moore, L.D., Le, T., Fan, G., 2013. DNA Methylation and Its Basic Function. Neuropsychopharmacology 38, 23–38. https://doi.org/10.1038/npp.2012.112

Muiznieks, L.D., Keeley, F.W., 2013. Molecular assembly and mechanical properties of the extracellular matrix: A fibrous protein perspective. Biochim. Biophys. Acta BBA - Mol. Basis Dis., Fibrosis: Translation of basic research to human disease 1832, 866–875. https://doi.org/10.1016/j.bbadis.2012.11.022

O'Shaughnessy, J., Schwartzberg, L., Danso, M.A., Miller, K.D., Rugo, H.S., Neubauer, M., Robert, N., Hellerstedt, B., Saleh, M., Richards, P., Specht, J.M., Yardley, D.A., Carlson, R.W., Finn, R.S., Charpentier, E., Garcia-Ribas, I., Winer, E.P., 2014. Phase III study of iniparib plus gemcitabine and carboplatin versus gemcitabine and carboplatin in patients with metastatic triple-negative breast cancer. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. 32, 3840–3847. https://doi.org/10.1200/JCO.2014.55.2984

Papoutsis, A.J., Borg, J.L., Selmin, O.I., Romagnolo, D.F., 2012. BRCA-1 promoter hypermethylation and silencing induced by the aromatic hydrocarbon receptor-ligand TCDD are prevented by resveratrol in MCF-7 Cells. J. Nutr. Biochem. 23, 1324–1332. https://doi.org/10.1016/j.jnutbio.2011.08.001

Parsons, M.J., Vertino, P.M., 2006. Dual role of TMS1/ASC in death receptor signaling. Oncogene 25, 6948–6958. https://doi.org/10.1038/sj.onc.1209684

Paydar, P., Asadikaram, G., Nejad, H.Z., Akbari, H., Abolhassani, M., Moazed, V., Nematollahi, M.H., Ebrahimi, G., Fallah, H., 2019. Epigenetic modulation of BRCA-1 and MGMT genes, and histones H4 and H3 are associated with breast tumors. J. Cell. Biochem. 120, 13726–13736. https://doi.org/10.1002/jcb.28645

Péterfi, Z., Donkó, A., Orient, A., Sum, A., Prókai, A., Molnár, B., Veréb, Z., Rajnavölgyi, E., Kovács, K.J., Müller, V., Szabó, A.J., Geiszt, M., 2009. Peroxidasin is secreted and incorporated into the extracellular matrix of myofibroblasts and fibrotic kidney. Am. J. Pathol. 175, 725–735. https://doi.org/10.2353/ajpath.2009.080693

Postovit, L.-M., Abbott, D.E., Payne, S.L., Wheaton, W.W., Margaryan, N.V., Sullivan, R., Jansen, M.K., Csiszar, K., Hendrix, M.J.C., Kirschmann, D.A., 2008. Hypoxia/reoxygenation: A dynamic regulator of lysyl oxidase-facilitated breast cancer migration. J. Cell. Biochem. 103, 1369–1378. https://doi.org/10.1002/jcb.21517

Prat, A., Pineda, E., Adamo, B., Galván, P., Fernández, A., Gaba, L., Díez, M., Viladot, M., Arance, A., Muñoz, M., 2015. Clinical implications of the intrinsic molecular subtypes of breast cancer. The Breast 24, S26–S35. https://doi.org/10.1016/j.breast.2015.07.008

Pritchard, K.I., Burris, H.A., Ito, Y., Rugo, H.S., Dakhil, S., Hortobagyi, G.N., Campone, M., Csöszi, T., Baselga, J., Puttawibul, P., Piccart, M., Heng, D., Noguchi, S., Srimuninnimit, V., Bourgeois, H., Gonzalez Martin, A., Osborne, K., Panneerselvam, A., Taran, T., Sahmoud, T., Gnant, M., 2013. Safety and efficacy of everolimus with exemestane vs. exemestane alone in elderly patients with HER2-negative, hormone receptor-positive breast cancer in BOLERO-2. Clin. Breast Cancer 13, 421-432.e8. https://doi.org/10.1016/j.clbc.2013.08.011

Redza-Dutordoir, M., Averill-Bates, D.A., 2016. Activation of apoptosis signalling pathways by reactive oxygen species. Biochim. Biophys. Acta BBA - Mol. Cell Res. 1863, 2977–2992. https://doi.org/10.1016/j.bbamcr.2016.09.012

Reik, W., Collick, A., Norris, M.L., Barton, S.C., Surani, M.A., 1987. Genomic imprinting determines methylation of parental alleles in transgenic mice. Nature 328, 248–251. https://doi.org/10.1038/328248a0

Rhyu, D.Y., Yang, Y., Ha, H., Lee, G.T., Song, J.S., Uh, S., Lee, H.B., 2005. Role of Reactive Oxygen Species in TGF-β1-Induced Mitogen-Activated Protein Kinase Activation and Epithelial-Mesenchymal Transition in Renal Tubular Epithelial Cells. J. Am. Soc. Nephrol. 16, 667–675. https://doi.org/10.1681/ASN.2004050425

Ricard-Blum, S., 2011. The Collagen Family. Cold Spring Harb. Perspect. Biol. 3. https://doi.org/10.1101/cshperspect.a004978

Rideout, W.M., Coetzee, G.A., Olumi, A.F., Jones, P.A., 1990. 5-Methylcytosine as an Endogenous Mutagen in the Human LDL Receptor and p53 Genes. Science 249, 1288–1290. https://doi.org/10.1126/science.1697983

Schrader, C., Schielke, A., Ellerbroek, L., Johne, R., 2012. PCR inhibitors – occurrence, properties and removal. J. Appl. Microbiol. 113, 1014–1026. https://doi.org/10.1111/j.1365-2672.2012.05384.x

Schüler-Toprak, S., Häring, J., Inwald, E.C., Moehle, C., Ortmann, O., Treeck, O., 2016. Agonists and knockdown of estrogen receptor β differentially affect invasion of triple-negative breast cancer cells in vitro. BMC Cancer 16, 951. https://doi.org/10.1186/s12885-016-2973-y

Selleck, M.J., Senthil, M., Wall, N.R., 2017. Making Meaningful Clinical Use of Biomarkers. Biomark. Insights 12. https://doi.org/10.1177/1177271917715236

Semenza, G.L., 2007. Hypoxia-Inducible Factor 1 (HIF-1) Pathway. Sci. STKE 2007, cm8–cm8. https://doi.org/10.1126/stke.4072007cm8

Shanle, E.K., Zhao, Z., Hawse, J., Wisinski, K., Keles, S., Yuan, M., Xu, W., 2013. Research Resource: Global Identification of Estrogen Receptor β Target Genes in Triple Negative Breast Cancer Cells. Mol. Endocrinol. 27, 1762–1775. https://doi.org/10.1210/me.2013-1164

Shargh, S.A., Sakizli, M., Khalaj, V., Movafagh, A., Yazdi, H., Hagigatjou, E., Sayad, A., Mansouri, N., Mortazavi-Tabatabaei, S.A., Khorram Khorshid, H.R., 2014. Downregulation of E-cadherin expression in breast cancer by promoter hypermethylation and its relation with progression and prognosis of tumor. Med. Oncol. 31, 250. https://doi.org/10.1007/s12032-014-0250-y

Shulman, L.N., Willett, W., Sievers, A., Knaul, F.M., 2010. Breast Cancer in Developing Countries: Opportunities for Improved Survival. J. Oncol. 2010. https://doi.org/10.1155/2010/595167

Sigurdardottir, A.K., Jonasdottir, A.S., Asbjarnarson, A., Helgudottir, H.R., Gudjonsson, T., Traustadottir, G.A., 2021. Peroxidasin Enhances Basal Phenotype and Inhibits Branching Morphogenesis in Breast Epithelial Progenitor Cell Line D492. J. Mammary Gland Biol. Neoplasia 26, 321–338. https://doi.org/10.1007/s10911-021-09507-1

Simpson, J.T., Workman, R.E., Zuzarte, P.C., David, M., Dursi, L.J., Timp, W., 2017. Detecting DNA cytosine methylation using nanopore sequencing. Nat. Methods 14, 407–410. https://doi.org/10.1038/nmeth.4184

Singer, J., Roberts-Ems, J., Riggs, A.D., 1979. Methylation of mouse liver DNA studied by means of the restriction enzymes msp I and hpa II. Science 203, 1019–1021. https://doi.org/10.1126/science.424726

Singer-Sam, J., LeBon, J.M., Tanguay, R.L., Riggs, A.D., 1990. A quantitative HpaII-PCR assay to measure methylation of DNA from a small number of cells. Nucleic Acids Res. 18, 687.

Singh, B., Sarli, V.N., Lucci, A., 2021. Inhibition of resistant triple-negative breast cancer cells with low-dose 6-mercaptopurine and 5-azacitidine. Oncotarget 12, 626–637. https://doi.org/10.18632/oncotarget.27922

Sitole, B.N., Mavri-Damelin, D., 2018. Peroxidasin is regulated by the epithelial-mesenchymal transition master transcription factor Snai1. Gene 646, 195–202. https://doi.org/10.1016/j.gene.2018.01.011

Sørlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Thorsen, T., Quist, H., Matese, J.C., Brown, P.O., Botstein, D., Lønning, P.E., Børresen-Dale, A.-L., 2001. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc. Natl. Acad. Sci. U. S. A. 98, 10869–10874. https://doi.org/10.1073/pnas.191367098

Stirzaker, C., Zotenko, E., Song, J.Z., Qu, W., Nair, S.S., Locke, W.J., Stone, A., Armstong, N.J., Robinson, M.D., Dobrovic, A., Avery-Kiejda, K.A., Peters, K.M., French, J.D., Stein, S., Korbie, D.J., Trau, M., Forbes, J.F., Scott, R.J., Brown, M.A., Francis, G.D., Clark, S.J., 2015. Methylome sequencing in triple-negative breast cancer reveals distinct methylation clusters with prognostic value. Nat. Commun. 6, 5899. https://doi.org/10.1038/ncomms6899

Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F., 2021. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA. Cancer J. Clin. 71, 209–249. https://doi.org/10.3322/caac.21660

Swain, S.M., Baselga, J., Kim, S.-B., Ro, J., Semiglazov, V., Campone, M., Ciruelos, E., Ferrero, J.-M., Schneeweiss, A., Heeson, S., Clark, E., Ross, G., Benyunes, M.C., Cortés, J., CLEOPATRA Study Group, 2015. Pertuzumab, trastuzumab, and docetaxel in HER2-positive metastatic breast cancer. N. Engl. J. Med. 372, 724–734. https://doi.org/10.1056/NEJMoa1413513

Tang, Q., Cheng, J., Cao, X., Surowy, H., Burwinkel, B., 2016. Blood-based DNA methylation as biomarker for breast cancer: a systematic review. Clin. Epigenetics 8. https://doi.org/10.1186/s13148-016-0282-6

Tauber, S., Jais, A., Jeitler, M., Haider, S., Husa, J., Lindroos, J., Knöfler, M., Mayerhofer, M., Pehamberger, H., Wagner, O., Bilban, M., 2010. Transcriptome analysis of human cancer reveals a functional role of Heme Oxygenase-1 in tumor cell adhesion. Mol. Cancer 9, 200. https://doi.org/10.1186/1476-4598-9-200

Theocharis, A.D., Manou, D., Karamanos, N.K., 2019. The extracellular matrix as a multitasking player in disease. FEBS J. 286, 2830–2869. https://doi.org/10.1111/febs.14818

Theocharis, A.D., Skandalis, S.S., Gialeli, C., Karamanos, N.K., 2016. Extracellular matrix structure. Adv. Drug Deliv. Rev., Extracellular Matrix (ECM) and ECM-like materials: Therapeutic Tools and Targets in Cancer Treatment 97, 4–27. https://doi.org/10.1016/j.addr.2015.11.001

Tindall, A.J., Pownall, M.E., Morris, I.D., Isaacs, H.V., 2005. Xenopus tropicalis peroxidasin gene is expressed within the developing neural tube and pronephric kidney. Dev. Dyn. 232, 377–384. https://doi.org/10.1002/dvdy.20226

Torii, S., Furumai, H., Katayama, H., 2021. Applicability of polyethylene glycol precipitation followed by acid guanidinium thiocyanate-phenol-chloroform extraction for the detection of SARS-CoV-2 RNA from municipal wastewater. Sci. Total Environ. 756, 143067. https://doi.org/10.1016/j.scitotenv.2020.143067

Uhlen, M., Zhang, C., Lee, S., Sjöstedt, E., Fagerberg, L., Bidkhori, G., Benfeitas, R., Arif, M., Liu, Z., Edfors, F., Sanli, K., von Feilitzen, K., Oksvold, P., Lundberg, E., Hober, S., Nilsson, P., Mattsson, J., Schwenk, J.M., Brunnström, H., Glimelius, B., Sjöblom, T., Edqvist, P.-H., Djureinovic, D., Micke, P., Lindskog, C., Mardinoglu, A., Ponten, F., 2017. A pathology atlas of the human cancer transcriptome. Science 357, eaan2507. https://doi.org/10.1126/science.aan2507

Varley, K.E., Gertz, J., Bowling, K.M., Parker, S.L., Reddy, T.E., Pauli-Behn, F., Cross, M.K., Williams, B.A., Stamatoyannopoulos, J.A., Crawford, G.E., Absher, D.M., Wold, B.J., Myers, R.M., 2013. Dynamic DNA methylation across diverse human cell lines and tissues. Genome Res. 23, 555–567. https://doi.org/10.1101/gr.147942.112

Veal, C.D., Freeman, P.J., Jacobs, K., Lancaster, O., Jamain, S., Leboyer, M., Albanes, D., Vaghela, R.R., Gut, I., Chanock, S.J., Brookes, A.J., 2012. A mechanistic basis for amplification differences between samples and between genome regions. BMC Genomics 13, 455. https://doi.org/10.1186/1471-2164-13-455

Vrtačnik, P., Ostanek, B., Mencej-Bedrač, S., Marc, J., 2014. The many faces of estrogen signaling. Biochem. Medica 24, 329–342. https://doi.org/10.11613/BM.2014.035

Walker, C., Mojares, E., Del Río Hernández, A., 2018. Role of Extracellular Matrix in Development and Cancer Progression. Int. J. Mol. Sci. 19. https://doi.org/10.3390/ijms19103028

Warnecke, P.M., Stirzaker, C., Melki, J.R., Millar, D.S., Paul, C.L., Clark, S.J., 1997. Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA. Nucleic Acids Res. 25, 4422–4426.

Wojdacz, T.K., Dobrovic, A., Hansen, L.L., 2008. Methylation-sensitive high-resolution melting. Nat. Protoc. 3, 1903–1908. https://doi.org/10.1038/nprot.2008.191

Wojdacz, T.K., Hansen, L.L., 2006. Reversal of PCR bias for improved sensitivity of the DNA methylation melting curve assay. BioTechniques 41, 274–278. https://doi.org/10.2144/000112240

Wolff, E.M., Byun, H.-M., Han, H.F., Sharma, S., Nichols, P.W., Siegmund, K.D., Yang, A.S., Jones, P.A., Liang, G., 2010. Hypomethylation of a LINE-1 Promoter Activates an Alternate Transcript of the MET Oncogene in Bladders with Cancer. PLoS Genet. 6, e1000917. https://doi.org/10.1371/journal.pgen.1000917

Wong, K.K., 2021. DNMT1: A key drug target in triple-negative breast cancer. Semin. Cancer Biol., Precision Medicine in Breast Cancer 72, 198–213. https://doi.org/10.1016/j.semcancer.2020.05.010

Worm, J., Aggerholm, A., Guldberg, P., 2001. In-tube DNA methylation profiling by fluorescence melting curve analysis. Clin. Chem. 47, 1183–1189.

Wouters, B.J., Delwel, R., 2016. Epigenetics and approaches to targeted epigenetic therapy in acute myeloid leukemia. Blood 127, 42–52. https://doi.org/10.1182/blood-2015-07-604512

Yang, X., Han, H., De Carvalho, D.D., Lay, F.D., Jones, P.A., Liang, G., 2014. Gene Body Methylation Can Alter Gene Expression and Is a Therapeutic Target in Cancer. Cancer Cell 26, 577–590. https://doi.org/10.1016/j.ccr.2014.07.028

Yegnasubramanian, S., Lin, X., Haffner, M.C., DeMarzo, A.M., Nelson, W.G., 2006. Combination of methylated-DNA precipitation and methylation-sensitive restriction enzymes (COMPARE-MS) for the rapid, sensitive and quantitative detection of DNA methylation. Nucleic Acids Res. 34, e19–e19. https://doi.org/10.1093/nar/gnj022

York, K.T., Smith, R.C., Yang, R., Melnyk, P.C., Wiley, M.M., Turk, C.M., Ronaghi, M., Gunderson, K.L., Steemers, F.J., 2012. Highly parallel oligonucleotide purification and functionalization using reversible chemistry. Nucleic Acids Res. 40, e4. https://doi.org/10.1093/nar/gkr910

Yu, J., Qin, B., Moyer, A.M., Nowsheen, S., Liu, T., Qin, S., Zhuang, Y., Liu, D., Lu, S.W., Kalari, K.R., Visscher, D.W., Copland, J.A., McLaughlin, S.A., Moreno-Aspitia, A., Northfelt, D.W., Gray, R.J., Lou, Z., Suman, V.J., Weinshilboum, R., Boughey, J.C., Goetz, M.P., Wang, L., 2018. DNA methyltransferase expression in triple-negative breast cancer predicts sensitivity to decitabine. J. Clin. Invest. 128, 2376–2388. https://doi.org/10.1172/JCI97924

Zafon, C., Gil, J., Pérez-González, B., Jordà, M., 2019. DNA methylation in thyroid cancer. Endocr. Relat. Cancer 26, R415–R439. https://doi.org/10.1530/ERC-19-0093

Zemach, A., McDaniel, I.E., Silva, P., Zilberman, D., 2010. Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science 328, 916–919. https://doi.org/10.1126/science.1186366

Zhang, J.-L., Huang, Y., Qiu, L.-Y., Nickel, J., Sebald, W., 2007. von Willebrand Factor Type C Domain-containing Proteins Regulate Bone Morphogenetic Protein Signaling through Different Recognition Mechanisms. J. Biol. Chem. 282, 20002–20014. https://doi.org/10.1074/jbc.M700456200

Zheng, Y.-Z., Liang, L., 2018. High expression of PXDN is associated with poor prognosis and promotes proliferation, invasion as well as migration in ovarian cancer. Ann. Diagn. Pathol. 34, 161–165. https://doi.org/10.1016/j.anndiagpath.2018.03.002

Zhou, X., Sun, Q., Xu, C., Zhou, Z., Chen, X., Zhu, X., Huang, Z., Wang, W., Shi, Y., 2022. A systematic pan-cancer analysis of PXDN as a potential target for clinical diagnosis and treatment. Front. Oncol. 12.

Zhu, H., Wang, G., Qian, J., 2016. Transcription factors as readers and effectors of DNA methylation. Nat. Rev. Genet. 17, 551–565. https://doi.org/10.1038/nrg.2016.83

7 Appendix

7.1 Recipes

7.1.1 0.5 M EDTA

Component	Mass (g)	Volume (mL)
Na2EDTA	18.2	
NaOH	Add until pH reaches 8.0	
dH_2O		Up to 100

Following preparation, buffer was autoclaved to ensure sterilisation.

7.1.2 50 x TAE Buffer

Component	Mass (g)	Volume (mL)
Tris	242	
Glacial acetic acid		57.1
0.5 M EDTA		100
dH_2O		Up to 1000

After the buffer was prepared pH was adjusted to 8.0. This was followed by autoclaving for sterilisation.

7.1.3 10 x PBS buffer

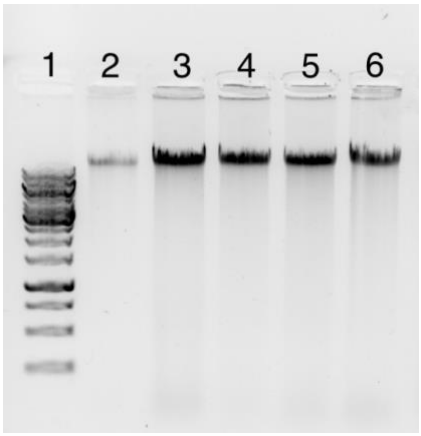
Component	Mass (g)	Volume (mL)
NaCl	80	
KCl	2	
8.1 mM Na2HPO4•12H2O	29.01	
1.8mM KH2PO4 pH 7	2.4	
dH_2O		Up to 1000

Buffer was autoclaved to ensure sterilisation. For experiments, 10 x PBS buffer was diluted with dH_2O to make 1 x PBS.

7.2 DNA Extraction Results

7.2.1 Phenol Chloroform-extracted gDNA

A

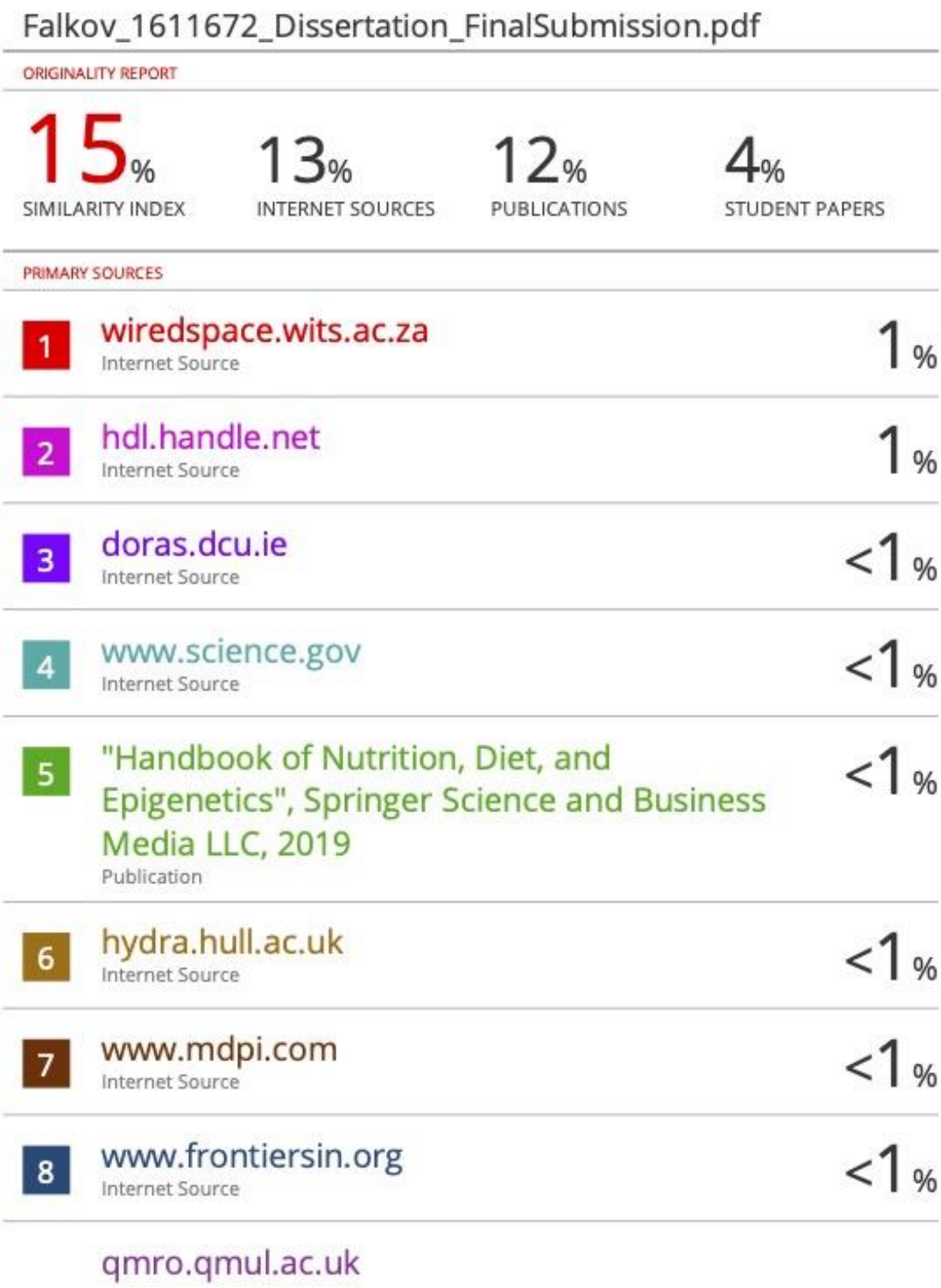


B

Lane	Sample	Nucleic Acid Concentration (ng/µL)	A260/280	A260/230
2	MDA-MB-231 dish 1	162.062	1.929	2.343
3	MDA-MB-231 dish 2	217.241	1.961	2.242
4	MDA-MB-231 dish 3	135.655	1.927	2.304
5	MDA-MB-231 dish 4	166.057	1.945	2.298
6	MDA-MB-231 dish 5	363.599	1.965	2.435

Appendix Fig. 1: Agarose electrophoresis gel indicating gDNA extracted from MDA-MB-231 cells using the phenol chloroform extraction protocol. A Lane 1 contains the NEB 1 kb MWM. Lanes 2 to 6 contain samples extracted from MDA-MB-231 cell using this protocol. **B** Absorbance ratios and concentrations of extracted DNA.

7.3 Turnitin Similarity Index



Falkov_1611672_Dissertation_FinalSubmission.pdf

ORIGINALITY REPORT

15%	13%	12%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	wiredspace.wits.ac.za Internet Source	1%
2	hdl.handle.net Internet Source	1%
3	doras.dcu.ie Internet Source	<1%
4	www.science.gov Internet Source	<1%
5	"Handbook of Nutrition, Diet, and Epigenetics", Springer Science and Business Media LLC, 2019 Publication	<1%
6	hydra.hull.ac.uk Internet Source	<1%
7	www.mdpi.com Internet Source	<1%
8	www.frontiersin.org Internet Source	<1%
	qmro.qmul.ac.uk	