THE UNIVERSITY OF HULL

Investigation of the regulatory role of CBX2 in breast cancer via *in silico* transcriptomic analysis and *in vitro* cistromic analysis.

being a Thesis submitted for the degree of

MSc by Research, Biomedical Science

in the University of Hull

by

Shannon Cartay Kalsi, BSc (Hons) Biomedical Science.

September 2021

Abstract

Chromobox 2 (CBX2) is a component of Polycomb repressive protein complex 1 (PRC1), which is an epigenetic regulatory complex that downregulates the expression of target genes. CBX2 has been linked to cancer progression and development in several cancer types, with evidence showing it may play a role in a particularly aggressive form of breast cancer called triple-negative breast cancer (TNBC). TNBC is highly metastatic, has a poor prognosis and suffers from a lack of targeted therapeutics.

This study aimed to identify oncogenic processes regulated by CBX2 in breast cancer to aid further understanding of its potential role in this disease. Oncogenic pathways putatively regulated by CBX2 were determined via bioinformatic analysis of publicly available patient datasets using Gene Set Enrichment Analysis (GSEA) and analysis of RNA-sequencing data from a TNBC cell line (MDA-MB-231).

In silico analysis showed that CBX2 was upregulated in TNBC compared to normal breast tissue and elevated levels of CBX2 expression were associated with a decrease in overall survival (OS) and disease-free survival (DFS) in breast cancer. Our research identified that CBX2 was associated with upregulation of E2F and mTORC1 signalling pathways. Analysis of RNA sequencing data from MDA-MB-231 cells, which were depleted of CBX2, showed upregulation of genes that code for inhibitors of E2F (*RBL2*) and mTORC1 (*TSC1, TSC2* and *PRKAA2*) signalling, indicating that CBX2 represses the expression of these putative tumour suppressor genes in TNBC. Chromatin immunoprecipitation and CUTandRUN analysis in MDA-MB-231 cells indicated that CBX2 was bound to the promotor regions of *RBL2, TSC1* and *PRKAA2,* indicating that CBX2 may directly regulate the expression of these genes.

Overall, this analysis identified oncogenic processes regulated by CBX2 in TNBC, and for the first time identified the potential mechanism by which CBX2 promotes these pathways. This knowledge is crucial to understand the specific role of CBX2 in TNBC and to aid investigations of CBX2 as a potentially novel therapeutic target.

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr Mark Wade for his patience, expertise, and support throughout the project. Thank you for all your help and for being a great supervisor. I would also like to thank Dr Pedro Beltran-Alvarez, for his support and knowledge throughout my research. Additionally, I would like to thank Rebecca Humphries, for her help, guidance, and support throughout my project. I would also like to thank Sabrina Samuel and Alistair Marsden, for the help and support they offered during my time in the laboratory. Furthermore, I would like to thank Dr Isabel Pires and Emily Pyne, for the extra resources and support during this project. Finally, I would like to thank my friend Jessica Lee and partner Nathaniel Jones, for the continued support and reassurance during this research.

Table of Contents

Table of Figures

Table of Tables

Table of Abbreviations

Abbreviation	Description
BLBC	Basal-like breast cancer
Bp	Base pairs
CBX	Chromobox
CBX2	Chromobox 2
CFC	Colony-forming cell
ChIP	Chromatin immunoprecipitation
CNA	Copy-number alterations
CoA	Acetyl coenzyme A
Co-A	Co-activator
Co-B	Co-repressor
DFS	Disease-free survival
DiH_2O	Deionised water
DMEM	Dulbecco's Modified Eagle Medium
E_2	Oestrogen
EED	Embryonic Ectoderm Development
ER	Oestrogen receptor
EREs	Oestrogen response elements
ERα	Oestrogen receptor alpha
Erβ	Oestrogen receptor beta
ES	Enrichment score
ESC	Embryonic stem cells
FDR	False discovery rate
GBM	Glioblastoma multiforme
GEPIA	Gene Expression Profiling Interactive Analysis 2
GSEA	Gene Set Enrichment Analysis
HATs	Histone acetylases
HDACs	Histone deacetylases
HER2	Human epidermal growth factor receptor 2
HGSOC	High grade serous ovarian carcinoma
HMTs	Histone methyltransferase
HPV	Human papilloma virus
HR	Hazard ratio
HSP	Heat shock proteins
HSPCs	Haematopoietic stem and progenitor cells
LTC-IC	Limiting-dilution long-term culture initiating cell
LUAD	Lung adenocarcinoma
MDS	Myelodysplastic syndrome
NES	Normalised enrichment score
OS	Overall survival
OV	Ovarian serous cystadenocarcinoma
PARPs	Poly(adenosine diphosphate-ribose) polymerase
PBS	Phosphate buffer saline

PcGs	Polycomb group complexes
pCR	Pathological complete response
PDS	Pathway deregulation score
PFS	Progression-free survival
PIC	Protease Inhibitor Cocktail
PR	Progesterone receptor
PRAD	Prostate adenocarcinoma
PRC1	Polycomb repressive protein complex 1
PRC2	Polycomb repressive protein complex 2
PTMs	Post translational modifications
RT-qPCR	Real-time quantitative polymerase chain reaction
SDS	Sodium dodecyl sulphate
SERDs	Selective oestrogen receptor downregulators
SERMs	Selective oestrogen receptor modulators
TBS	Tris-buffered saline
TBST	Tris-buffered saline with TWEEN
TCGA	The Cancer Genome Atlas
TGCT	Testicular germ cell tumours
TNBC	Triple-negative breast cancer
TPM	Transcripts per million

1. Introduction

Cancer is defined as a group of diseases characterised by an unregulated, uncontrollable hyperproliferation of cells with the potential to metastasise to other organs or tissues. Cancer is a leading health concern, with an increasing prevalence and incidence rate each year worldwide (Uhlen *et al*., 2017). The most common cancer types in the UK are lung, bowel, breast, and prostate cancer, which are accountable for 53% of all cancer deaths (Cancer Research, 2017). Cancer is caused by an accumulation of mutations in the genome via DNA damage not being repaired. DNA damage can be induced by exogenous agents (such as ionising radiation or chemical carcinogens) or due to errors in endogenous DNA replication or repair processes (Chatterjee & Walker, 2017).

Cancer development and progression is a multiphase process in which the alteration of a normal cell to malignancy is characterised. Hanahan & Weinberg identified six specific hallmarks of cancer in 2000. Subsequently, four further hallmarks have been acknowledged (Figure 1.1): pro-tumour inflammation, immune destruction avoidance, deregulation of cellular energetics and genomic instability and mutations (Hanahan & Weinberg, 2011). These hallmarks allow us to characterise malignant cells and provide the basis for the advancement of cancer therapies as well as providing an understanding of therapeutic resistance.

Figure 1.1. Hallmarks of Cancer
The ten hallmarks of cancer identified by Hanahan & Weinberg (2011).

1.1. Breast Cancer

One of the leading causes of death in women is breast cancer, which is responsible for 450,000 deaths per year worldwide (The Cancer Genome Atlas Network, 2012; Sun *et al*., 2017). Breast cancer is a malignancy originating from breast tissue, which in some cases, metastasises to distant organs. Breast cancer is broadly categorised into two groups: invasive and non-invasive breast carcinoma. Non-invasive cancers are characterised as an intraductal proliferation of epithelial cells within ducts and lobules, while invasive cancers are characterised by an abnormal growth of cells that have expanded from the duct lobular unit into the surrounding tissue (Sharma *et al*., 2010; Akram *et al*., 2017). Sun *et al*., (2017) identified that sex, age, hormone levels, gene mutations and poor lifestyle are all risk factors that can contribute to breast cancer development.

The breast is comprised of glandular tissues, which hold the lobules (milk-producing glands), ductal tissue, adipose tissue and stromal tissues which are fatty and fibrous connective tissues (Zhu & Nelson, 2013). Most breast cancers begin in the line of the ducts, with some starting in the lobules and a small number beginning in the other tissues (Sharma *et al*., 2010). Figure 1.2 shows the anatomical sites of breast cancer, with over half the cancers not developing in a specific area or overlapping in more than one (Cancer Research, 2017).

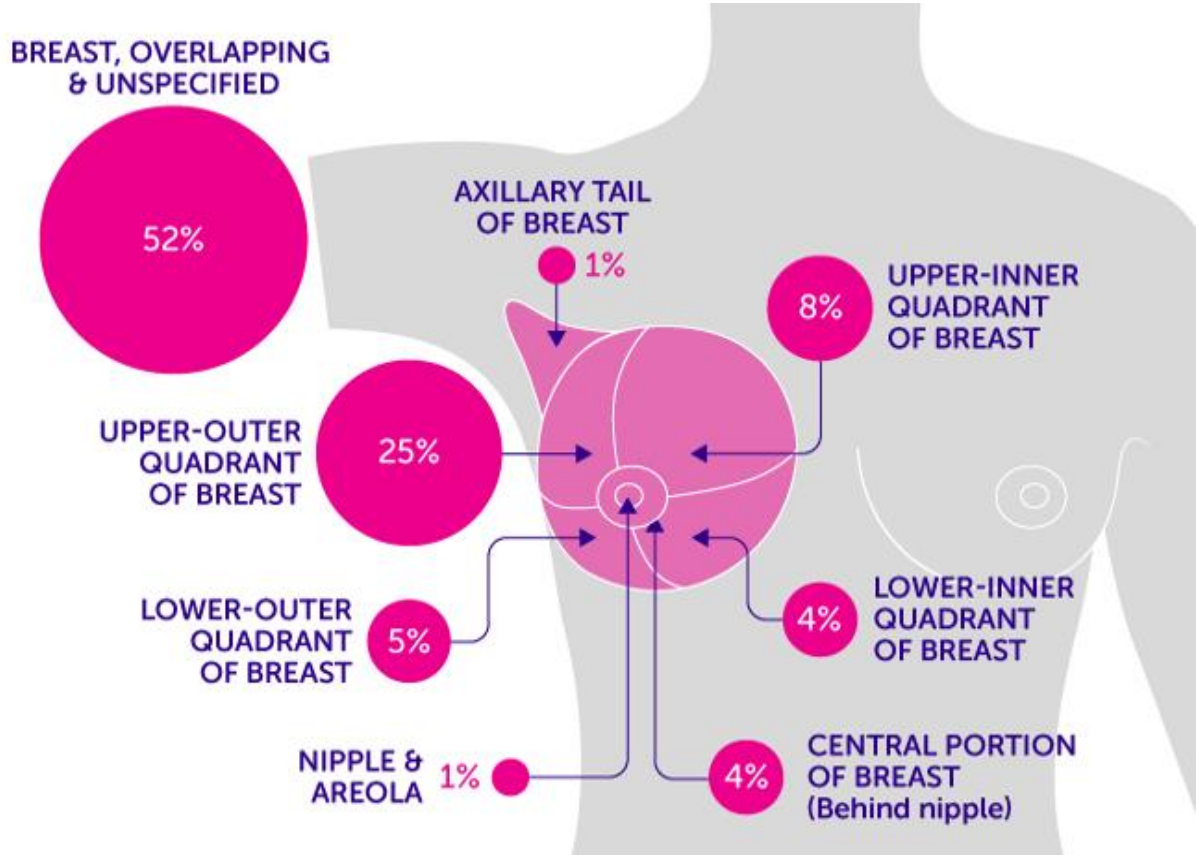

Figure 1.2. Anatomical sites of breast cancer.
Diagram showing the percentage distribution of breast cancer origin in patients, between 2010-2012, from the cancer registry data. Taken from Cancer Research UK (2017).

Breast cancer is graded as Stage I-IV, to indicate the severity of each case. The Tumour, Nodes, and Metastases (TNM) staging system is used in conjunction with this. The TNM system considers the size of the tumours (T) if it has spread to the lymph nodes (N) or if it has metastasised (M). Stage I cancer is categorised by the absence or presence of a 2 cm diameter tumour with, or without, small cancerous areas found in the lymph node. Stage II is when the tumour either grows to between 2-5 cm with, or without, cancerous cells spreading to one-three lymph nodes. Late-stage breast cancer (Stage III and IV) are more aggressive. Stage III is characterised by a tumour larger than 5 cm with or without, the tumour spreading to three lymph nodes. It can also have more clusters of cancer cells in four-nine lymph nodes. The cancer may spread into the skin of the breast (causing an ulcer/abscess), the chest or other structures around the breast. These tumours are determined as operable or inoperable. The final stage (IV) is a tumour of any size, with cancerous or non-cancerous lymph nodes, and has

metastasised to other areas of the body (like the brain or lungs). Most patients will receive a stage at diagnosis (shown in Appendix) with most patients (79-87%) being identified at an early stage (I or II) (Cancer Research, 2017).

1.2. Subtypes of breast cancer

Breast cancer is traditionally classified histologically by tumour grade, type, and lymph node stage. This system of classification is descriptive and relatively subjective, with the histological appearance of tumours being unable to provide insight on the genetic alteration or the progression of cancer (Abd El-Rehim *et al*., 2005). However, the use of traditional clinicopathological variables used in conjunction with immunohistochemistry markers is now widely used for prognosis and treatment. The common immunohistochemistry markers used for identification are oestrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2). Breast cancer has also been defined into several different molecular subtypes. To redefine breast cancer classification, gene expression profiling classified breast cancer tumours into five subtypes; luminal A, luminal B, HER2 over-expression, basal and normal-like (Sørlie *et al*., 2001). Two other studies analysing gene expression profiles of breast cancer patients confirmed the stratification of breast cancer into these five molecular subtypes (Hu *et al*., 2006; Gnant *et al*., 2014). These subtypes were also identified in another study using tissue microarray (Abd El-Rehim *et al*., 2005). However, other classifications of breast cancer exist. Two studies, done by Dai *et al* (2015) and The Cancer Genome Atlas Network (2012), used an integrative data analysis approach, incorporating multiple platforms of information e.g., DNA copy numbers, DNA methylation miRNA sequencing, which classify breast cancer into luminal A, luminal B, HER2 positive and triple-negative (The Cancer Genome Atlas Network, 2012; Dai *et al*., 2015). Despite the different classification systems, it is agreed that breast cancer separates into three main subtypes, luminal, HER2 and triple-negative/basal-like.

1.2.1. Luminal

Around three-quarters of breast cancer are positive for ER or PR and can be classified via immunohistological staining (Anderson *et al*., 2002). A study of 1944 primary operable invasive breast

cancers used immunohistological staining to determine the receptors on the tumours. Of the 1944 patients, 71% were ER positive, 59% were PR positive and 55.3% of patients (n=963) were ER and PR positive, indicated the majority of luminal breast cancers are ER/PR positive (Rakha *et al*., 2007). Luminal B tumours can also be distinguished from luminal A via the presence of proliferation marker Ki67 (Cheang *et al*., 2009). Luminal A has a high expression of ER related genes, but lower expressions of genes related to cell cycle and a lower recurrence score compared to luminal B (Gnant *et al*., 2014). Luminal B tends to have tumours at a more advanced stage and a worse prognosis (Gnant *et al*., 2014; Dai *et al*., 2015). This indicates patients who have luminal B tumours have a worse prognosis and higher recurrence than luminal A tumours. Endocrine or hormonal therapy, such as Tamoxifen is the most used treatment for luminal breast cancer. Using tamoxifen for 5 years can reduce breast cancer mortality by 34% (Abe *et al*., 2005).

There are two types of ER: ERα and ERβ. ERβ is expressed in both normal and malignant breast tissue and has been found to have a possible tumour-suppressing role by interacting with p53. Mice with a conditional p53 knockout and Esr2 knockout (the mouse Erβ gene) displayed more malignant phenotypes compared to mice with just p53 knockout (Feng *et al*., 2018). ERα (Figure 1.4a) is expressed in 50-80% of breast cancers and is responsible for increased cell proliferation that is mediated by its ligand oestrogen (E_2) (Paterni *et al.,* 2014; Huang *et al*., 2015). E_2 and its metabolites, such as 17β-estradiol, can diffuse into tumour cells. ERs have a heat shock protein (HSP) bound to them, which dissociates when E_2 binds and forms the E_2-ER complex. The E_2-ER complex then dimerises, allowing co-repressor and co-activator molecules to be recruited (Figure 1.4b) (Osborne & Schiff, 2011). The complex (promoted by Forkhead protein FOXA1) can bind to the oestrogen response elements (EREs) located in promotor regions of oestrogen responsive genes, thereby activating or repressing transcription (Hurtado *et al*., 2011). There are two hypotheses for the method of interaction between ER and EREs: direct binding or indirect (tethering). Direct binding is where the E_2-ER binds to the EREs with co-factors, thereby recruiting the RNA polymerase II transcriptional complex resulting in transcription; whereas tethering is when the ER does not bind to DNA, but it interacts with

a DNA-bound transcription factor to stabilises and/or recruit coactivators to the complex (Stice & Knowlton, 2008). These mechanisms drive the expression of pro-proliferative genes to drive cancer growth.

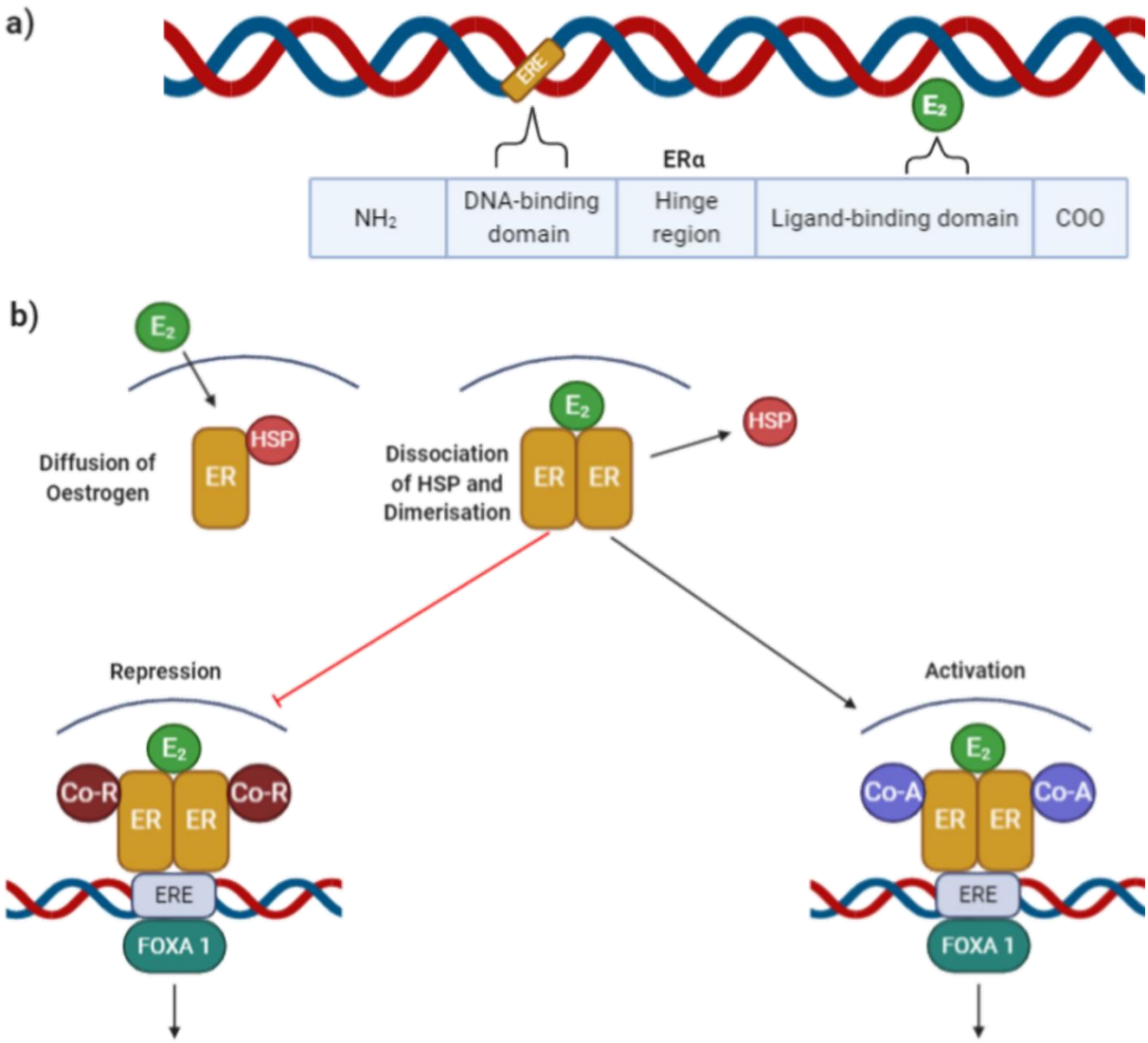

Figure 1.3. a) Functional components of ERα. b) Oestrogen stimulates proliferation within breast cancer cells. a) DNA-Binding domain binds oestrogen response elements (ERE) onto DNA. The DNA-binding domain is between the N-terminal domain and the ligand-binding domain. The ligand-binding domain allows binding of oestrogen to the ER. b) As E_2 diffuses into the cells, heat shock proteins (HSP) dissociate from the receptors and allow E_2 binding. E_2 binds to the ER, which dimerises to attach to EREs on the genome. Activation or repression of ER target genes depend on where co-activator or co-repressors bind to the complex. Created using Biorender.com, based on Osborne & Schiff (2011).

Endocrine therapies are based on targeting and blocking ER signalling. The pathways are targeted via reducing oestrogen levels, antagonising ER function and down-regulating ER levels (Schiff *et al*., 2003). Selective oestrogen receptor modulators (SERMs) such as Tamoxifen interfere with ER signalling. It binds to ER, inhibiting E_2 binding and preventing the recruitment of co-activator molecules, leading to

a reduction in the expression of ER regulated genes (Williams & Lin, 2013). Selective oestrogen receptor downregulators (SERDs) such as Fulvestrant, competitively bind to the ER and instigate degradation by activating the ubiquitin-proteosome pathway, reducing ER levels in the cell (Long & Nephew, 2006). Tamoxifen has been demonstrated to cause a 40-50% reduction in the annual recurrence of breast cancer and reduced the incidence of secondary tumours. However, 50% of breast cancers fail to respond to tamoxifen, and even patients who initially respond can acquire resistance (Normanno *et al*., 2005). For example, according to Osborne & Schiff (2011) ER loss occurs in approximately 20% of patients which causes endocrine resistance. These tumours are no longer driven by oestrogen and instead can be dependent on escape pathways that take over (e.g., upregulation of HER2).

1.2.2. HER2 overexpression

HER2 is a membrane tyrosine kinase. Overexpression of HER2 accounts for approximately 15-30% of breast cancers (Burstein, 2005). Overexpression of HER2 can be caused by increased amplification of the HER2 gene locus or due to dysregulation of HER2 gene expression (Arteaga *et al*., 2012). Cheang *et al.,* (2009) study showed that HER2+ tumours (n=222) are more aggressive (69.8% of tumours were graded at stage III) and carry a poor prognosis with 36% of patients relapsing within 10 years. HER2+ cancers are treated with targeted endocrine therapy such as Trastuzumab (a monoclonal antibody that binds to HER2, preventing proliferation in the cancer cells) Patients (n=1073) that received standard therapy (doxorubicin and cyclophosphamide followed by docetaxel every 3 weeks) had a 5 year disease-free survival rate of 75% and an overall survival of 87%, while patients who received trastuzumab with standard therapy (n=1074) had a 5 year disease-free survival rate of 84% and an overall survival of 91% (Slamon *et al*., 2011).

1.2.3. Triple-negative

Triple-negative breast cancer (TNBC) is defined as tumours that lack expression of ER, PR and HER2. Basal-like breast cancer (BLBC) is one of the intrinsic subtypes of breast cancer (Perou *et al*., 2000), classified by its gene expression profile which is similar to myoepithelial cells (Cheang *et al*., 2008;

Carey *et al*., 2010; Badowska-Kozakiewicz & Budzik, 2016; Milioli *et al*., 2017). Triple-negative and basal-like cells share many features since the majority of BLBC are triple-negative (around 80%) (Carey *et al*., 2010; Foulkes *et al*., 2010; Badowska-Kozakiewicz & Budzik, 2016). However, there is no single set of markers that specifically defines basal-like tumours and so the terms triple-negative and basal-like have frequently become synonyms for each other, due to the overlap.

TNBC are an aggressive form of breast cancer, making up 15% of all invasive breast cancer cases (Badve *et al*., 2011; Milioli *et al*., 2017). TNBC patients have a poorer prognosis compared to other breast cancer subtypes, with a mortality rate of 40% in the first five years from diagnosis (Dent *et al*., 2007), and with distant metastasis exhibiting in 46% of patients, with the average relapse time of 19-40 months (Yin *et al*., 2020). TNBC are typically identified with a high histological grade, high apoptotic rate, central tumour necrosis (centre of tumour exhibits dead breast cancer cells), hormone-receptor negativity (ER -, PR -, HER2 -), TP53 mutations (~83% of TNBC), EGFR expression, adverse prognosis and are more likely to metastasise to viscera, particularly the lung and brain (Cheang *et al*., 2008; Carey *et al*., 2010; Foulkes *et al*., 2010; Milioli *et al*., 2017). Less than 30% of women with metastatic breast cancer will survive five years after diagnosis, while virtually all women diagnosed with TNBC die due to the disease (Bianchini *et al*., 2016). TNBC is more frequent in younger patients (under 40 years old), women with abdominal obesity, and has a higher prevalence in African Americans (Carey *et al*., 2010; Foulkes *et al*., 2010). 15-20% of all TNBC cases share clinical and pathological features with hereditary BRCA1 related breast cancer. Dysregulation of BRCA1 and other defects in homologous recombination pathways (e.g., P53 and PALB2) has been implicated in the tumorigenesis of some TNBC (O'Shaughnessy *et al*., 2011). Badve *et al* (2011) also state that around 30% of TNBC show a lack of pRB expression, overexpression of p16 and p53, all of which are established tumour suppressor proteins.

Due to the lack of steroid receptors, endocrine therapy is not an option for TNBC. Current treatment for patients includes chemotherapy agents such as anthracycline, taxanes and platinum agents (Hudis

& Gianni, 2011). Patients with TNBC have an elevated response rate to chemotherapy than non-TNBC, with approximately 30-40% of patients in early-stage TNBC achieving pathological complete response (pCR) after treatment (Bianchini *et al*., 2016). Below is an outline of current chemotherapeutic regimens being used and trialled to treat TNBC.

A meta-analysis of different trials comparing chemotherapy agents (cyclophosphamide, methotrexate, and fluorouracil) with anthracycline (an antibody that causes DNA damage) containing regimens suggested fluorouracil with anthracycline-containing regimen was most effective to treat TNBC (Foulkes *et al*., 2010). A randomised phase II study investigated the effects of taxane-based chemotherapy compared to anthracycline based chemotherapy in patients with hormone receptor negative breast cancer (Narui *et al*., 2019). The study compared six cycles of docetaxel (75 mg/m^2) and cyclophosphamide (600 mg/m^2) (TC6) and three cycles of 5-fluorouracil, epirubicin and cyclophosphamide (500 mg/m^2 each) followed by three cycles of docetaxel (100 mg/m^2) (FEC-D), with 103 patients randomly assigned to each treatment (97 completed the study). pCR rate was higher for patients in the FEC-D group (36%) compared to the TC6 (25.5%) (p-value= 0.265), and the FEC-D group had a slightly higher overall clinical response (94%) than the TC6 group (93.6%) (p-value= 0.938). Disease-free survival (DFS) and overall survival (OS) (for TNBC patients) was improved in FEC-D group (p-value= 0.016) compared to TC6 (p-value= 0.034), indicating anthracycline-based chemotherapy may be more efficient when treating TNBC. Anti-EGFR drugs have been considered to treat TNBC, due to TNBC overexpressing EGFR. However, TNBC displays abnormal amounts of PTEN60 which is associated with resistance to anti-EGFR therapies (Foulkes *et al*., 2010). Atezolizumab is a type of immunotherapy drug which selectively targets tumours expressing PD-L1 to prevent interaction with the receptors and reversing T-cell suppression (Schmid *et al*., 2018). Results from a phase III trial (n=451) showed that atezolizumab plus nab-paclitaxel led to a longer progression-free survival (by 2.5 months, P<0.001) among patients with metastatic TNBC and has been approved to treat PD-L1 positive TNBC patients (Schmid *et al*., 2018).

Poly(adenosine diphosphate-ribose) polymerase (PARPs) are a family of nuclear enzymes involved in the detection and repair of DNA damage (Herceg & Wang, 2001). PARP inhibitors have been shown to enhance the effects of a range of DNA-damaging agents, resulting in double-strand breaks in replicating cells (O'Shaughnessy *et al*., 2011). PARP inhibitor Olaparib's overall patient response with *BRCA*-deficient metastatic breast cancer ranged from 22% (200 mg) to 41% (800 mg) (Bianchini *et al*., 2016).

Despite the continued development of chemotherapeutic regimens to treat TNBC, patients diagnosed with this breast cancer subtype still have a poor prognosis with a relative lack of targeted treatment options. Due to this, new therapeutic options are required to treat this deadly subtype of breast cancer.

2. Epigenetics

Epigenetics was introduced by Conrad Waddington in 1942 and was defined as the interactions between genes and their products that allow for phenotypic expression (Waddington, 2012). The term has since been redefined with different variations and, is agreed that epigenetics is the chemical modification of DNA and histone proteins, which regulate gene expression. Some of the central epigenetic processes are DNA methylation and chromatin remodelling via post-translational modifications (PTMs) of histone proteins.

DNA is tightly wrapped around a histone core, which is comprised of eight histones (2x H2A, 2x H2B, 2x H3 and 2x H4) see Figure 2.1. Approximately 145-147 base pairs (bp) of DNA are wound around the core, with a linker histone (H1) bound to provide stability, thereby forming the nucleosome (Lawrence *et al*., 2015). Basic histone amino N-terminal tails protrude out of each nucleosome, which can directly contact adjacent ones (Tropberger & Schneider, 2013). These tails can undergo PTMs which affect the direct interaction between the nucleosomes and the structure of the chromatin, allowing accessibility for transcriptional machinery (Figure 2.1). Euchromatin refers to the genome when it is transcriptionally active, with open regions which allows transcriptional machinery such as RNA

polymerase II to transcribe the expression of genes within these regions. Heterochromatin describes the genome when the chromatin is in a compact state, preventing transcriptional machinery to access these regions, which prevents transcription. These phases of chromatin structure are interchangeable, with chromatin modulation proteins determining the function of PTMs, blocking, or activating downstream signalling or influencing recruitment of transcriptional factors and chromatin modifiers (Badeaux & Shi, 2013).

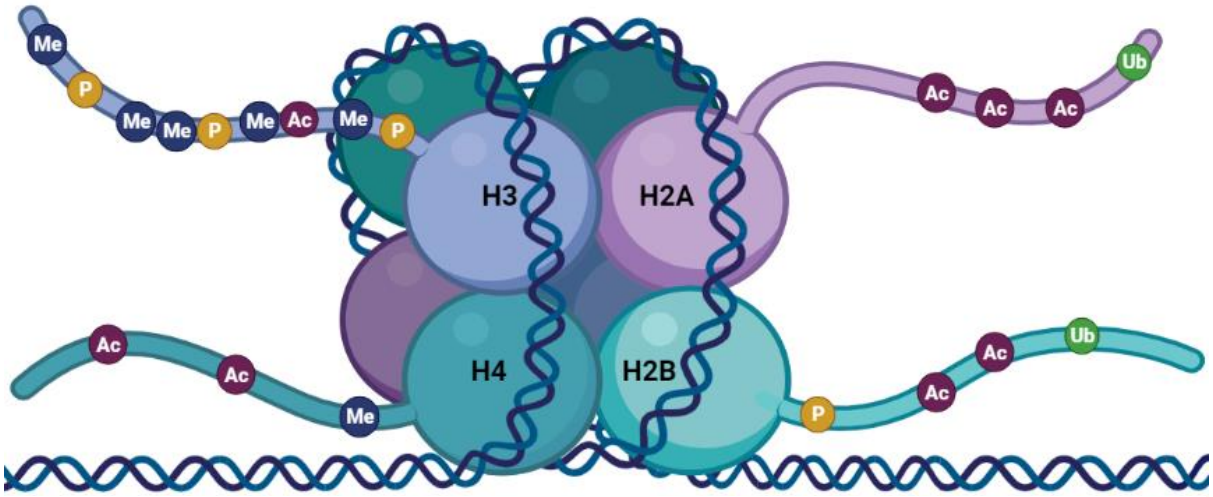

Figure 2.1. Nucleosome structure.
Nucleosomal histone core, containing two copies of H2A, H2B, H3 and H4 (only one of each shown for clarity), with histone tails protruding from each that can be modified by chemical groups being added or removed from specific amino acids (Me = methylation, P = phosphorylation, Ac = acetylation and Ub = ubiquitination). Adapted from Simó-Riudalbas & Esteller (2015). Created using Biorender.com.

2.1. PTMs

Gene expression profiles are regulated by the PTMs of histone proteins. These influence key cellular events including DNA damage response, gene expression, cell cycle, cell signalling pathways and metabolic pathways (Shanmugam *et al*., 2018). Various disease states, such as cancer, can occur due to disruption in proper maintenance of epigenetic mechanisms resulting in activation or inhibition of critical pathways (Khan *et al*., 2015). Examples of histone PTMs include phosphorylation, acetylation, methylation, ubiquitination, and poly ADP ribosylations. PTMs can occur to different degrees (e.g., monomethyl, dimethyl) while some chemical modifications, such as ubiquitin, can be modified themselves (Macek *et al*., 2019).

The mechanism behind key regulations of processes by histone PTMs can be generalised into two categories: 1) addition of PTM on histone proteins, which affect the nucleosomal interactions and their binding to DNA, 2) addition of PTMs to histone protein which inhibits or facilitates binding of various proteins to chromatin (Khan *et al*., 2015). An example of the first mechanism is histone acetylation on H4K16 which inhibits interactions between the H4 tail and the acidic patch (on the surface of the nucleosome) preventing the formation of fibres and chromatin structures, which shifts the chromatin and nucleosome, creating a more open structure (Dhar *et al*., 2017). An example of the second mechanism involves histone "readers" which bind to specific modifications and affect nuclear processes, such as transcription (Khan *et al*., 2015). Histone modification can also regulate and effect one another by crosstalk. For example, CREB binding protein can trigger methylation of H3R17 (via H3K18 and H3K23 acetylation) allowing activation of oestrogen-responsive gene expression (Khan *et al*., 2015; Zhao & Garcia, 2015).

Histone PTMs are added and removed by enzymes called "writers" and "erasers". Epigenetic writer enzymes include histone acetylases (HATs), histone methyltransferase (HMTs) and ubiquitin ligases and erasers include histone lysine/arginine demethylases and histone deacetylases (HDACs) (Simó-Riudalbas & Esteller, 2015). HATs can act as both oncogenes and tumour suppressors, and influence chromatin state by utilising acetyl coenzyme A (CoA) and catalysing the addition of acetyl groups to lysines on histone tails, thereby neutralising the positive charge of the lysine (Bannister & Kouzarides, 2011; Audia & Campbell, 2016). As the overall positive charge of the histone tail is reduced, it decreases the attraction between itself and negatively charged DNA, loosening the chromatin structure, and creating euchromatic regions. Different pro-oncogenic or tumour suppressor outcomes occur via histone acetylation, which is dependent on the genes whose expression is increased by HAT activity. The modification of chromatin structure is dependent on which chemical modification occurs (e.g., methylation, acetylation etc.) and the site where the modification arises on the histone tail. Table 1 shows a range of specific histone modifications, the proteins which are known to regulate their addition/removal and the transcriptional outcome associated with these modifications.

Table 1. Examples of Histone PTMs, the proteins responsible for adding (Writer protein) and removing the PTM (Eraser protein), and the gene expression regulatory function of the PTM.

Histone PTMs	Writer protein	Eraser protein	Function
H2AK119ub	RING1A, RING1B	BAP1, USP16, USP21, 2A-DUB, USP3, USP22	Transcriptional repression
H2BK120ub	RNF20, RNF40	USP3, USP7, USP22	Transcriptional activation, DNA damage response and meiosis
H3K4me	SETD1A SETD1B, ASH1L, MLL, MLL2, MLL3, MLL4, SETD7	KDM1A, KDM1B, KDM5B, NO66	Transcriptional activation
H3K4me2	SETD1A, SETD1B, MLL, MLL2, MLL3, MLL4, SMYD3	KDM1A, KDM1B, KDM5A, KDM5B, KDM5C, KDM5D, NO66	Transcriptional activation
H3K4me3	SETD1A, SETD1B, ASH1L, MLL, MLL2, MLL3, MLL4, SMYD3, PRMD9	KDM2B, KDM5A, KDM5B, KDM5C, KDM5D, NO66	Transcriptional activation
H3K9me	SETDB1, G9a, EHMT1, PRDM2	KDM3A, KDM3B, PHF8, JHDM1D	Transcriptional activation
H3K9me2	SUV39H1, SUV39H2, SETDB1, G9a, EHMT1, PRDM2	KDM3A, KDM3B§, KDM4A, KDM4B, KDM4C, KDM4D, PHF8, KDM1A, JHDM1D	Transcriptional repression, X-inactivation and imprinting
H3K9me3	SUV39H1, SUV39H2, SETDB1, PRDM2	KDM3B, KDM4A, KDM4B, KDM4C, KDM4D	Transcriptional repression
H3S10p	Aurora B, MSK/RSK/Jil-1	PP2A, PP1	Transcriptional activation, mitosis and meiosis
H3K27ac	CBP/p300	HDACs	Transcriptional activation
H3K27me	EZH1, EZH2	JHDM1D	Transcriptional activation
H3K27me2	EZH1, EZH2	KDM6A, KDM6B, KDM7A, PHF8	Transcriptional silencing, X-inactivation and bivalent genes
H3K27me3	EZH1, EZH2	KDM6A, KDM6B, KDM7A, PHF8	Transcriptional silencing, X-inactivation and bivalent genes
H3K36me	NSD1, NSD2, NSD3, SETD2, SETD3, ASH1L, SETMAR, SMYD2	KDM2A, KDM2B, KDM4A, KDM4B, KDM4C, KDM4D, JHDM1A	Transcriptional elongation, transcriptional repression and DNA repair
H3K36me2	NSD1, NSD2, NSD3, SETD2, SETD3, ASH1L, SETMAR, SMYD2	KDM2A, KDM2B, KDM4A, KDM4B, KDM4C, KDM4D, JHDM1A	Transcriptional elongation, transcriptional repression and DNA repair
H3K36me3	NSD1, NSD2, NSD3, SETD2, SETD3, ASH1L, SETMAR, SMYD2	KDM2A, KDM2B, KDM4A, KDM4B, KDM4C, KDM4D, JHDM1A	Transcriptional elongation, transcriptional repression and DNA repair
H3K79me	DOT1L	N/A	Transcriptional activation
H3K79me2	DOT1L	N/A	Transcriptional activation
H3K79me3	DOT1L	N/A	Transcriptional activation
H4K20me	PR-Set7	LSD1n	Transcriptional activation

H4K20me2	SUV4-20H1, SUV4-20H2	LSD1n, DPY-21	Transcriptional silencing and heterochromatin
H4K20me3	SUV4-20H1, SUV4-20H2	LSD1n, DPY-21	Transcriptional silencing and heterochromatin
H4K16ac	MOF	HDACs, Sirt2	Transcriptional activation and DNA repair
H2AK5ac	Tip60, p300, Hat1	N/A	Transcriptional activation
H2AK7ac	Hat1, Esa1	N/A	Transcriptional activation
H2AK13ub	Rnf168	N/A	DNA damage response
H2AK15ub	Rnf168	N/A	DNA damage response
H2AK63ub	Rnf8	N/A	DNA damage response
H2AK119ub	dRing, RING1B	N/A	Polycomb silencing and UV damage response
H2BK5ac		N/A	Transcriptional activation
H3K4ac	GCN5, RTT109, Sir2, Hst1	N/A	Transcriptional activation

Adapted from Zhao & Garcia, (2015), Alaskhar Alhamwe *et al*., (2018) and Zhao & Shilatifard, (2019).

Various diseases develop due to atypical histone modifications. PTMs have a known role in cancer progression. Audia & Campbell (2016) outline that genetic or epigenetic abnormalities that affect HAT and HDAC expression have been detected in a variety of cancers, which may not be limited to the histones, but act on proteins involved in tumour migration, metastasis, and growth. Khan *et al*., (2015) discovered atypical gene silencing in different cancers was due to alterations in H3K9 and H3K27 methylation patterns. Epigenetic silencing of *BRAC1* is observed in sporadic breast cancer and *c-Myc* is deregulated and overexpressed in multiple cancer types via dysregulation of epigenetic mechanisms at the *MYC* locus (Shanmugam *et al*., 2018). PTMs have mostly been studied as potential prognostic markers. Mouse models of multistage skin carcinogenesis showed that a loss of H4K16ac and H4K20me3 was present in cancerous tissue compared to normal tissue, and were associated with DNA hypomethylation, allowing them to be a common hallmark of tumour cells (Fraga *et al*., 2005). However, modulation of epigenetic regulatory proteins dysregulated in cancer is a growing field in cancer therapeutics. For example, KDMs have emerged as a potential therapeutic target, demonstrated by studies which showed knockdown of KDM3A and KDM4B (both established ER histone demethylase co-factors) together reduced proliferation in breast cancer cells more than either

on their own (Gaughan *et al*., 2013; Wade *et al*., 2015; Jones *et al*., 2019), but which did not affect non-transformed breast cell growth.

2.2. PRC1

Polycomb group complexes (PcGs) are essential regulators of developmental gene expression, including proliferation, stem-cell self-renewal and cell differentiation (Kundu *et al*., 2017; Zhu *et al*., 2018a; Blackledge *et al*., 2019). PcGs functions by forming chromatin-associated protein complexes and repressing target gene expression. PcGs belong to one of two repressive complexes: PRC1 and PRC2. These complexes are major epigenetic regulators for the monoubiquitination of histone H2A (Zhu *et al*., 2018).

PRC2 contains the core components EZH1/EZH2, Suz12 and EED (Embryonic Ectoderm Development, which is a polycomb group protein) (Yu *et al*., 2012; Huang & Hornyak, 2015; Zhu *et al*., 2018), while multiple variations of PRC1 exists (canonical vs non-canonical). Variations of PRC1 are dictated by the homologs for PRC1 components and the presence or absence of CBX proteins (Gil & O'Loghlen, 2014). Canonical PRC1 is composed of an HPH family member (HPH1-HPH3), a PCGF family member (PCGF1-PCGF6), a chromobox domain proteins (CBX2,4,6,7,8) and a RING1 protein (RING1a or RING1b) (Yu *et al*., 2012; Huang & Hornyak, 2015; Zhu *et al*., 2018). Non-canonical (aka variant) PRC1 can utilise all six PCGFs but RYBP/YAF2 in place of a CBX protein (Blackledge *et al*., 2019; Fursova *et al*., 2019).

PRC2 is involved in gene silencing through the recruitment of PRC1. PRC2 tri-methylates lysine 27 (via the EZH1/EZH2 component) on the amino acid tail of histone H3 (H3K27me3) (Scelfo *et al*., 2015). This methylation mark is recognised by the CBX protein within canonical PRC1. The chromodomain within the CBX protein recognises and interacts with the methyl group and therefore recruits the PRC1 complex to regions of chromatin marked by H3K27me3 (Gil & O'Loghlen, 2014; Pherson *et al*., 2017). Lysine 119 on the histone tail of H2A is then ubiquitinated by the RING protein of the PRC1 complex, which condenses the chromatin structure and causes transcriptional silencing of the genes within these regions of chromatin.

Non-canonical PRC1 recruitment to the target genes relies on DNA-binding activities, for example, PCGF6-PRC1 associates with KDM2B which recognises CGIMGA/MAX DNA-binding factors (Blackledge *et al*., 2019). Non-canonical PRC1 complexes may also play a role in gene repression, despite not having a CBX component. RYBP-PRC1 complexes can mediate H2AK119Ub independently of H3K27me3, which is essential for global maintenance of embryonic stem cells (ESC) (Gil & O'Loghlen, 2014). PRC1 has recently been implicated in transcriptional activation as well as PRC1-independent roles in ESC and somatic cells (Gil & O'Loghlen, 2014). Findings indicate PRC1 influences the transcription of active genes and may suppress transcription of non-promoter regulatory sequences, although the mechanisms behind this are unclear (Pherson *et al*., 2017). Importantly, it is also known that different compositions of PRC1 regulate different gene targets and that PRC1 composition is also cell type specific (Morey *et al*., 2015; Chan *et al*., 2018).

PcGs have been implicated as both oncogenic and tumour suppressive complexes. An example of PRC1 oncogenic activity is the fact it may promote cancer progression, by repressing INK1a/ARF expression, a well-characterised tumour suppressor gene and cell cycle regulator, there by inhibiting senescence (Gil & O'Loghlen, 2014). Individual components of the PcG complexes have also been implicated in cancer. For example, upregulation of PRC1 gene PCGF4 (also known as *BMI1*) is associated with reduced apoptosis and increased proliferation and stem cell self-renewal (Barbour *et al*., 2020). In myelodysplastic syndrome (MDS) and leukaemia patients, inactivating mutations found in EED and other PRC2 gene loci have been implicated to play a role in the progression of these diseases, and PRC2 activity is essential in the development of acute myeloid leukaemia (Scelfo *et al*., 2015). PcGs inhibition has also become a potential therapeutic target. EZH2 inhibitors are in a trial as a single treatment for lymphoma, due to EZH2 frequently mutating in large B cell lymphomas and follicular lymphomas (Scelfo *et al*., 2015).

2.2.1. CBX2

CBX proteins are epi-reader proteins and a component of the PRC1 complex. Eight different CBX proteins regulate heterochromatin by the N-terminal chromodomain, which recognise histone

modifications, specifically H3K27me3 (Nichol *et al*., 2016). However, only five of these proteins (CBX2, CBX4, CBX6, CBX7 and CBX8) have a c-terminal Polycomb repressor box which makes them a part of the PRC1 complex, and therefore involved in transcriptional silencing (Clermont *et al*., 2014; Di Costanzo *et al*., 2018).

CBX proteins play a role in tumorigenesis and stem cell differentiation. Klauke *et al*., (2013) showed that overexpression of certain CBX proteins has an effect on the cell cycle. Self-renewal (the ability to proliferate and maintain multipotency (Fuchs & Chen, 2013; Shenghui *et al*., 2009)) was measured by colony-forming cell (CFC) assay in haematopoietic stem cells, with overexpression of CBX7 increasing colony-forming ability, while CBX2, CBX4 and CBX8 had diminished self-renewal, due to a lower number of secondary colonies. CBX2 decreased the proliferation of secondary colonies and has a knock-on effect with stem cell self-renewal (Klauke *et al*., 2013). The article also showed that CBX proteins affected haematopoietic stem and progenitor cells (HSPCs) repopulating *in vivo*. Overexpression of CBX2, CBX4 or CBX8 resulted in exhaustion (over-proliferation causing damage, reducing proliferation potential, and increasing tissue atrophy and premature ageing (Fuchs & Chen, 2013)) of HSPCs, due to not contributing to long-term haematopoietic reconstitution. This was validated by a CFC assay and investigated re-plating efficiency (van den Boom *et al*., 2013). The study showed via a limiting-dilution long-term culture initiating cell (LTC-IC) assay (hematopoietic stem cells are measured on the ability to generate myeloid progeny (Liu *et al*., 2012)) that CBX2 knockdown caused a prominent ~5-fold reduction of LTC-IC frequency. This indicates CBX2 expression controls HSPCs compartments (the hierarchy of regulated cells that maintain and function the blood systems and bone marrow. Overall, CBX2 was shown to affect not only the rate of proliferation but also the differentiation of stem cells. CBX proteins also have an essential role in both the cell cycle and tumorigenesis, with certain members of the protein family, such as CBX2, being identified to have important roles in tumour cell growth and survival. Clermont *et al*., (2014) did a meta-analysis investigating a potential oncogenic role of CBX2 within cancers. Using the Oncomine database analysis, twenty-five studies (3848 patients) were identified showing significant upregulation of CBX2

in a variety of cancer tissues compared to normal tissues, with the most represented types being colon, breast, stomach, and lung. However, using the same criteria, no downregulation was reported. This indicates CBX2 has an essential role within the body, which is also used to promote tumorigenesis.

The effect of CBX2 to repress tumour suppressive loci was also investigated within the dataset, and it was found that CBX2 did not downregulate p14ARF & p16INK4A (loci for CDKN2A, a tumour suppressor gene). However, in ten of the twenty-five studies, p15INK4B (loci for CDKN2B, a tumour suppressor gene) was downregulated (Clermont *et al*., 2014). Analysis revealed that out of the ten studies that p15INK4B was downregulated in, eight of them specifically occurred in colorectal cancer, but spearman's correlation showed no significant correlation.

Further analysis was done using MANOVA for one cohort of breast, lung, and colorectal cancers from the Oncomine database. Analysis of the colorectal cancer group found CBX2 expression was elevated in females. In the lung, elevated expressions of CBX2 were detected in squamous cell carcinoma compared with lung adenocarcinoma. High CBX2 expression was also correlated with younger age, elevated tumour grade and basal-like subtype in the breast cancer cohort, which is associated with poor prognosis. Clermont *et al*., (2016) also showed CBX2 expression has a strong association with metastatic prostate cancer compared with non-metastatic prostate cancer. A PDX model, using LTL313H (metastatic) and LTL313B (non-metastatic) xenograft tissues (both lines are derived from two biopsies of the same tumour), observed CBX2 was most highly upregulated in LTL313H compared to LTL313B. This was validated via qRT-PCR, which showed CBX2 expression 3.2-fold higher in LTL313H versus LTL313B. CBX2 depletion induced prostate cancer cell death and promoted proliferation arrest in androgen-independent late-stage prostate cancer, meaning CBX2 could be used as a prognostic and therapeutic target in this disease setting (Clermont *et al*., 2016).

CBX2 expression has also been shown to be upregulated in high grade serous ovarian carcinoma (HGSOC). CBX2 expression examined in HGSOC samples (via TCGA database) showed that high expression of CBX2 contributed to both worse DFS (11.7 vs 17.6 months) and OS (34 vs 44.8 months)

(p-values = <0.05), and in an independent HGSOC data set, high CBX2 was correlated with poorer survival at three years (Wheeler *et al*., 2018). Wheeler *et al*., also found via reverse-phase protein array that, in high CBX2 expressing tumours, phosphorylated serine 318 and FOXO3 (a known tumour suppressor) was notably reduced, again suggesting a role for CBX2 in driving tumour progression. CBX2 knockdown in OVCAR4, PEO1 and OVCAR8 HGSOC cell lines reduced the rate of colony formation and proliferation (as measured by Gaussia luciferase (gLuc) activity), in both 2D tissue culture and suspension culture. In the suspension tissue culture, in all cell lines, cell viability significantly decreased after the knockdown of CBX2 and showed that CBX2 knockdown increased anoikis (anchorage-independent cell death). This indicates CBX2 has a significant role in HGSOC cell proliferation and in preventing anoikis.

Due to the majority of HGSOC patients developing chemoresistance for platinum-based chemotherapy, CBX2 expression was compared in carboplatin sensitive HGSOC tumours and platinum-resistant tumours. Results showed there was an increase in CBX2 in platinum-resistant tumours. This, combined with the CBX2 role against anoikis, suggests it reduces chemotherapy response. A gene expression dataset was also used to compare platinum-sensitive HGSOC tumours, to platinum-resistant HGSOC. Results showed CBX2 expression was elevated in platinum-resistant tumours, contributing to the association of cancer progression and more aggressive HGSOC. Due to the evidence of CBX2 being involved in chemoresistance and cancer progression, meaning CBX2 could be a predictive marker for advanced HGSOC. Both distant metastasis and ascites-associated tumour cells were associated with increased CBX2 expression in three out of five patients (matched with primary tumours). CBX2 is therefore a potentially crucial factor in driving HGSOC progression.

CBX2 has also been strongly implicated in breast cancer progression. In a study, 216 paired breast cancer tissues and matched normal tissues immunostaining showed CBX2 expression was higher in breast cancer tissues (81.01%) compared to normal tissues (Chen *et al*., 2017). The protein expression of CBX2 was also higher in breast cancer tumour tissue, than the matched normal (p<0.001). Chen *et*

al., also showed that high cytoplasmic CBX2 expression associated with tumour size ($p<0.001$), lymph node metastasis ($p=0.008$), TNM classification ($p<0.001$) and HER2 positive status ($p=0.048$). Stratified Kaplan-Meier survival curves indicated that high CBX2 was associated with poor prognosis ($p=0.025$). Zheng *et al*., (2019) investigated the impact of CBX2 on breast tumorigenesis. Nude mice were injected with MCF-7 cells which had been transfected with sh-CBX2 (an shRNA construct targeting CBX2 mRNA) or Sh-NC (a negative control shRNA) and tumours were observed for five weeks. Results showed that tumour growth and tumour weight dramatically reduced in the sh-CBX2 group compared to the control (sh-NC group) therefore indicating a role for CBX2 in breast epithelial cell tumourigenesis.

CBX2 has also been implicated in oncogenic signalling pathways in the breast. Zheng *et al*., (2019) showed that CBX2 acts in regulating the cell cycle, DNA replication and that the PI3K/AKT signalling pathway functions downstream of CBX2 in breast cancer tumourigenesis and/or progression. These mirrored findings seen by Clermont *et al* (2016) that showed that CBX2 promoted PI3K/AKT signalling in prostate cancer via downregulating PI3K antagonists (PIK3RI and INPP5A). Correlation analysis of pathway deregulation score (PDS) of signalling hallmarks (from METABRIC and TCGA datasets) showed a strong correlation with CBX2 with mTORC1 signalling in both datasets (Iqbal *et al*., 2021). This shows that not only does CBX2 have a variety of roles within the body, but has a variety of roles in cancer progression, as it affects pathways downstream.

Overall, while the mechanisms of CBX2 are not fully understood, CBX2 has been shown to affect stem cell differentiation and appear to play a crucial role in cancer progression. This indicates that CBX2 could be used as a prognostic marker and a potential therapeutic target, preventing chemoresistance for a variety of cancers.

2.3. Hypothesis and Aims

The hypothesis of this project is that CBX2 plays a significant role in TNBC progression and that it could be a potential therapeutic target for this hard-to-treat disease.

Aims:

1. To investigate the expression of *CBX2* in different cancers and to determine biological processes/signalling pathways potentially regulated by CBX2 via gene set enrichment analysis (GSEA) of publicly available patient datasets.

2. To analyse previously generated RNA-sequencing data from MDA-MB-231 cells (a TNBC cell line which is depleted of CBX2 expression) and compare this with GSEA of patient datasets (Aim 1).

3. To knockdown CBX2 in breast cancer cell lines and use chromatin immunoprecipitation and CUTandRUN techniques to determine the direct regulatory role of CBX2 on identified tumour suppressor genes putatively regulated by CBX2 (tumour suppressor gene identified in Aim 2).

Overall, this study aims to identify which oncogenic processes and pathways CBX2 may regulate in TNBC and to elucidate potential mechanisms by which CBX2 influences these processes. Please note that the aims of this project have been modified due to the COVID-19 pandemic which began part way through the degree.

2.3.1. Ethical Statement

This project uses cultured human cancer cell lines (MDA-MB-231 and MDA-MB-468) purchased from American Type Culture Collection (ATCC, USA), which are sourced in accordance with ethical guidelines, with the cell donor remaining anonymous for the duration of the project and onwards. Control of Substances Hazardous to Health (COSHH) forms were completed before all experiments involving these cell lines. Analysis of patient datasets are from publicly available gene expression databases in which samples have been collected with ethical approval and informed consent and which are anonymised to prevent the identification of individuals.

3. Materials and methods

General materials, such as plastic consumables, were purchased from Sarstedt, Germany. General chemicals were purchased from Fisher Scientific, UK unless otherwise stated.

3.1. GEPIA2

Comparative expression of *CBX2* in tumour and normal tissues was analysed using the GEPIA2 (Gene Expression Profiling Interactive Analysis 2) online database (http://gepia2.cancer-pku.cn/) (Tang *et al*., 2019). GEPIA2 was used for gene expression analysis between tumour vs normal tissue to produce boxplots comparing expression between different subtypes of breast cancer and other cancers based on The Cancer Genome Atlas (TCGA) and GTEx datasets, as well as survival curves including OS and DFS. CBX2 gene expression was examined across different datasets (Breast invasive carcinoma, glioblastoma multiforme, lung adenocarcinoma, ovarian serous cystadenocarcinoma, prostate adenocarcinoma, and testicular germ cell tumours) and different stages of cancer within the datasets. Gene expression data in the form of transcripts per million (TPM) was log-transformed to allow a more standardised comparison between data points (log2(TPM + 1)). Parameters for gene expression comparison were using the default settings: statistical analysis method for differential gene expression analysis was one-way ANOVA, expression data were first log2(TPM+1) transformed for analysis and the Log2FC cut off between samples for statistical analysis was defined as median(Tumour) – median(Normal) = 1. The p-value cut off for defining a statistical difference between cohorts was defined as <0.01 (Tang *et al*., 2019). Survival analysis (OS and DFS) parameters were also set at default settings. The Mantel-cox test was used for hypothesis testing and the cox proportional hazard model was used to determine the hazard ratio (HR) (Tang *et al*., 2019). Analysis was again based on TCGA and GTEx databases. Patients were grouped as having a high or low expression of CBX2 with the median expression used as the cut-off.

3.2. cBio Cancer Genomics Portal

The cBio Cancer Genomics Portal (cBioportal) is an open-access resource allowing for the exploration, visualisation, and analysis of multidimensional cancer genomic data sets (http://cbioportal.org)

(Cerami *et al*., 2012; Gao *et al*., 2013). This software was used to identify genes significantly positively and negatively correlated with CBX2 expression in chosen datasets. CBX2 was queried in the chosen datasets, with mutations, copy-number alterations (CNA) and mRNA expression parameters selected. Only complete samples were selected (samples with mutation, DNA CNA and expression data). The query was submitted, and the co-expression data was log-transformed and downloaded. Datasets downloaded were Breast Cancer (METABRIC, Nature 2012 & Nat Commun 2016) (Pereira *et al*., 2016), Breast Invasive Cancercinoma (TCGA, PanCancer Atlas), Lung Adenocarcinoma (TCGA, PanCancer Atlas) and Glioblastoma Multiforme (TCGA, PanCancer Atlas) (Hoadley *et al*., 2018). Those labelled TCGA, PanCancer Atlas here are based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga. All downloaded correlation analyses were ranked via spearman's correlation value (most positive to most negative). Genes with a spearman's correlation p-value over 0.05 were removed from the ranked list.

3.3. Gene Set Enrichment Analysis

GSEA was used to analyse the ranked list of positively and negatively correlated genes generated using cBioPortal (Mootha *et al*., 2003; Subramanian *et al*., 2005). Pre-ranked co-expression data obtained from cBioportal co-expression analysis (Section 3.2) was compared against the GSEA hallmark gene sets (H), ontology gene sets (C5) and oncogenic signature gene sets (C6) to identify genesets that were enriched with genes that either positively or negatively correlate with CBX2 expression (Ashburner *et al*., 2000; Edgar *et al*., 2002; Barrett *et al*., 2012; Liberzon *et al*., 2015; The Gene Ontology Consortium *et al*., 2019). GSEA gene sets were from the molecular signature database (v7.1 MSigDB) which contains 25,724 gene sets for use with GSEA. Nominal p-value, false discovery rate (FDR) q-value and normalised enrichment score (NES) were obtained from the analyses.

3.4. Tissue Culture

Two ER+ breast cancer cell lines (MCF-7 and T47D) and two TNBC cell lines (MDA-MB-231 and MDA-MB-468) were used in the study. These cell lines were purchased from American Type Culture Collection (ATCC, USA), and were cultured in RPMI medium 1640 (Gibco, UK) containing 10% foetal

bovine serum (Biowest, USA) and 1% penicillin/streptomycin (Lonza, UK). MDA-MB-468 were cultured in high glucose Dulbecco's Modified Eagle Medium (DMEM) (Gibco, UK) which included 10% foetal bovine serum and 1% penicillin/streptomycin. Cells were incubated at 37°C with a 5% CO_2 atmosphere.

For standard culture, cells were grown in T75 flasks or T175 flasks supplemented with 15 ml or 25 ml of media respectively, which was changed every 3 days. Cells were trypsinised when approximately 70% confluent. Trypsin (Lonza, UK) was diluted at 1:10 (for the MDA-MB-468 cell line) and 1:30 (for the other cell lines) in phosphate buffer saline (PBS) solution. The solution was made from 1x PBS tablets dissolved in 200 ml of deionised water, sterilised by an autoclave. During cell sub-culturing, media was removed, and flasks were washed with 5 ml PBS and then 3 ml (for T75) or 5 ml (for T175) of diluted trypsin was added and incubated at 37°C for 3-5 minutes to allow the cells to detach. Once detached, trypsin was neutralised by adding 12 ml or 15 ml of culture media to T75 and T175 flasks, respectively. Cells were centrifuged (Centrifuge 5702, Eppendorf tubes, UK) at 400 g for 3 minutes and then media aspirated without disturbing the cell pellet. Cells were resuspended in 5 ml of fresh media and mixed by pipetting up and down to remove clumps. The appropriate volume of cell suspension was added to the new flasks for continued culture. All cell culture was performed under a class II Biohazard safety cabinet (ESCOglobal, UK).

3.4.1. siRNA transfections

Cell lines were transfected with siRNA duplexes using RNAiMAX lipofectamine (Thermo Fisher Scientific, UK) following manufacturer instructions. Transfections were performed using a non-silencing siRNA control (siScr) or individual siRNA sequences (see Table 2) targeted against *CBX2* (Sigma, UK). For transfections in 6 well plates, a transfection mix of 1 µl of siRNA (50 µM stock), 100 µl basal media and 2 µl of RNAiMax lipofectamine was used per well. Cells were incubated in 2 ml of culture media per well, meaning a final siRNA concentration of 25 nM per well. For 150 mm dishes, a transfection mix of 10 µl of SiRNA, 500 µl basal media and 20 µl RNAiMax lipofectamine was used per plate at a final concentration of 25 nM (20 ml of culture media per plate). siRNA transfection mixes were added to the centre of the well or plate and left at room temperature for 20 minutes. Cells were

trypsinised (as described above) and following resuspension were counted three times using a single chamber haemocytometer to determine the correct cell concentration. Cells were plated directly onto the siRNA transfection mixtures. In 6 wells plates, 150,000 cells were plated per well. In 150 mm dishes, 5,000,000 cells were plated per dish. 6 well plates were used for protein lysis for western blot analysis and cell counts. Cells were incubated for 72-96 hours, depending on the experiment. 150 mm dishes were used for chromatin immunoprecipitation (ChIP) Cells were incubated for 48-72 hours, depending on the experiment.

Table 2. siRNA sequences.

siRNA	**Sequence 5'-3'**
siScr - non-silencing siRNA	UUCUCCGAACGUGUCACGU (Horizon Discovery, UK)
si*CBX2* #1	AGGAGGUGCAGAACCGGAA (Sigma, UK)
si*CBX2* #2	GCAAGGGCAAGCUGGAGUA (Sigma, UK)
si*CBX2* #3	CAAGGAAGCUCACUGCCAU (Sigma, UK)

3.4.2. Protein harvesting via SDS lysis buffer

Cells were transfected in 6-well plates (as described in Section 3.4.1) and incubated for 72 hours. Following incubation, approximate cell confluence was observed by microscopy to determine the amount of sodium dodecyl sulphate (SDS) lysis buffer (125mM Tris-HCl pH 6.8, 10% Glycerol, 2% (w/v) SDS) to add to each well. These were adjusted in accordance with the control (siScr) transfected cells. For example, 200 µl SDS lysis buffer was added to cells transfected with siScr, whereas cells with 50% confluency compared to the siScr sample would have 100 µl of SDS lysis buffer added. Media was aspirated, wells washed with 1 ml PBS before SDS lysis buffer was added. Cells were incubated at room temperature for 5 minutes on a rocker. Lysates were then transferred to fresh tubes for western blot analysis or stored at -20^{o}C.

3.5. Western blotting

3.5.1. Gel electrophoresis/SDS-PAGE

In a vertical gel cast, approximately 7.5 ml of 10% acrylamide composition separating gel (Table 3) was added. To ensure the gel set evenly, 100 µl isopropanol (VWR, UK) was pipetted above the gel line. After the gel polymerised, the isopropanol was removed. Above the separating gel, a 5% acrylamide

composition stacking gel (Table 4) was added. A lane comb (comprising of 10 or 15 wells) was positioned on top of the stacking gel. After polymerisation, the gels and 5x Laemli buffer (Table 5) were placed into an electrophoresis tank (BioRad, UK). Protein lysates (protein extraction described above) were boiled on a heat block at 100^{o}C for 10 minutes, before being centrifuged and mixed to make homogenous. Protein lysate was loaded into the wells (10 µl for 15 well comb and 20 µl for a 10 well comb). Spectra B Multicolour broad range protein ladder (Thermo Fisher Scientific, UK) was loaded in at 3 µl or 5 µl, respectively. Gels were run at 140 V until all the dye from the lysis buffer had run off the bottom of the gel. All experiments were prepared in the same way.

Table 3. Composition of 10% acrylamide separating gels.

Separating gel components	**Amount for 10ml mixture**
DiH_2O	1.67 ml
Stacking buffer 1.5 M Tris-HCl with SDS (Tris 18.16 g/L, SDS 0.4 g/L, pH 8.8)	5 ml
30% acrylamide (Sigma, UK)	3.3 ml
10% Ammonium persulphate (APS) solution	100 µl
N,N,N',N'-tetramethylethane-1,2-diamine (TEMED) (Sigma, UK)	12 µl

Table 4. Composition of 5% acrylamide stacking gel

Stacking gel components	**Amount for 5ml mixture**
DiH_2O	1.67 ml
Stacking buffer 1.5 M Tris-HCl with SDS (Tris 18.16 g/L, SDS 0.4 g/L, pH 8.8)	2.5 ml
30% acrylamide (Sigma, UK)	830 µl
10% Ammonium persulphate (APS) solution	50 µl
N,N,N',N'-tetramethylethane-1,2-diamine (TEMED) (Sigma, UK)	7.5 µl

Table 5. Composition of 5x Laemli buffer.

5x Laemli buffer in 1L of DiH_2O	**Amount**
Tris-HCl	30 g
Glycine	144 g
SDS	5 g

3.5.2. Gel transfer

After gel electrophoresis, the separated proteins were transferred onto a nitrocellulose blotting membrane (GE Healthcare Life Sciences, UK). A transfer cassette was constructed in the following order: two pieces of Whatman paper (GE Healthcare Life Sciences, UK), acrylamide gel, nitrocellulose blotting membrane, two pieces of Whatman paper, sponge. This was assembled from anodes to cathodes, with a roller being used to prevent bubbles after the membrane was placed on the gel. All components were bathed in 1x transfer buffer, which was made up of 100 ml of 10x transfer buffer (Table 6), 100 ml methanol (VWR, UK), 800 ml DiH2O. The cassette was then placed into a gel transfer tank (BioRad, UK) filled with 1x transfer buffer which was placed above a magnetic stirrer (VELP Scientifica, Italy). A small magnet and ice pack was added to prevent overheating. The transfer occurred at 100 V for 1 hour.

Table 6. Composition of 10x transfer buffer.

Transfer buffer in 1L of DiH_2O	**Amount**
Tris	24.2 g
Glycine	111.75 g
SDS	10 g

3.5.3. Protein visualisation

Once protein transfer was completed, the membrane was blocked for 1 hour on a rocker (PM130, Grant-bio, UK) in a 5% (w/v) milk solution created from dried milk powder (Marvel, UK) mixed with 1x Tris-buffered saline (TBS) (Table 7) with 0.1% TWEEN (Fisher Scientific, UK) (TBST). 1x TBST is made up 900 ml DiH2O, 100 ml 10x TBS and 1 ml TWEEN. After the blocking, the membrane was incubated overnight at 4^{o}C on a roller mixer (Grant-bio, UK) in primary antibody (Table 8), diluted in 5% milk solution (made with TBST). Following primary incubation, the membrane was washed three times in TBST for 10 minutes each. Incubation in a secondary antibody (Table 8), diluted in 5% milk solution (made with TBST), took place for one hour at room temperature on a roller mixer. Afterwards, the membrane was washed three times again for 10 minutes, then once in TBS for 5 minutes. Clarity™ western enhanced chemiluminescence ECL substrate (Bio-Rad, UK) was mixed in the ratio 1:1:2 of

substrate one, two and 1x TBS, respectively. 1 ml of this mixture was added to the membranes. Membranes were visualised using a ChemiDoc (Bio-Rad, UK).

Table 7. Composition of 10x Tris-buffered saline solution.

10x TBS in 1L of DiH_2O	**Amount**
Tris-HCl	24.2 g
NaCl (VWR, UK)	8.8 g

Table 8. Primary and Secondary antibody concentrations when analysed by Western blot.

Protein	**Dilution of primary antibody**	**Dilution of secondary antibody**
Alpha tubulin	1:20,000 alpha tubulin antibody 260.0 µg/150 µL (Catalogue number: 66031-1-1g, Proteintech, UK)	1:20,000 secondary antibody rabbit anti-mouse 1 mg/mL (Abcam, UK)
CBX2	1:5,000 anti-*CBX2* primary antibody 1 mg/mL (Catalogue number: ab80044, Abcam, UK)	1:5,000 Goat anti-rabbit 1 mg/mL (Abcam, UK)

3.6. Chromatin immunoprecipitation (ChIP)

3.6.1. Formaldehyde fixation

Cells were transfected in 150 mm dishes (as described in Section 3.4.1) and incubated for 48 hours. After incubation, fresh 11% formaldehyde solution is made according to Table 9. 2 ml of the formaldehyde solution was added to each plate, and the plate was swirled to ensure it was evenly mixed. The plate was then incubated for 7 minutes at room temperature before 1.1 ml of 2.5 M glycine was added to quench the formaldehyde. The plate was swirled and incubated at room temperature for 5 minutes. The glycine-formaldehyde-media solution was then removed, and the cells were washed in ice-cold sterile PBS twice. 5 ml of PBS was then added, and the cells were scrapped with a cell scraper and transferred to a 15 ml conical tube. The tubes were centrifuged at 2000 RCF at 4^{o}C for 4 minutes. The supernatant was then removed. The cell pellet was either snap-frozen in liquid nitrogen and stored at -80^{o}C at this point before the cell lysis step (Section 3.6.2), or cell lysis was performed immediately.

Table 9. Formaldehyde solution components.

11% formaldehyde solution	Stock solution	Amount (5 plates = 10ml)
50 mM HEPES-KOH	0.5 M	1 ml
100 mM NaCl	4 M	0.25 ml
1 mM EDTA	0.5 M	0.02 ml
0.5 mM EGTA	0.5 M	0.01 ml
11% formaldehyde	37%	3 ml
DiH_2O		5.72 ml

3.6.2. Cell lysis and chromatin extraction

The cell pellet was resuspended in 5 ml of LB1 buffer (Table 10) and placed on a rocker on ice for 10 minutes. The suspension was then centrifuged at 2000 RCF for 5 minutes at 4^{o}C. The supernatant was removed, and the cell pellet was resuspended in 5 ml LB2 buffer (Table 11) and placed on a rocker on ice for 5 minutes. The sample was then centrifuged at 2000 RCF at 4^{o}C for 5 minutes followed by the removal of the supernatant. The cell pellet was resuspended in 300 µl of LB3 (Table 12) and placed into RNase/DNase free Eppendorf tubes. The samples were left on ice for 30 minutes before being stored at -20^{o}C or sonicated (Section 3.6.3).

Table 10. LB1 components.

LB1	Stock solution	Amount (500ml)
50 mM HEPES-KOH	0.5 M	50 ml
140 mM NaCl	4 M	17.5 ml
1 mM EDTA	0.5 M	1 ml
10% glycerol	*Neat*	50 ml
0.5% NP40	*Neat*	2.5 ml
0.25% Triton X-100	*Neat*	1.25 ml
DiH_2O		377.75 ml

Table 11. LB2 components.

LB2	Stock solution	Amount (500ml)
10 mM Tris-HCl, pH 8	1 M	5 ml
200 mM NaCl	4 M	25 ml
1 mM EDTA	0.5 M	1 ml
0.5 mM EGTA	0.5 M	0.5 ml
DiH_2O		468.5 ml

Table 12. LB3 components.

LB3	Stock solution	Amount (250ml)
10 mM Tris-HCl, pH 8	1 M	2.5 ml
100 mM NaCl	4 M	6.25 ml
1 mM EDTA	0.5 M	0.5 ml
0.5 mM EGTA	0.5 M	0.25 ml
0.1% Na-deoxycholate		250 mg
0.5% N-lauroylsarcosine		1.25 mg
DiH_2O		240.5 ml

3.6.3. Sonicator

Before sonicating, the sonicator was pre-cooled with ice for 15 minutes. Chromatin lysates (described above) were thawed on ice. The sonicator was filled with ice-cold water and the samples were placed in the holder, ensuring they were balanced. The sonicator was set to be switched on and off every 30 seconds, with a run time of 5 minutes. After 5 minutes, the samples were vortexed to ensure DNA fragments did not settle. Water in the sonicator was removed and replaced with ice-cold water. The sonicator was run again in these 5-minute increments until it reached 30 minutes. Samples were then stored at -20^{o}C.

3.6.4. Chromatin immunoprecipitation

40 µl of Dynabeads (Invitrogen) per sample was added to an Eppendorf tube along with 600 µl of blocking solution (0.5g BSA (Sigma, UK) dissolved in 100 ml sterile PBS, filtered through a 0.44 µm filter). Using a magnetic rack, the beads were collected, and the blocking solution was removed (without disturbing the beads). The beads were washed twice more with 700 µl blocking solution and then resuspended in 700 µl of blocking solution before 2 µg of antibody was added (refer to Table 13 for details). The Eppendorf tubes were then sealed with parafilm and incubated for at least 6 hours on a rotating wheel at 4^{o}C. This could be extended to overnight incubation if required.

Samples were analysed using a Nanodrop 2000 (Thermo Fisher Scientific, UK) to determine DNA yield. The volume of the samples, which equalled 150 µg of chromatin, was calculated. Using the volume of sample needed, LB3 was added to make up a volume of 630 µl (for example, if the sample of chromatin required was 70 µl, 560 µl of LB3 would be added). 70 µl of 10% triton (giving a final triton volume of

1%) was added to the chromatin sample, giving a total volume of 700 µl. The sample was mixed by gently pipetting up and down before 70 µl was transferred into a fresh Eppendorf tube to be used as an input sample (stored at -20°C). The beads were collected from the rotating wheel and using a magnetic rack, the supernatant was removed without disturbing the beads. The remaining sample (630 µl) was added to the beads and gently mixed. The lids were sealed with parafilm and placed on a rotating wheel at 4°C overnight.

Table 13. Antibodies used in chromatin immunoprecipitation

Antibody target	**Manufacturer**
CBX2 (polyclonal rabbit)	Diagenode, Cat number: C15410339
H2AK119Ub (polyclonal rabbit)	Diagenode, Cat number: C15410002
Rabbit IgG	Diagenode, Cat number: C15410206

3.6.5. RIPA Wash

The following steps were done at 4°C. The samples were collected from the rotating wheel and placed on a magnetic rack. The supernatant was removed, without removing or disturbing the beads. The samples were then taken off the magnet and washed with 1 ml of RIPA buffer (Table 14) by gently pipetting up and down. The samples were then placed back on the magnet, and the supernatant was removed without disturbing the beads. This was repeated at least 4 more times. After the last RIPA wash, the beads were resuspended with 1 ml of ChIP wash buffer (1x TBS). Samples were then centrifuged at 960 RCF for 5 minutes at 4°C.

The following steps were done at room temperature. The beads were placed on the magnetic rack, and the TBS were removed without disturbing the beads. 200 µl of ChIP elution buffer (Table 15) was then added to the samples. Input samples were thawed on ice and 130 µl of ChIP elution buffer was added. Samples were placed on a heat block at 65°C for 7 hours to elute DNA from the beads. The beads were resuspended for the first 15 minutes by gently pipetting up and down every 5 minutes, before the Eppendorf tubes were sealed with parafilm. After incubation, samples were pulsed down, and the beads magnetised. The elute was removed and placed into fresh Eppendorf tubes. Chromatin DNA samples were then stored at -20°C or continued to the next step (Section 3.6.6).

Table 14. RIPA buffer components.

RIPA Buffer	Stock Solution	Amount (200ml)
50 mM HEPES-KOH, pH 7.5	0.5 M	20 ml
500 mM LiCl	2 M	50 ml
1 mM EDTA	0.5 M	0.4 ml
1% NP40		2 ml
0.7% Na-deoxycholate		1.4 g
DiH_2O		127.6 ml

Table 15. ChIP elution buffer components.

Elution Buffer	Stock Solution	Amount (100ml)
50 mM Tris-HCl, pH 8	1 M	5 ml
10 mM EDTA	0.5 M	2 ml
1% SDS	20%	5 ml
DiH_2O		88 ml

3.6.6. Protein and RNA digestion

Samples (if frozen) were thawed on ice. 200 µl of TE buffer (Table 16) was added to each sample. 1 µl of RNAseA (8 mg/ml) was added to each sample. Samples were then incubated at 37°C for 30 minutes, and then pulsed down. 4 µl of Proteinase K (20 mg/ml) was then added to each sample and samples were incubated at 55°C for 1 hour.

Table 16. TE buffer components.

TE Buffer	Amount (500ml)
Tris-HCl	5 ml
EDTA	1 ml
DiH_2O	494 ml

3.6.7. DNA Purification

DNA purification was performed using the PureLink Genomic DNA Mini Kit (Thermo Fisher Scientific, UK) following the manufacturer's instructions. 200 µl of genomic lysis binding buffer was added to the samples and mixed by pipetting up and down, to obtain a homogenous solution. Samples were then incubated at 55°C for 10 minutes. 200 µl of 96-100% ethanol was added to the lysates. The samples were once again homogenised by pipetting. A spin column was placed into a collection tube. The lysate was added into the spin column (done in 500 µl batches) and centrifuged at 10,000 g for 1 minute at

room temperature. Any flow through from the collection tube was discarded between spins. 500 µl of wash buffer 1 was then added to the column which was centrifuged at 10,000 g for 1 minute at room temperature. The flow-through was discarded and the collection tube was replaced with a new one. 500 µl of wash buffer 2 was added to the spin column and centrifuged at max speed for 3 minutes at room temperature. The collection tube (with flow-through) was discarded, and the spin column was placed in sterile RNase free Eppendorf tubes. 140 µl of elution buffer was added to the spin column and incubated for 5 minutes at room temperature. Samples were then centrifuged at maximum speed for 1 minute at room temperature. And purified genomic DNA was collected ready for quantitative-polymerase chain reaction (Section 3.8). If not immediately used, genomic DNA was stored at -20^{o}C.

3.7. CUTandRUN

3.7.1. Cell preparation and primary antibody binding

The CUTandRUN kit provided by Cell Signal Technology was used for this procedure following the manufacturer's instructions. First, digitonin solution was incubated at 90^{o}C-100^{o}C for 5 minutes to prepare it for use and 200x Protease Inhibitor Cocktail (PIC) and 100x Spermidine was thawed on ice. 1x wash buffer was prepared (Table 17) in accordance with how many reactions were being performed. For example, if there were 5 reactions, 2 ml would be made for each cell line with an additional 100 µl per reaction (500 µl in total). Concanavalin A Magnetic beads were resuspended by gently pipetting up and down. For each reaction, 10 µl of activation buffer was then added per sample and the beads were placed on ice.

Table 17. 1x wash buffer components.

1x Wash Buffer	**Amount (2.5ml)**
10x Wash Buffer	250 µl
100x Spermidine	25 µl
200x Protease Inhibitor Cocktail	12.5 µl
DiH_2O	2212.5 µl

6-well cell cultures were harvested by trypsinisation at room temperature to minimise stress on the cells. 100,000 cells were required per reaction and 100,000 cells for the input sample. Cells were

trypsinised and counted as above (Section 3.4) and the appropriate number of cells were pipetted into an Eppendorf tube and centrifuged at 600 g for 3 minutes at room temperature. The media was then aspirated without disturbing the cell pellet. The cell pellet was resuspended with 1 ml of 1x wash buffer. 100 µl of cell suspension was placed into a new RNA-free Eppendorf tube as an input sample and stored at -20°C. The magnetic beads were resuspended by gently pipetting up and down. 10 µl of activated beads were added (per 100,00 cells) to the cell suspension and then rotated on a rotating wheel at room temperature for 5 minutes. The sample was centrifuged at 100 g for between 30 seconds to 1 minute to remove cell:bead suspension from the cap of the tube. The samples were then placed on a magnetic rack and the supernatant was removed without disturbing the beads. The tube was then placed on the magnetic rack and 100 µl of antibody binding buffer (Table 18) was added. 100 µl of the cell:bead suspension was then aliquoted into separate tubes for each reaction and placed on ice. Relevant amounts of antibody (this may vary depending on the antibody) (Table 19) was then added to each tube and gently pipetted up and down. Tubes were sealed with parafilm and rotated on a rotating wheel for 2 hours at 4°C. During this step, the antibody binds to its target on the DNA.

Table 18. Antibody binding buffer components.

Antibody binding buffer	**Amount per reaction (100µl)**
100x Spermidine	1 µl
200x Protease Inhibitor Cocktail	0.5 µl
Digitonin Solution	2.5 µl
Antibody Binding Buffer	96 µl

Table 19. Antibodies used in CUTandRUN.

Antibody target	**Manufacturer**
CBX2 (polyclonal rabbit)	See Table 13
H2AK119Ub (polyclonal rabbit)	Diagenode, Cat number: C15410002
Rabbit (DA1E) mAb IgG	Cell signaling technology, Cat number: #66362

3.7.2. Binding of pAG-MNase enzyme

1.05 ml of digitonin buffer (components in Table 20) and a pAG-MNase pre-mix (50 µl of digitonin buffer and 1.5 µl of pAG-MNase enzyme) for each reaction was prepared. The samples were collected and centrifuged at 100 g for between 30 seconds to 1 minute. The samples were then placed on a

magnetic rack and the supernatant was removed without disturbing the beads. Samples were then removed from the rack and the beads resuspended in 1 ml of digitonin buffer. The samples were then placed on the magnetic rack and the digitonin solution was removed without disturbing the beads. The samples were removed from the rack and the pAG-MNase pre-mix was added. Tubes were then sealed with parafilm and placed on a rotating wheel for 1 hour at 4°C. During this step, the pAG-MNase will bind to the antibodies that are complexed with the protein targets on the DNA.

Table 20. Digitonin buffer components for enzyme binding.

Digitonin Buffer	**Amount per reaction (1.05ml)**
10x Wash Buffer	105 µl
100x Spermidine	10.5 µl
200x Protease Inhibitor Cocktail	5.25 µl
Digitonin Solution	26.25 µl
DiH_2O	903 µl

3.7.3. DNA digestion and diffusion

For each reaction, 2.15 ml Digitonin buffer (refer to Table 21) and 150 µl 1x stop buffer (refer to Table 22) was prepared. Calcium chloride was placed in an ice bath, ensuring the temperature stays below 4°C. Samples were collected and centrifuged at 100 g for between 30 seconds and 1 minute, to remove cell:bead suspension from the cap of the tubes. The samples were placed on the magnetic rack and the supernatant was removed without disturbing the beads. Samples were removed from the rack and the beads were resuspended in 1 ml of digitonin buffer. The tubes were placed back on the magnetic rack and the buffer was removed. This wash was repeated once more. The beads were then resuspended in 150 µl of digitonin buffer before being placed into the ice bath to cool for 5 minutes. After the incubation, 3 µl of pre-cooled calcium chloride was added to activate the MNase to cut DNA approximately 80 bp on either side of where the antibody-MNase complex was bound, the tubes were then placed back in the ice bath immediately. The samples were left to incubate in the ice bath for 30 minutes. After incubation, 150 µl of 1x stop buffer was added and the tubes were incubated at 37°C for 10 minutes. The samples were centrifuged at 16,000 g at 4°C for 2 minutes and placed on a

magnetic rack. The supernatant was removed and placed into new tubes ready for DNA purification (Section 3.7.5). If not used immediately genomic DNA was stored at -20°C.

Table 21. Digitonin buffer components for DNA digestion.

Digitonin Buffer	**Amount per reaction (2.15ml)**
10x Wash Buffer	215 µl
100x Spermidine	21.5 µl
200x Protease Inhibitor Cocktail	10.75 µl
Digitonin Solution	53.73 µl
DiH_2O	1849 µl

Table 22. 1x stop buffer components.

1x Stop Buffer	**Amount per reaction (155µl)**
4x Stop buffer	37.5 µl
Digitonin Solution	3.75 µl
RNAse A	0.75 µl
Spike-In DNA	5 µl
DiH_2O	108 µl

3.7.4. Preparation of input sample

For each input sample, 200 µl of DNA extraction buffer (comprised of 2 µl Proteinase K, 0.5 µl RNase and 197.5 µl DNA extraction buffer) was prepared. 200 µl of DNA extraction was added to the input samples and the tubes were placed on a heat block at 55°C for 1 hour with shaking. After incubation, the tubes were placed on ice for 5 minutes to cool before being sonicated for 25 minutes (refer to Section 3.6.3). Samples were centrifuged at 18,500 g for 10 minutes at 4°C to clarify the lysate. The supernatant was transferred into a new tube, without disturbing the pellet. If not immediately used, genomic DNA was stored at -20°C.

3.7.5. DNA purification

DNA purification was done using the DNA Purification Buffer and Spin Columns kit (Cell Signalling Technologies, UK) for both the input and MNase digested samples. 1.5 ml of DNA binding buffer was added to each sample and mixed by pipetting. A DNA spin column was placed into a collection tube (1 per sample), and 600 µl of the sample was transferred to the spin column. The column was centrifuged at 18,500 g for 30 seconds, and the flow-through was discarded. This was repeated until all the

samples had been spun through the spin column. The spin column was placed into an empty collection tube. 750 µl of DNA wash buffer was added to the column and centrifuged at 18,500 g for 30 seconds. The flow-through was discarded and the spin column was placed back into the collection tube. The column was centrifuged at 18,500 g for 30 seconds. The collection tube was discarded, and the spin column was placed into an RNA-free Eppendorf tube. 50 µl of DNA elution buffer was added and the tube centrifuge at 18,500g for 30 seconds to elute DNA ready for quantitative-PCR (Section 3.8).

3.8. Real-time quantitative polymerase chain reaction

All DNA samples were analysed in triplicate. For a 20 µl reaction (CUTandRUN experiment only), a PCR mastermix (18 µl per well) of 5.8 µl of molecular grade water, 2 µl of forward primer, 2 µl of reverse primer (both primers were initially diluted to a 1 µg/µl stock and then diluted again to a 1:20/1:40 working solution in molecular grade H_2O), 10 µl SYBR-Green (Sigma, UK) and 0.2 µl of ROX dye was prepared. 18 µl of mastermix was added to each relevant well of the 96 well qPCR plate (ABI, UK).

For a 10 µl reaction (conventional ChIP), a PCR mastermix (8 µl per well) of 2.1 µl of molecular grade water, 0.4 µl of forward primer, 0.4 µl of reverse primer (both primers are diluted to a 1 µg/µl stock and then diluted again to a 1:20/1:40 solution in molecular grade H_2O), 5µl SYBR-Green (Sigma, UK) and 0.1 µl of ROX dye was prepared. 8 µl of mastermix was added to each relevant well of the 96 well qPCR plate.

2 µl of DNA was then added to each reaction well and the plate covered. qPCR was performed (see Table 23 for details) using a StepOne 96 well RealTime quantitative-polymerase chain reaction (RT-qPCR) Machine (ABI, UK). Primers used were known as TSC1 66, PPKAA2 475, RBL2 120 (see Table 24 for details). A standard curve was established for each primer pair using a serial dilution of neat input DNA (1:10, 1:20, 1:50 dilutions) to determine the efficiency of the primers. Negative control of molecular grade water was also included. StepOne™ software version 2.0 was used to analyse the data.

Following RT-qPCR analysis of ChIP samples and input samples, the Ct value for the input sample was adjusted by subtracting 3.3 (to reflect the fact $1/10^{th}$ of chromatin was used to prepare the input sample compared to the amount of chromatin used in the IP; $\log_2 0.1 = -3.3$). The Ct value for the IP sample was then subtracted from the relevant "adjusted input" value. To convert this value into percentage DNA content compared to input DNA content (% input) 2 was multiplied to the power of the calculated Ct which was then multiplied by 100. For the CUTandRUN analysis, no adjustment was needed for the input samples as the same number of cells (and therefore DNA) was used to generate the input and CUTandRUN samples. Subsequent Ct values for the CUTandRUN samples were subtracted from the Ct value for the relevant input. To convert the value into percentage DNA content compared to input DNA content (% input) 2 was multiplied to the power of the calculated Ct which was then multiplied by 100. These are optimised approaches as used in papers by Wade *et al*., (2015) and Jones *et al.,* (2019).

Table 23. RT-qPCR cycle information.

Stage	**Temperature (°C)**	**Gradient (%)**	**Time (s)**
Holding Stage	95	100	600
Cycling stage step 1	95	100	15
Cycling stage step 2	60	100	60

Steps 2 and 3 are repeated for 40 cycles.

Table 24. Primer sequence in RT-qPCR.

Primer	**Sequence (5'-3')**	
TSC1 66	Forward	GCCGTCTATCCTTCCTTTCGA
	Reverse	CGCCAGGAAAAAGAGTCCC
PPKAA2 475	Forward	TTCCCTTTTACAGCCCCTCG
	Reverse	TGGAAGAAGAGACGGGCCT
RBL2 120	Forward	CATGATTTTTGGCCCCCTTGA
	Reverse	CAGGCACCCGTAGTCTTGA

Primers designed by Mark Wade based on the promotor sequences of these genes. Numbers relate to how many base pairs upstream the promoters are from the transcriptional start site.

4. Results

4.1. *In silico* analysis of CBX2 gene expression in tumour and normal patient samples.

4.1.1. Analysis of CBX2 expression in breast cancer

CBX2 has been identified to be associated with breast cancer development and progression, (Liang *et al.*, 2017; Zheng *et al.*, 2019) with preliminary data (not shown) in the Wade lab supporting this. To investigate this further, *CBX2* expression was compared between tumour and normal breast tissue, and across breast cancer subtypes using GEPIA2 software (Tang *et al.*, 2019). The effect of *CBX2* expression on OS and DFS in breast cancer (and different breast cancer subtypes) was also analysed. The datasets analysed were from the TCGA and GTEx gene expression databases. The samples collected for GTEx and TCGA datasets were processed using different protocols, however, GEPIA2 analysis combines the raw data from both datasets to produce a combined dataset analysis. Using both datasets allow us to have a larger sample size of normal samples for which to compare tumour gene expression as for some cancers very few normal samples are included in the TCGA datasets.

Gene expression analysis on RNA extracted from tumours and normal breast tissue was performed on both the combined TCGA and GTEx normal breast tissue datasets (Figure 4.1.A Tumour n = 1085; Normal n = 291) and the TCGA dataset alone (Figure 4.1.B Tumour n = 1085; Normal n = 112) to determine if there was a difference in CBX2 expression in cancer tissue compared to normal breast tissue. Both analyses identified a distinct trend of elevated CBX2 expression in tumour samples compared to normal samples; however only analysis of the TCGA dataset alone showed statistical significance (Figure 4.1. $p = <0.01$).

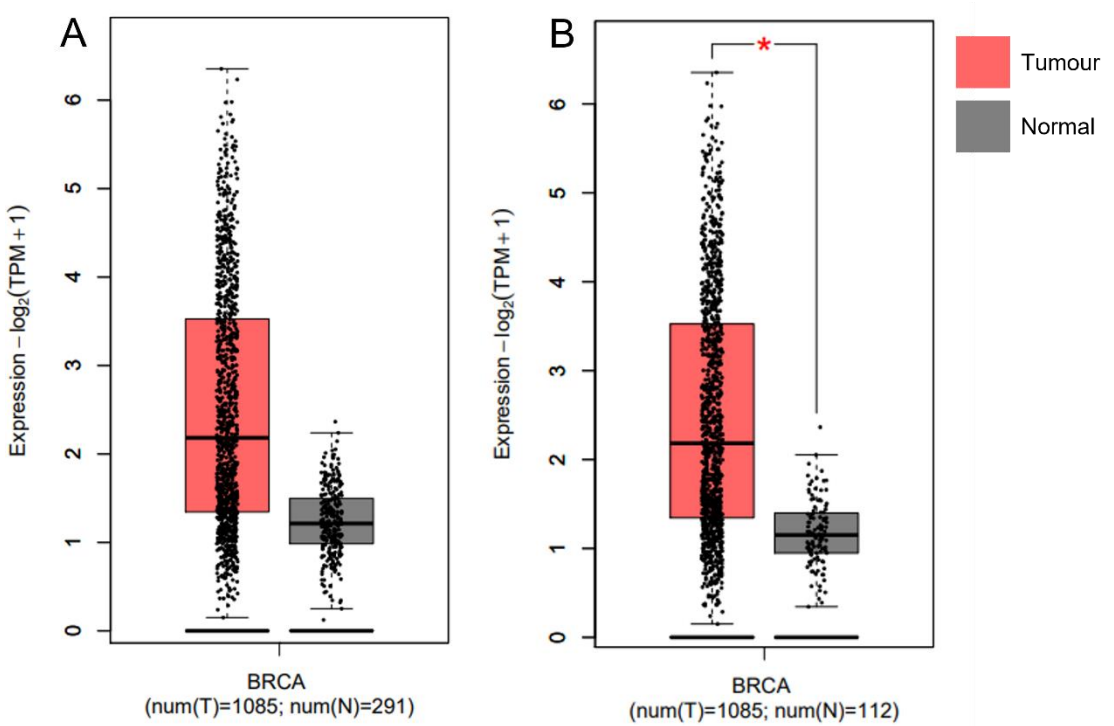

Figure 4.1. CBX2 mRNA expression is higher in breast cancer versus normal breast tissue.
A) Comparison between TCGA *CBX2* expression from tumour samples (n = 1085) versus *CBX2* expression in normal tissue from TCGA and GTEx datasets (n = 291). B) Comparison between TCGA *CBX2* expression from tumour samples (n = 1085) versus *CBX2* expression in normal tissue from the TCGA dataset alone (n = 112). T = Tumour samples; N = Normal samples. * = p-value < 0.01

Different breast cancer subtypes were then analysed to see if differences in expression were subtype-specific. Figure 4.2A shows data from combined TCGA and GTEx normal sample datasets while Figure 4.2B shows data from the TCGA dataset alone. The figure shows there is a trend of elevated levels of CBX2 expression in tumour samples compared to normal samples, especially in basal-like and HER2 breast cancer subtypes. In both analyses, basal-like, HER2 and luminal B subtypes showed statistically significant increased CBX2 expression compared to normal breast tissue ($p<0.01$).

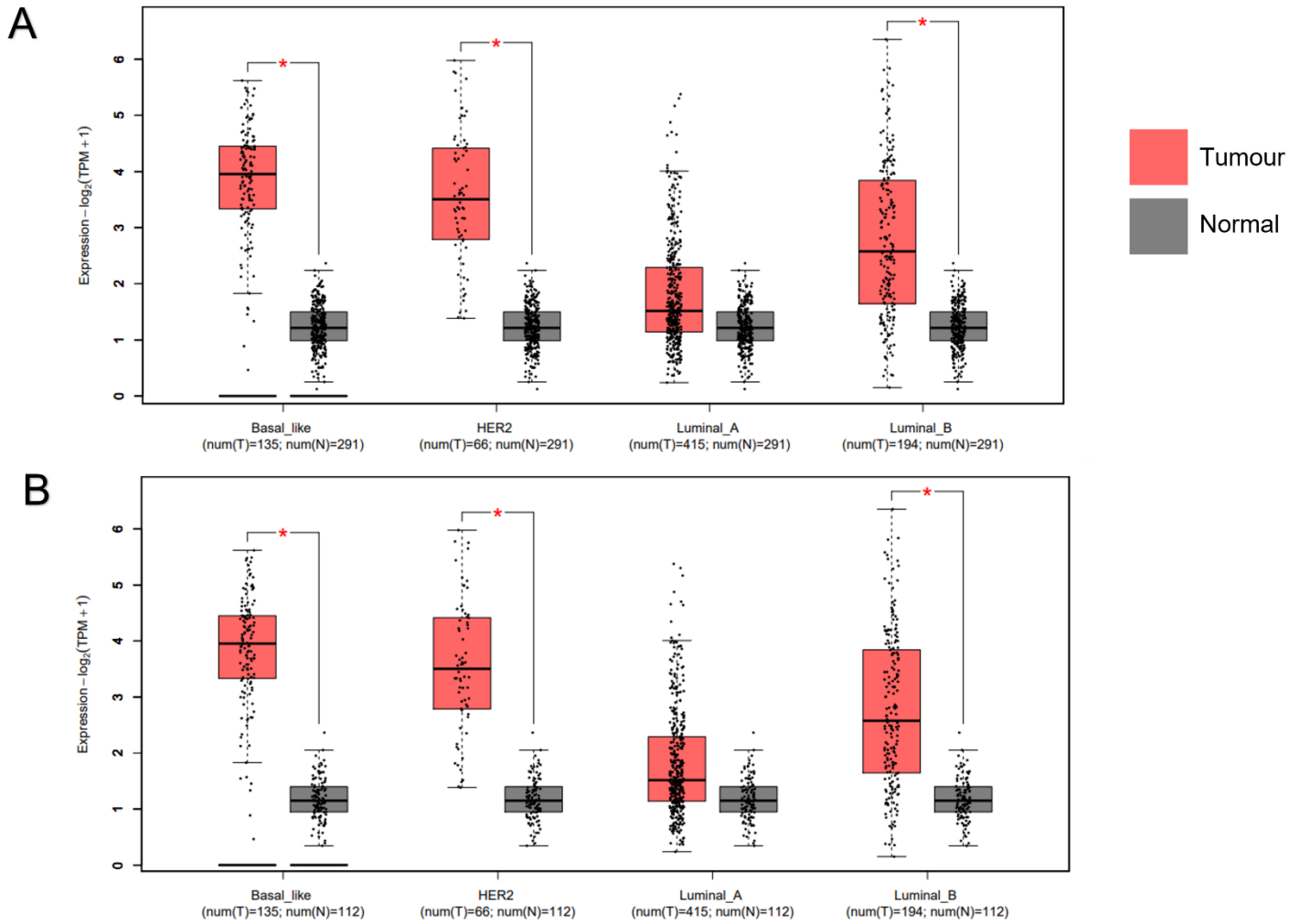

Figure 4.2. CBX2 mRNA expression is higher in different breast cancer subtypes vs normal breast tissue.

A) Comparison between TCGA *CBX2* expression from tumour samples versus *CBX2* expression in normal tissue from TCGA and GTEx datasets.

B) Comparison between TCGA *CBX2* expression from tumour samples versus *CBX2* expression in normal tissue from the TCGA dataset alone.

T = Tumour samples; N = Normal samples. * = p-value < 0.01

4.1.2. Survival Analysis

Considering CBX2 expression is elevated in breast cancer compared to normal breast tissue, a survival curve analysis was performed to determine if the level of CBX2 expression had an impact on both OS (Figure 4.3) and DFS (Figure 4.4) for breast cancer patients from the TCGA dataset. For OS in breast cancer patients (n = 1070), a HR of 1.4 for samples with "high" CBX2 expression (tumour samples whose expression was above the median expression) was identified, which was statistically significant (p(HR) = <0.05). This indicates a 1.4-fold increase in the risk of those in the elevated expression group dying throughout the study.

For OS in different subtypes of breast cancer (Figure 4.3B-E), HR for the high expression group were luminal A = 1.3, luminal B = 1.3, HER2 = 0.81 and basal-like = 0.84, although none of these HR calculations was statistically significant.

For DFS in breast cancer, a HR for the high group of 1 (indicating no increased risk for either group) was identified (Figure 4.4A) while analysis of DFS for each breast cancer subtype identified, HR of luminal A = 0.99, luminal B = 0.84, HER2 = 0.18 and basal-like = 1.7 (Figure 4.4B-E). There is an apparent increased risk of dying of breast cancer for patients in the basal-like subtype as it indicates a 1.7-fold increase in those in the higher expression group dying, however, all the p-values are over 0.05, indicating the increased risk is not significant.

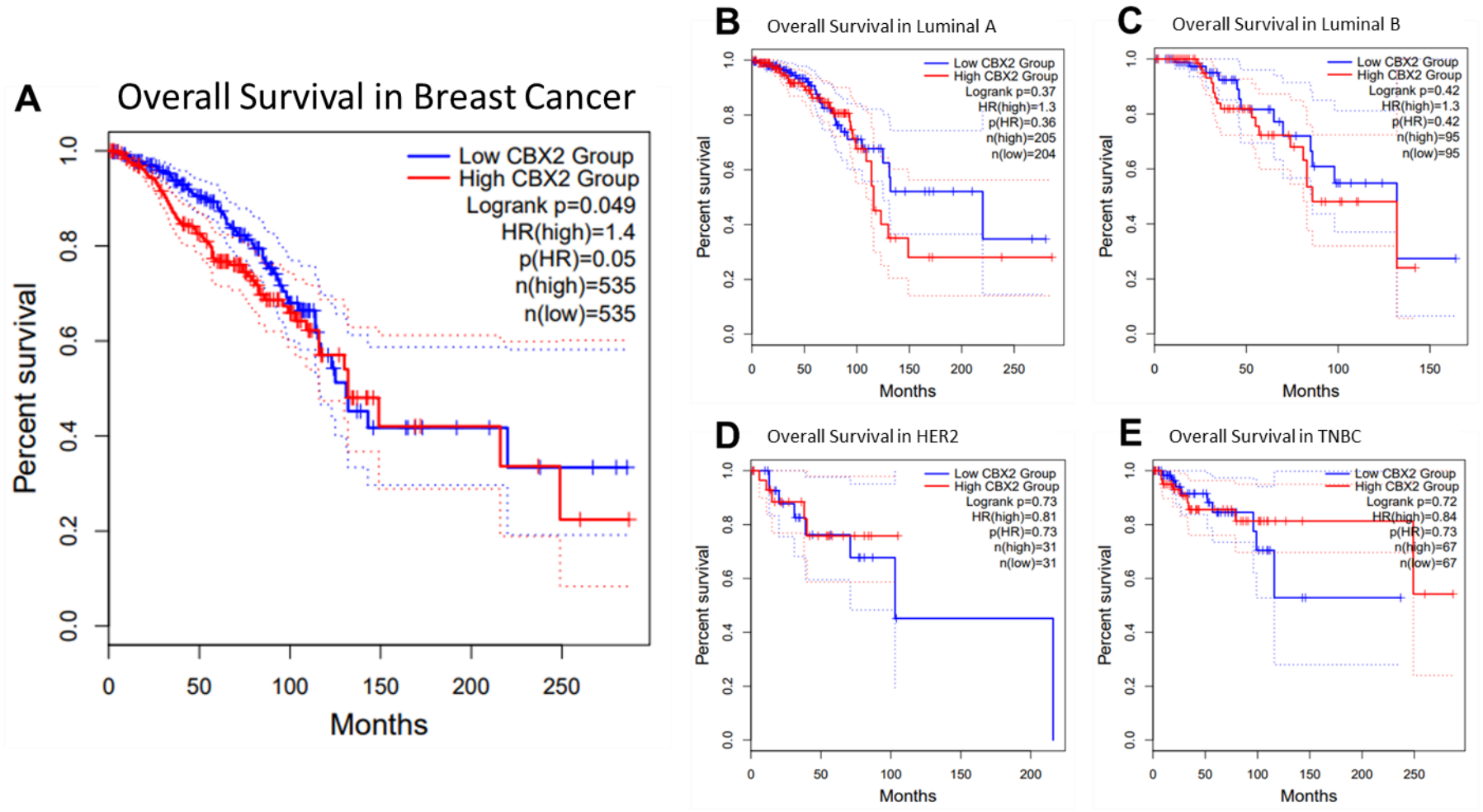

Figure 4.3. High CBX2 expression decreases OS in breast cancer.

A) Comparison of OS between high and low *CBX2* expression patients (as defined as above or below the median expression within the patient cohort) in breast cancer (n = 1070). B) Comparison in Luminal A patient (n = 409). C) Comparison in Luminal B patients (n = 190). D) Comparison in HER2 overexpression subtype patients (n = 62). E) Comparison in Basal-like patients (n = 134). Blue = low CBX2 expression group, red = high CBX2 expression group. Dotted lines = confidence intervals. Group cut-off was 50%. Log rank p-value must be less than 0.05 to be significant. HR = Hazard ratio.

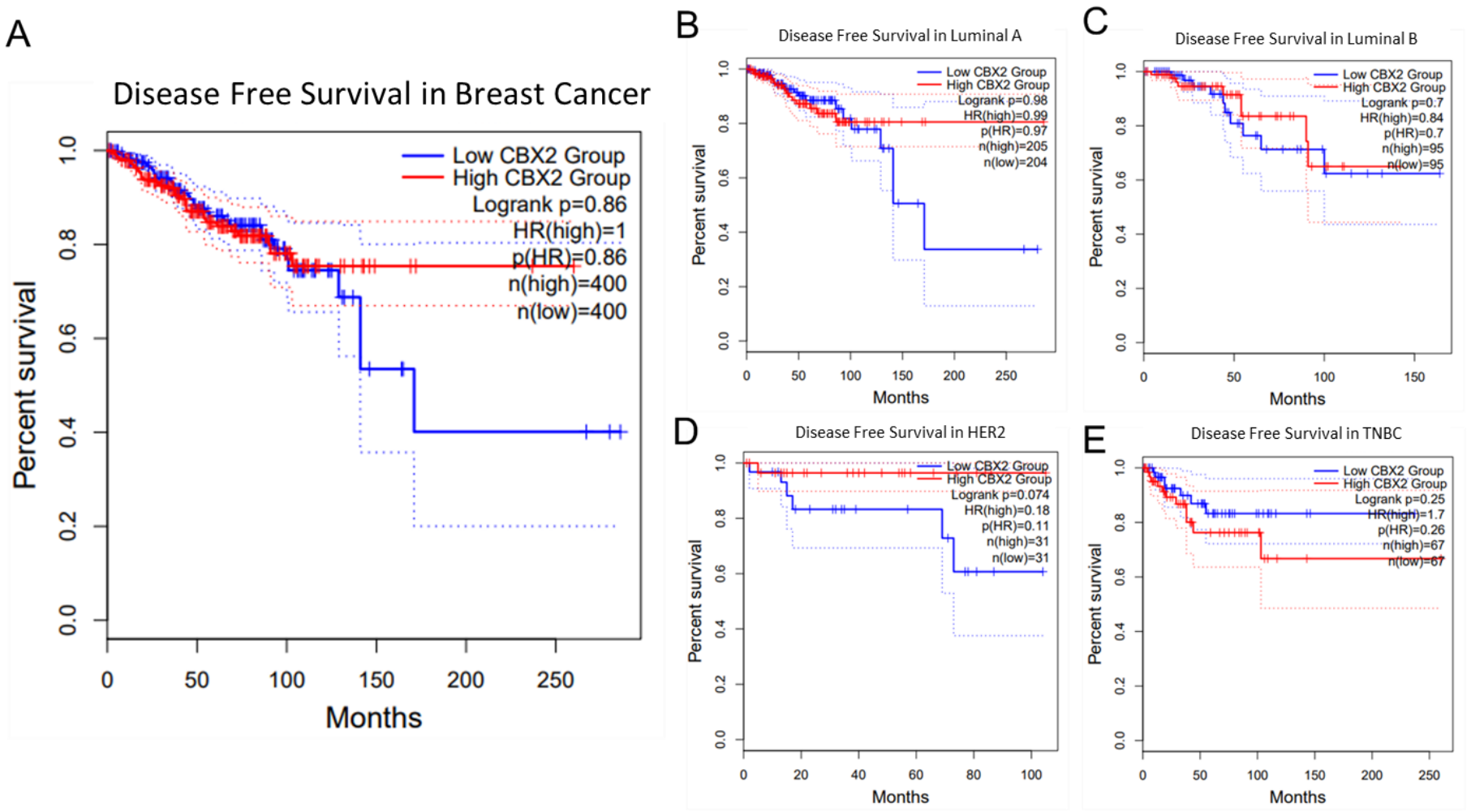

Figure 4.4. High CBX2 expression in decreases DFS in TNBC compared to other breast cancer subtypes

A) Comparison of DFS between high and low *CBX2* expression patients (as defined as above or below the median expression within the patient cohort) in breast cancer (n = 1070). B) Comparison in Luminal A patient (n = 409). C) Comparison in Luminal B patients (n=190). D) Comparison in HER2 overexpression subtype patients (n = 62). E) Comparison in Basal-like patients (n = 134). Blue = low CBX2 expression group, red = high CBX2 expression group. Dotted lines = confidence intervals. Group cut-off was 50%. Log rank p-value must be less than 0.05 to be significant. HR = Hazard ratio.

4.1.3. CBX2 expression analysis in tumour vs normal samples in other cancers

Our analysis indicates CBX2 is elevated in breast cancer compared to normal tissue, with evidence that breast cancer patients with elevated expression of CBX2 have a slightly poorer prognosis. Due to this, we further explored the expression of CBX2 in tumour vs normal tissue in other cancer types. Cancer types chosen for analysis were glioblastoma multiforme (GBM), lung adenocarcinoma (LUAD), ovarian serous cystadenocarcinoma (OV), prostate adenocarcinoma (PRAD), and testicular germ cell tumours (TGCT). These were chosen due to my preliminary research from published work indicating a possible role for CBX2 and these cancer types (Zhang *et al*., 2012; Clermont *et al*., 2016; Wheeler *et al*., 2018; Piqué *et al*., 2019). The cancers were analysed using data from the combined TCGA and GTEx normal sample dataset (Figure 4.5A). LUAD and PRAD were also analysed using data from the TCGA dataset alone (Figure 4.5B). The analysis showed that all cancers showed a trend of increased *CBX2* expression in tumours compared to normal tissue samples. Differences in LUAD when comparing against TCGA and GTEx normal expression datasets, and GBM when comparing with TCGA alone were statistically significantly ($p<0.01$).

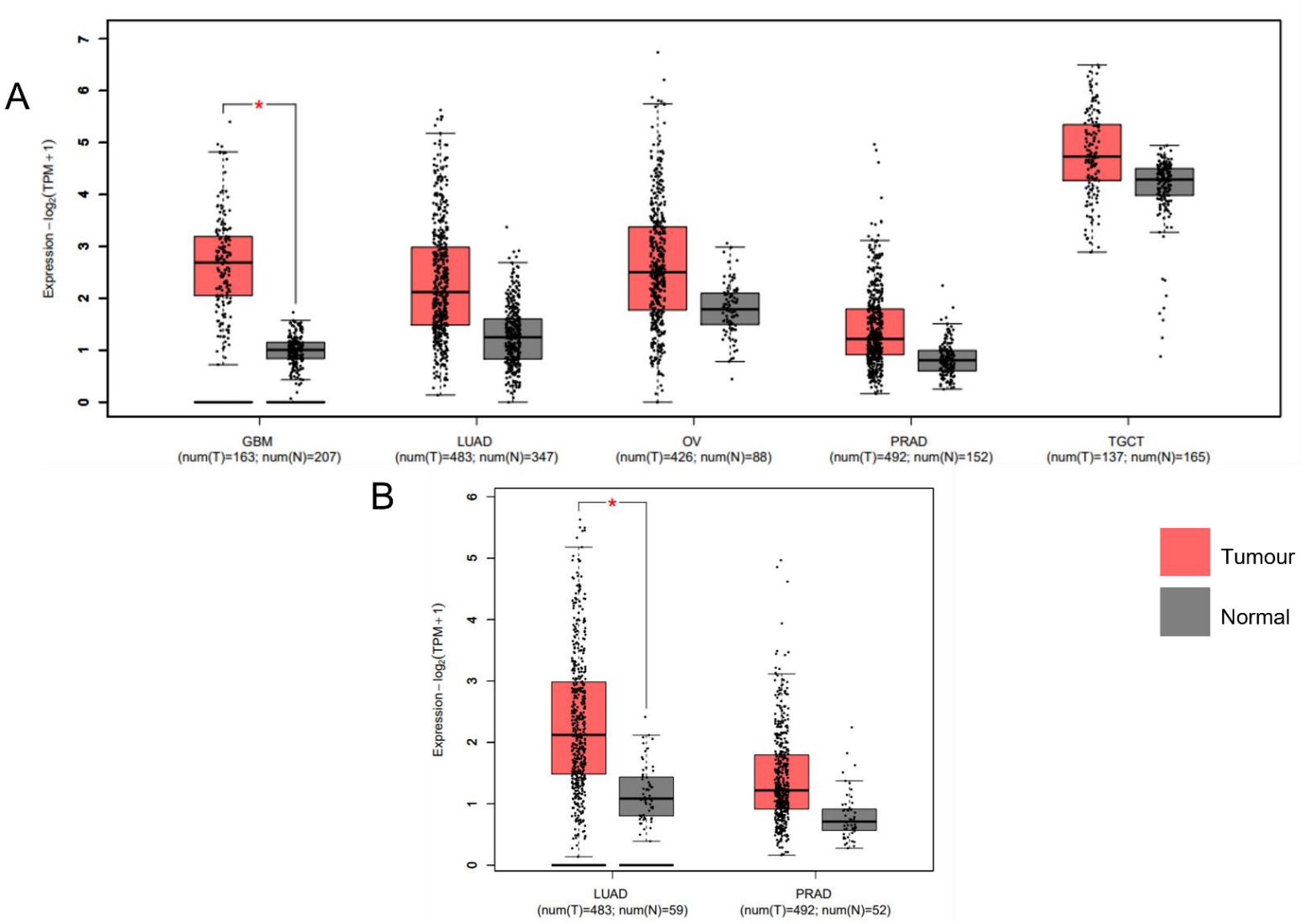

Figure 4.5. CBX2 mRNA expression is higher in different cancer tissues, compared to normal tissues.

A) Comparison between TCGA *CBX2* expression from tumour samples versus *CBX2* expression in normal tissue from TCGA and GTEx datasets. (B) Comparison between TCGA *CBX2* expression from tumour samples versus *CBX2* expression in normal tissue from the TCGA dataset alone. T = Tumour samples; N = Normal samples. * = p-value < 0.01. GBM = Glioblastoma multiforme, LUAD = Lung adenocarcinoma, OV = Ovarian serous cystadenocarcinoma, PRAD = Prostate adenocarcinoma, TGCT = Testicular Germ Cell Tumours.

4.1.4. Survival analysis of CBX2 expression in other cancer types

Survival curve analysis was performed to determine if the level of *CBX2* expression has an impact on both OS (Figure 4.6) and DFS (Figure 4.7) in GBM, LUAD, OV, PRAD and TGCT. In the OS analysis, the HR for the high groups for each cancer were: GBM = 1.1, LUAD = 1.2, OV = 0.77, PRAD = 0.97 and TGCT = 0.33. Both GBM and LUAD, show that the high expression group has a 1.1-fold risk and 1.2-fold risk (respectively) of dying, but this is not statistically significant. In OV, the "high" *CBX2* expression group had a better survival rate than those in the low expression group. All p-value are over 0.05 except for OV (p(HR) = 0.037).

For disease free survival, the HR for the high groups for each cancer were: GBM = 0.66, LUAD = 1.2, OV = 0.82, PRAD = 2.1 and TGCT = 0.96. In LUAD and PRAD, those in the high expression group have a 1.2-fold risk and 2.1-fold risk (respectively) of dying, however, only PRAD is statistically significant (p-value = 0.00063). All p-values were over 0.05 except for GBM, which showed better DFS in the high expression cohort (p(HR) = 0.05) and PRAD (p(HR) = 0.00063).

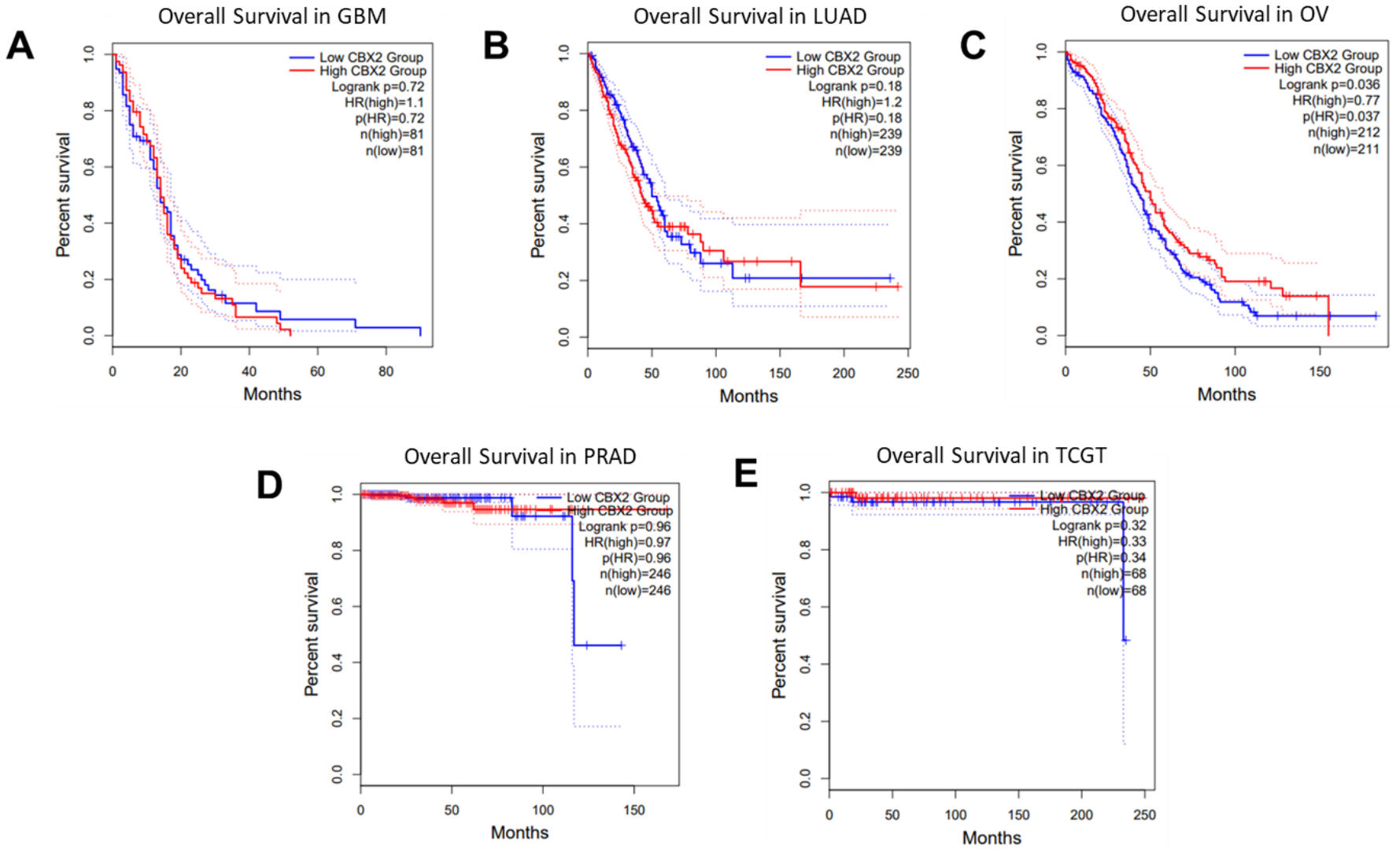

Figure 4.6. High CBX2 expression decreases OS in GBM and LUAD, compared to different cancer types.

Comparison of OS between high and low *CBX2* expression patients (as defined as above or below the median expression within the patient cohort) in GBM (n = 162) (A), LUAD (n = 478) (B), in OV (n = 423) (C), in PRAD (n = 492) (D), and in TCGT (n = 136) (E). Blue = low CBX2 expression group, red = high CBX2 expression group. Dotted lines = confidence intervals. Group cut-off was 50%. Log rank p-value must be less than 0.05 to be significant. HR = Hazard ratio. GBM = Glioblastoma multiforme, LUAD = Lung adenocarcinoma, OV = Ovarian serous cystadenocarcinoma, PRAD = Prostate adenocarcinoma, TGCT = Testicular Germ Cell Tumours.

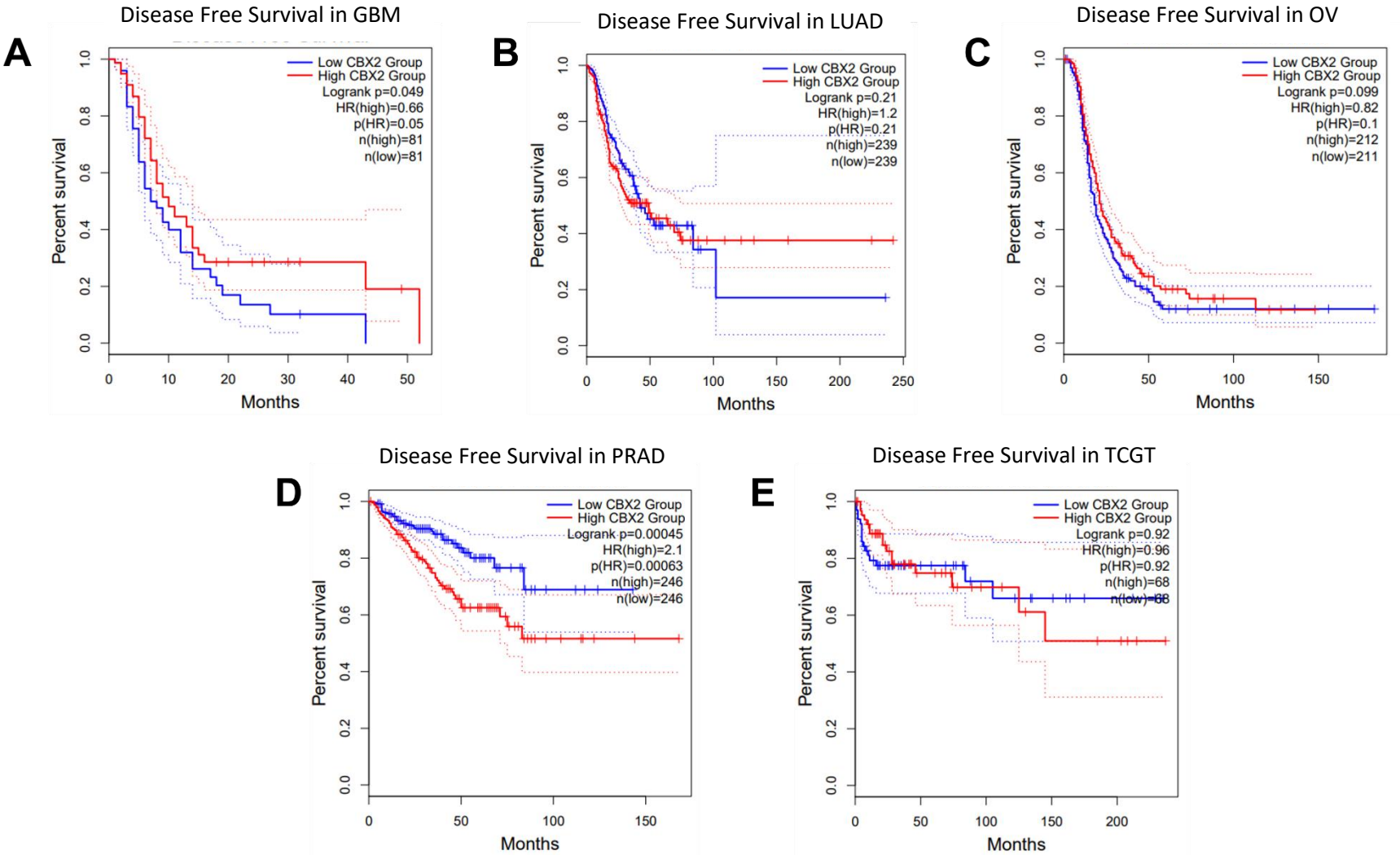

Figure 4.7. High CBX2 expression decreases DFS in LUAD and PRAD compared to different cancer types.

Comparison of DFS between high and low *CBX2* expression patients (as defined as above or below the median expression within the patient cohort) in GBM (n = 162) (A), LUAD (n = 478) (B), in OV (n = 423) (C), in PRAD (n = 492) (D), and in TCGT (n = 136) (E). Blue = low CBX2 expression group, red = high CBX2 expression group. Dotted lines = confidence intervals. Group cut-off was 50%. Log rank p-value must be less than 0.05 to be significant. HR = Hazard ratio. GBM = Glioblastoma multiforme, LUAD = Lung adenocarcinoma, OV = Ovarian serous cystadenocarcinoma, PRAD = Prostate adenocarcinoma, TGCT = Testicular Germ Cell Tumours.

4.2. In silico analysis of the potential regulatory role of CBX2 in cancer

4.2.1. Generation of lists of genes positively and inversely correlated with CBX2 expression in patient gene expression datasets

The results from the GEPIA2 analysis gave further evidence of *CBX2* having a potential role in development and progression in different cancer types, including breast cancer. This was demonstrated by the fact *CBX2* was overexpressed in cancer compared to normal tissue and that in some cases, elevated expression led to poorer disease outcome. To explore the role of *CBX2* further, *CBX2* expression was queried in two different breast invasive carcinoma datasets (TCGA, Pancancer (n = 994) and METABRICS (n = 1904)) using the cBio Cancer Genomics Portal (cBioPortal) bioinformatics tools. Two of the other cancers investigated by GEPIA2 analysis were also analysed using cBioPortal (glioblastoma (n = 145) and lung adenocarcinoma (n = 503) TCGA, Pancancer datasets) due to their expression being significantly elevated in cancer versus normal tissue.

cBioPortal analysis was used to identify genes whose expression was positively and negatively correlated with *CBX2* expression in each different dataset. Identified genes with a spearman's correlation p-value over 0.05 were not used for subsequent analysis.

Significantly correlated gene lists were compared between all four datasets analysed to determine the number of overlapping genes, and therefore a potentially common CBX2 correlation gene signature. From the 4 different datasets, 3795 genes overlapped (Figure 4.8) – indicating a large number of genes which overlap between different cancers. Further analysis showed 1229 genes were commonly positively correlated with CBX2 expression (Figure 4.9) and 843 genes were commonly negatively correlated with CBX2 expression (Figure 4.10) across the four different datasets.

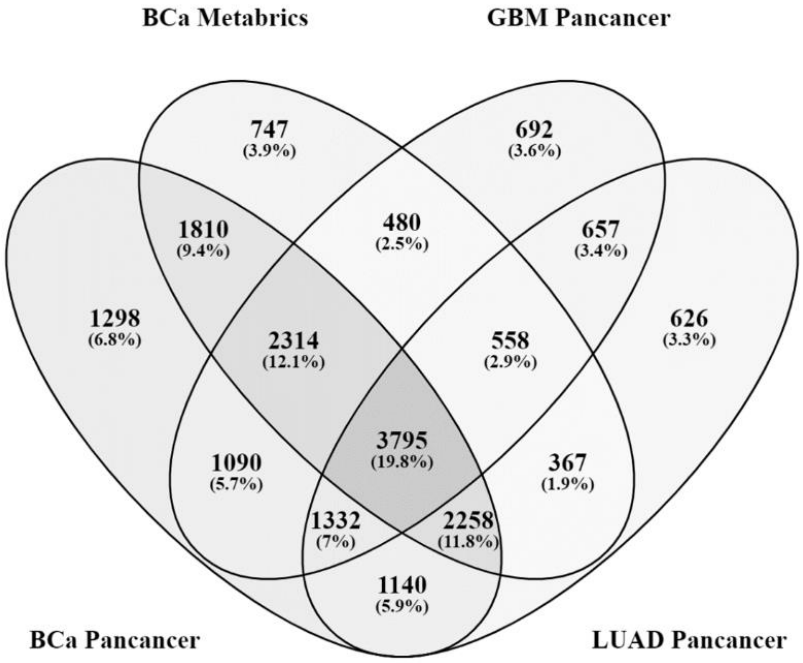

Figure 4.8. Venn diagram comparing the number of genes which both positively and negatively correlate with CBX2 gene expression list from 4 different patient datasets.

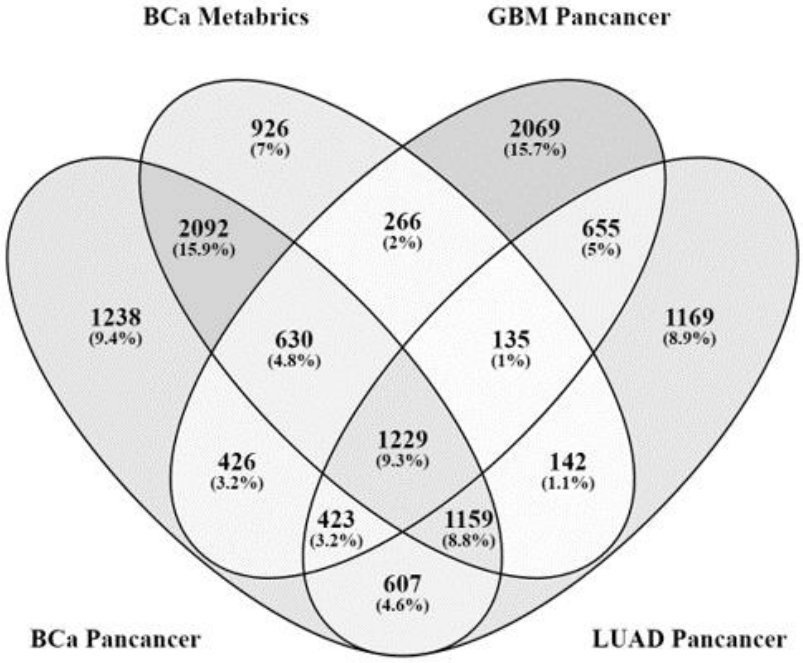

Figure 4.9. Venn diagram indicating 1229 genes which positively correlate with CBX2 gene expression from 4 different patient datasets

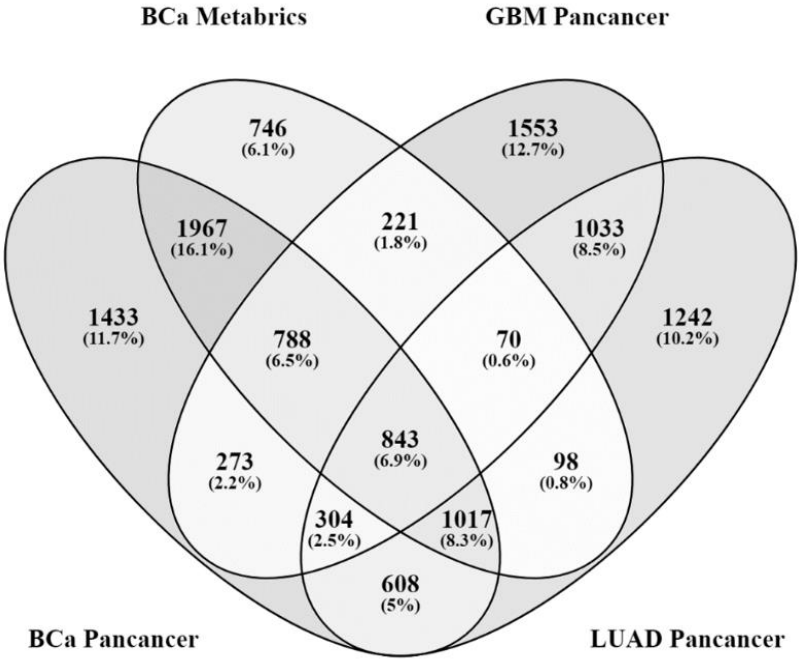

Figure 4.10. Venn diagram indicating 843 genes which negatively correlate with CBX2 gene expression from 4 different patient datasets.

4.2.2. Gene set enrichment analysis of genes positively and inversely correlated with CBX2 expression

To investigate the potential regulatory role of CBX2 in cancer further, the correlation gene lists obtained from cBioPortal were ordered by Spearman correlation coefficient (strongest positive correlation to strongest inverse correlation) and compared against three different GSEA curated datasets (hallmarks gene sets, ontology gene sets and oncogenic gene signature gene set). This analysis aimed to determine which biological and oncogenic processes were enriched for genes that CBX2 positively and negatively correlated with, thereby potentially indicating which processes CBX2 plays a role in regulating.

The hallmark gene sets represent well-defined biological states or processes (Subramanian *et al*., 2005; Liberzon *et al*., 2015). Oncogenic signature gene sets represent signatures of cellular pathways which are often dysregulated in cancer. These signatures were generated directly from microarray data from NCBI GEO (Subramanian *et al*., 2005; Liberzon *et al*., 2015).

4.2.2.1. Hallmark enrichment plots for breast cancer

GSEA identified several hallmark gene sets enriched for genes identified to be positively and inversely correlated with *CBX2* expression from the TCGA Pancancer database (Table 25 and Table 26). Enrichment score (ES) reflects the degree to which a gene set is overrepresented at the top or bottom of a ranked list of genes, represented as a deviation from zero. Normalized enrichment score (NES) is the primary statistic, by normalizing the ES, to compare analysis results across different genes sets. It is determined by; ES/mean (against all permutations of a dataset). False discovery rate (FDR) is the estimated probability that a gene set with a given NES represents a false positive e.g., an FDR of 25% indicates that the result is likely to be valid 3 out of 4 times. The nominal p-value estimates the statistical significance of the ES for a single gene set.

Gene sets enriched with genes that positively correlate with *CBX2* expression are indicated by a positive NES. Four hallmark gene sets were identified as of interest in the database. These were G2M checkpoint (genes that are involved in G2/M checkpoint and the cell cycle), E2F signalling (genes that are targets of E2F transcription factors), MYC target (subgroup of genes related to MYC) and MTORC1 signalling (genes upregulated through activation of mTORC1) (Subramanian *et al*., 2005; Liberzon *et al*., 2015). These have been indicated to be associated with cancer development or progression. The positive NES are E2F signalling (NES = 9.796536), G2M checkpoint (NES = 8.404441), MYC targets (NES = 7.7615705) and MTORC1 signalling (NES = 5.409676) (Figure 4.11). The NES values indicate a strong enrichment for the positively correlated genes. This indicates CBX2 may have role in E2F and mTORC1 signalling. Figure 4.12 shows hallmark gene sets that are enriched for genes inversely correlated with *CBX2* expression. Two other hallmark gene set where apparent in the database, which were Estrogen response early and estrogen response late (genes which define an early, or late response to oestrogen) (Subramanian *et al*., 2005; Liberzon *et al*., 2015). Negative NES are estrogen response early (NES = -5.0916157) and estrogen response late (NES = -3.8914905). All q-values were less than 0.001.

Table 25. Top 10 Hallmarks with a positive enrichment score in breast cancer, TCGA Pancancer.

Name of Hallmark	Number of Genes	ES	NES	Nominal p-value	FDR q-value	FWER p-value
E2F_TARGETS	183	0.619	9.797	<0.001	<0.001	<0.001
G2M_CHECKPOINT	180	0.543	8.404	<0.001	<0.001	<0.001
MYC_TARGETS_V1	168	0.511	7.762	<0.001	<0.001	<0.001
ALLOGRAFT_REJECTION	156	0.384	5.666	<0.001	<0.001	<0.001
MTORC1_SIGNALING	172	0.354	5.410	<0.001	<0.001	<0.001
INTERFERON_GAMMA_RESPONSE	163	0.315	4.719	<0.001	<0.001	<0.001
MYC_TARGETS_V2	56	0.495	4.306	<0.001	<0.001	<0.001
MITOTIC_SPINDLE	159	0.247	3.522	<0.001	<0.001	<0.001
GLYCOLYSIS	167	0.220	3.292	<0.001	<0.001	<0.001
INTERFERON_ALPHA_RESPONSE	70	0.323	3.283	<0.001	<0.001	<0.001

Table 26. Top 10 hallmarks with a negative enrichment score in breast cancer, TCGA Pancancer.

Name of Hallmark	Number of Genes	ES	NES	Nominal p-value	FDR q-value	FWER p-Value
ESTROGEN_RESPONSE_EARLY	169	-0.339	-5.092	<0.001	<0.001	<0.001
ESTROGEN_RESPONSE_LATE	176	-0.247	-3.891	<0.001	<0.001	<0.001
UV_RESPONSE_DN	114	-0.289	-3.637	<0.001	<0.001	<0.001
BILE_ACID_METABOLISM	89	-0.203	-2.237	0.002	0.005	0.029
MYOGENESIS	145	-0.153	-2.146	0.000	0.007	0.049
FATTY_ACID_METABOLISM	129	-0.114	-1.498	0.059	0.215	0.841
TGF_BETA_SIGNALING	38	-0.195	-1.435	0.088	0.250	0.911
HEME_METABOLISM	152	-0.098	-1.429	0.094	0.224	0.915
EPITHELIAL_MESENCHYMAL_TRANSITION	154	-0.096	-1.394	0.109	0.230	0.947
ANDROGEN_RESPONSE	77	-0.123	-1.257	0.180	0.353	0.986

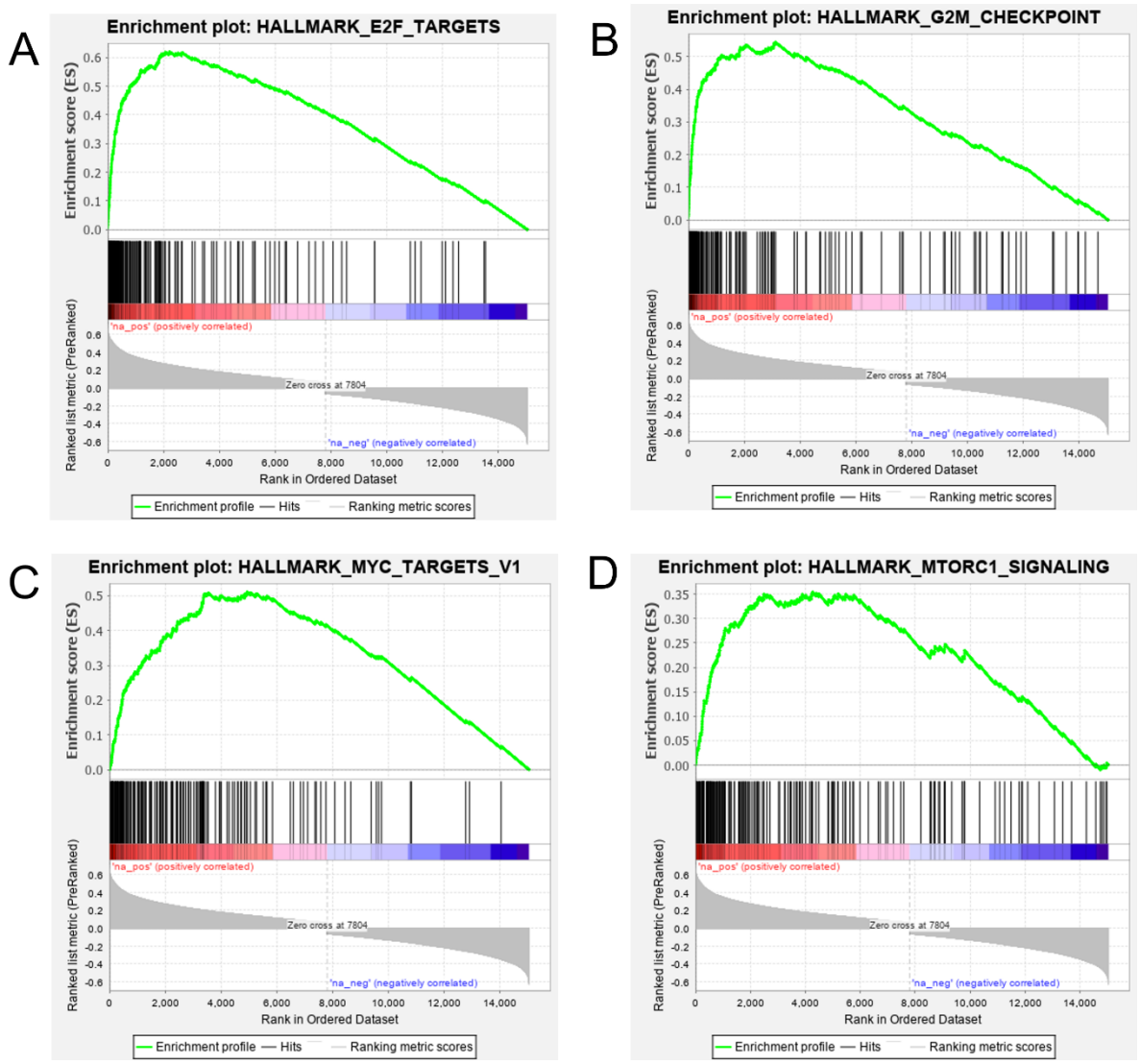

Figure 4.11. Positively correlated GSEA hallmark enrichment plots in breast cancer, using TCGA, Pancancer database.

The black lines indicate the different genes within the hallmark set. Lines within the red section indicate genes positively correlated with *CBX2* while the blue section show inversely correlated with *CBX2* expression. Genes are ranked by spearman's coefficient. A-D are hallmarks enrichment points that are positively correlated by *CBX2*. A) FDR = <0, NES = 9.796536 , Nominal P-value = <0. B) FDR = <0, NES = 8.404441 , Nominal P-value = <0. C) FDR = <0, NES = 7.7615705, Nominal P-value = <0. D) FDR = <0, NES = 5.409676, Nominal P-value = <0.

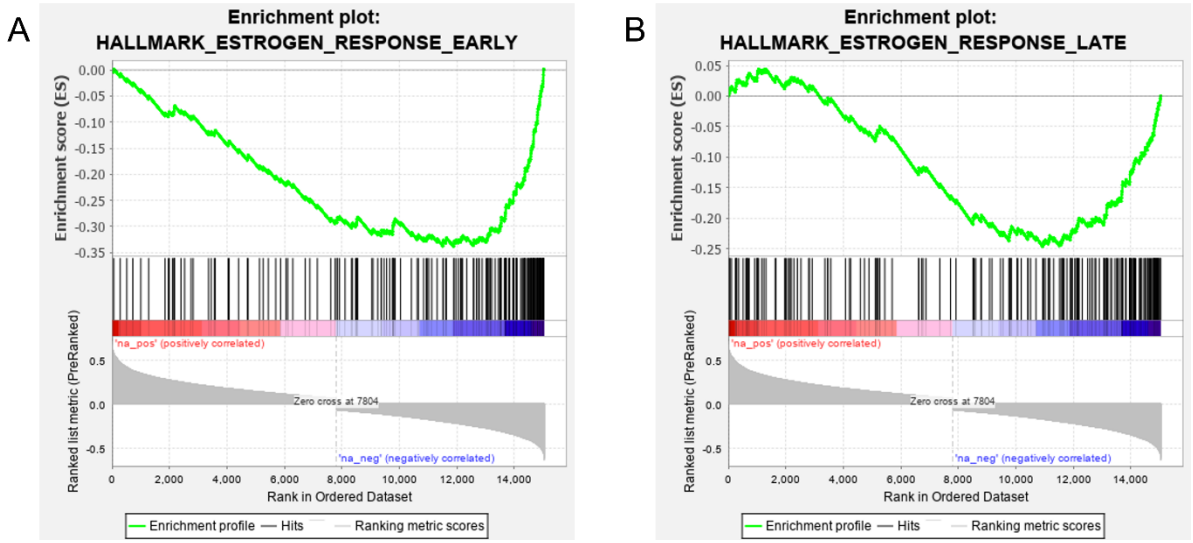

Figure 4.12. Negatively correlated GSEA hallmark enrichment plots in breast cancer, using TCGA, Pancancer database.

GSEA of Breast Cancer TCGA, PanCancer database. The black lines indicate the different genes within the hallmark set. Lines within the red section indicate genes positively correlated with *CBX2* while the blue section show inversely correlated with *CBX2* expression. Black lines in the red area are genes positively correlated, while lines in the blue are genes negatively correlated. Genes are ranked by spearman's coefficient. A and B are hallmarks enrichment points that are negatively correlated by CBX2. A) FDR = <0, NES =-5.0916157, Nominal P-value = <0. B) FDR = <0, NES = -3.8914905, Nominal P-value = <0.

GSEA also identified numerous hallmark gene sets enriched for genes identified to be positively and inversely correlated with CBX2 expression from the METABRICS database (Table 27 and Table 28). Gene sets that are associated with cancer development or progression and E2F or mTORC1 signalling were apparent in this database. Positive NES are E2F signalling (NES = 3.850345), G2M checkpoint (NES = 3.7101395), MTORC1 signalling (NES = 2.8559391) and MYC targets (NES = 2.8950305) (Figure 4.13). The NES values scores again show strong enrichment for the positively correlated genes. All q-values were less than 0.01. Genes that negatively correlate with CBX2 expression, indicated by the negative NES are estrogen response early (NES = -2.8874424) and estrogen response late (NES = -2.0190015) (Figure 4.14). All q-values were less than 0.01. This analysis shows that common hallmark gene sets identified from two different breast cancer databases, indicate CBX2 may have a role in E2F or mTORC1 signalling, as well as cancer progression. All q-values were less than 0.001, except for estrogen response late which has a q-value of 0.003.

Table 27. Top 10 hallmarks with a positive enrichment score in breast cancer in METABRICS database

Name of Hallmark	Number of Genes	ES	NES	Nominal p-value	FDR q-value	FWER p-Value
E2F_TARGETS	183	0.619	9.797	<0.001	<0.001	<0.001
G2M_CHECKPOINTS	180	0.543	8.404	<0.001	<0.001	<0.001
MYC_TARGETS_V1	168	0.511	7.762	<0.001	<0.001	<0.001
ALLOGRAFT_REJECTION	156	0.384	5.666	<0.001	<0.001	<0.001
MTORC1_SIGNALING	172	0.354	5.410	<0.001	<0.001	<0.001
MYC_TARGETS_V2	45	0.685	2.940	<0.001	<0.001	<0.001
INTERFERON_GAMMA_RESPONSE	156	0.492	2.661	<0.001	<0.001	<0.001
MITOTIC_SPINDLE	158	0.450	2.449	<0.001	<0.001	<0.001
UNFOLDED_PROTEIN_RESPONSE	86	0.455	2.224	<0.001	<0.001	<0.001
INTERFERON_ALPHA_RESPONSE	70	0.458	2.179	<0.001	<0.001	<0.001

Table 28. Top 10 hallmarks with a negative enrichment score in breast cancer in METABRICS database

Name of Hallmark	Number of Genes	ES	NES	Nominal p-value	FDR q-value	FWER p-Value
ESTROGEN_RESPONSE_EARLY	157	-0.515	-2.887	<0.001	<0.001	<0.001
ESTROGEN_RESPONSE_LATE	173	-0.356	-2.019	<0.001	0.003	0.007
UV_RESPONSE_DN	117	-0.344	-1.836	<0.001	0.006	0.024
FATTY_ACID_METABOLISM	125	-0.264	-1.415	0.023	0.141	0.533
BILE_ACID_METABOLISM	81	-0.277	-1.378	0.036	0.148	0.621
KRAS_SIGNALING_DN	121	-0.253	-1.353	0.037	0.153	0.697
ADIPOGENESIS	151	-0.219	-1.217	0.094	0.338	0.956
MYOGENESIS	136	-0.215	-1.182	0.149	0.375	0.977
HEME_METABOLISM	147	-0.212	-1.171	0.156	0.356	0.979
P53_PATHWAY	157	-0.183	-1.031	0.360	0.710	1.000

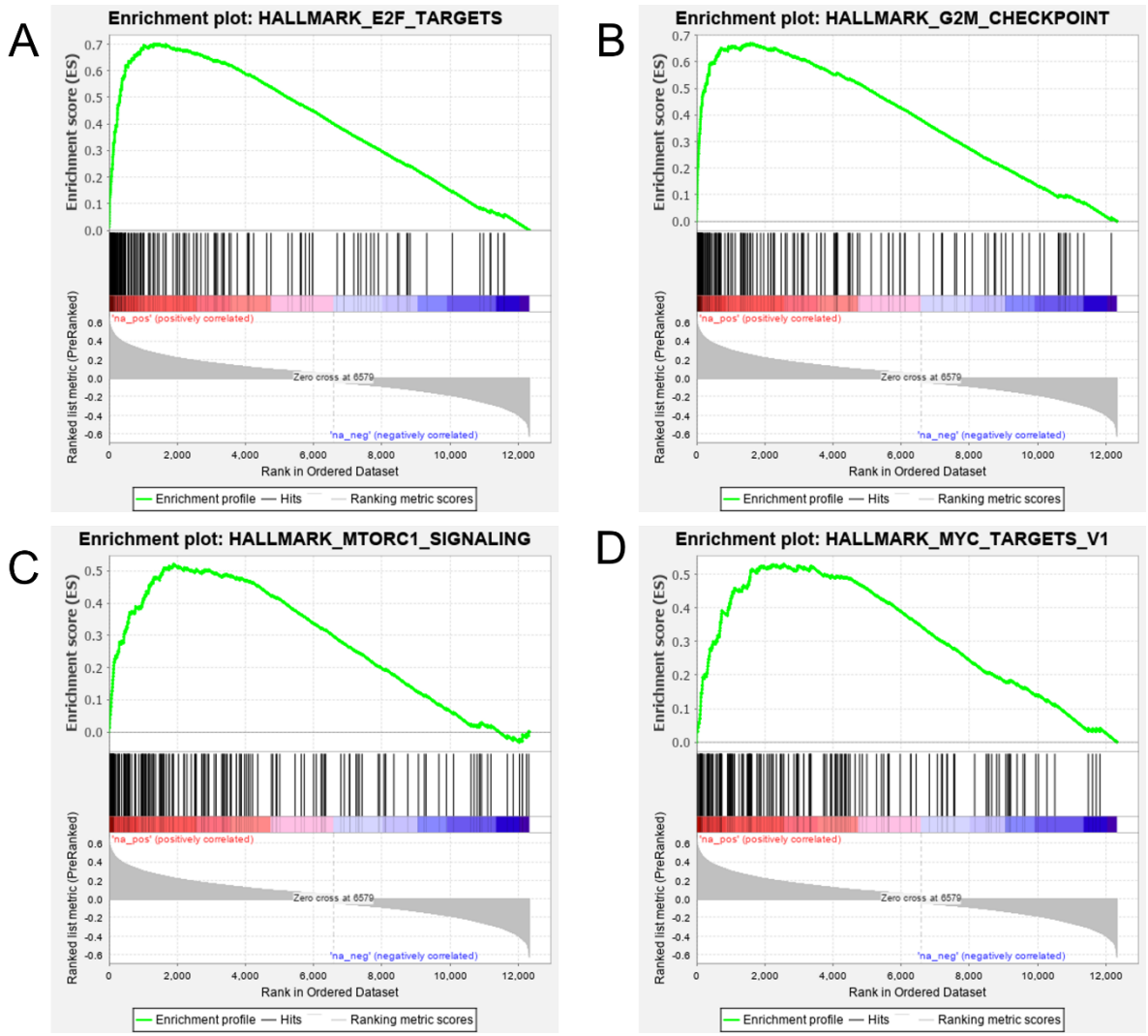

Figure 4.13. Positively correlated GSEA hallmark enrichment plots in breast cancer, using METABRICS database.

The black lines indicate the different genes within the hallmark set. Lines within the red section indicate genes positively correlated with *CBX2* while the blue section show inversely correlated with *CBX2* expression. A) FDR = <0, NES = -2.8874424, Nominal P-value = <0. B) FDR = <0, NES = -2.0190015, Nominal P-value = <0. Black lines in the red area are genes positively correlated, while lines in the blue are genes negatively correlated. Genes are ranked by spearman's coefficient. A and B are hallmarks enrichment points that are negatively correlated by *CBX2*.

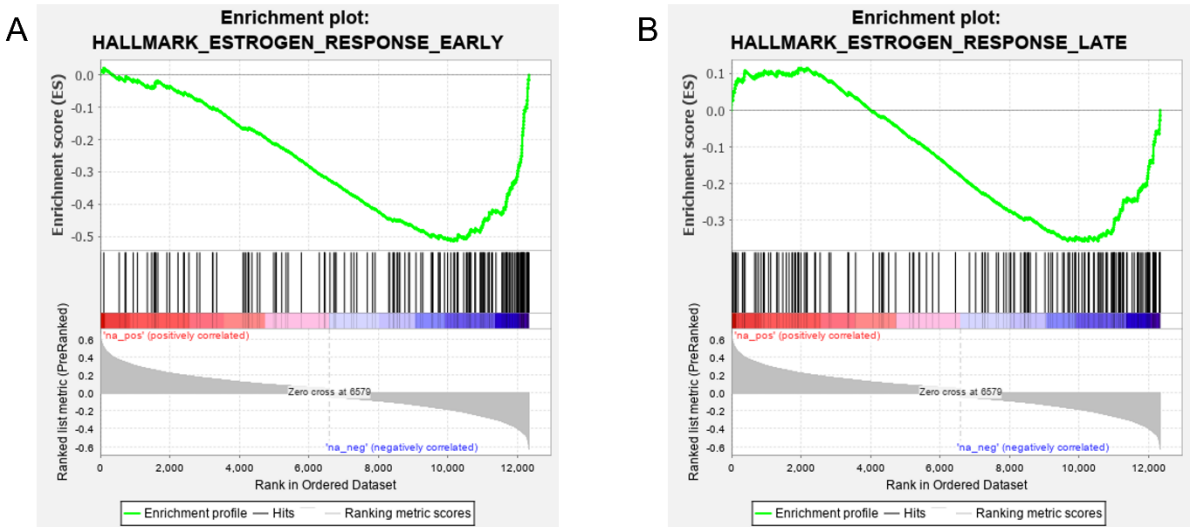

Figure 4.14. Negatively correlated GSEA hallmark enrichment plots in breast cancer, using METABRICS database.

GSEA of METABRICS database. The black lines indicate the different genes within the hallmark set. Lines within the red section indicate genes positively correlated with *CBX2* while the blue section show inversely correlated with *CBX2* expression. A) FDR = <0, NES = -2.8874424, Nominal P-value = <0. B) FDR = <0, NES = -2.0190015, Nominal P-value = <0. Black lines in the red area are genes positively correlated, while lines in the blue are genes negatively correlated. Genes are ranked by spearman's coefficient. A and B are hallmarks enrichment points that are negatively correlated by *CBX2*.

4.2.2.2. Oncogenic gene signature enrichment plots for breast cancer

GSEA identified several oncogenic gene signature sets enriched for genes identified to be positively and inversely correlated with *CBX2* expression from the TCGA Pancancer database (Table 29 and Table 30). Gene sets enriched for genes positively correlated with *CBX2* expression, indicated by the positive NES. Five oncogenic signature gene sets were apparent in the database which were RSP14_DN.V1_DN (genes downregulated after RPS14 knockdown, which is a protein involved in the small ribosome subunit [40s]), RB_P107_DN.V1_UP (genes upregulated in keratinocytes from mice with RB1 and RBL1 knockdown, which are involved in E2F signalling), E2F1_UP.V1_UP (genes upregulated in mouse fibroblast which overexpress E2F1), RAF_UP.V1_DN (genes upregulated in MCF-7 cells positive for ESR1) and MTOR_UP.N4.V1_DN (genes downregulated in CEM-C1 cells with Rapamycin, an mTOR pathway inhibitor) (Subramanian *et al*., 2005; Liberzon *et al*., 2015). These gene sets are implicated in E2F and mTORC1 signalling. Positive NES are RPS14_DN.V1_DN (NES = 4.403209), RB_P107_DN.V1_UP (NES = 4.863799) and E2F1_UP.V1_UP (NES = 3.889796) (Figure 4.15), which indicate and association between CBX2 and E2F signalling. Figure 4.15 also shows oncogenic gene signature sets which are

enriched for genes inversely correlated with *CBX2* expression. Negative NES include RAF_UP.V1_DN (NES = -4.3853335) and MTOR_UP.N4.V1_DN (NES = -4.2671657). All q-values were less than 0.001.

Table 29. Top 10 oncogenic signatures with a positive enrichment score in TCGA Pancancer database.

Name of Oncogenic Signature	Number of Genes	ES	NES	Nominal p-value	FDR q-value	FWER p-Value
CSR_LATE_UP.V1_UP	151	0.378	5.419	<0.001	<0.001	<0.001
RB_P107_DN.V1_UP	107	0.402	4.864	<0.001	<0.001	<0.001
RPS14_DN.V1_DN	152	0.309	4.403	<0.001	<0.001	<0.001
E2F1_UP.V1_UP	154	0.270	3.890	<0.001	<0.001	<0.001
MYC_UP.V1_UP	148	0.274	3.861	<0.001	<0.001	<0.001
CSR_EARLY_UP.V1_UP	118	0.293	3.715	<0.001	<0.001	<0.001
GCNP_SHH_UP_LATE.V1_UP	151	0.252	3.625	<0.001	<0.001	<0.001
MTOR_UP.N4.V1_UP	169	0.238	3.551	<0.001	<0.001	<0.001
HOXA9_DN.V1_DN	155	0.240	3.497	<0.001	<0.001	<0.001
PRC2_EZH2_UP.V1_DN	151	0.215	3.023	<0.001	<0.001	<0.001

Table 30. Top 10 oncogenic signatures with a negative enrichment score in TCGA Pancancer database.

Name of Oncogenic Signature	Number of Genes	ES	NES	Nominal p-value	FDR q-value	FWER p-Value
RAF_UP.V1_DN	153	-0.310	-4.385	<0.001	<0.001	<0.001
MTOR_UP.N4.V1_DN	138	-0.316	-4.267	<0.001	<0.001	<0.001
CSR_EARLY_UP.V1_DN	103	-0.275	-3.222	<0.001	<0.001	<0.001
STK33_UP	220	-0.161	2.728	<0.001	0.001	0.004
LEF1_UP.V1_DN	150	-0.187	-2.649	<0.001	0.002	0.009
KRAS.KIDNEY_UP.V1_UP	99	-0.220	-2.584	<0.001	0.002	0.015
ESC_J1_UP_EARLY.V1_UP	138	-0.186	-2.559	<0.001	0.002	0.015
MEK_UP.V1_DN	153	-0.178	-2.548	<0.001	0.002	0.016
PGF_UP.V1_UP	147	-0.171	-2.423	<0.001	0.004	0.032
CSR_LATE_UP.V1_DN	108	-0.198	-2.388	<0.001	0.004	0.041

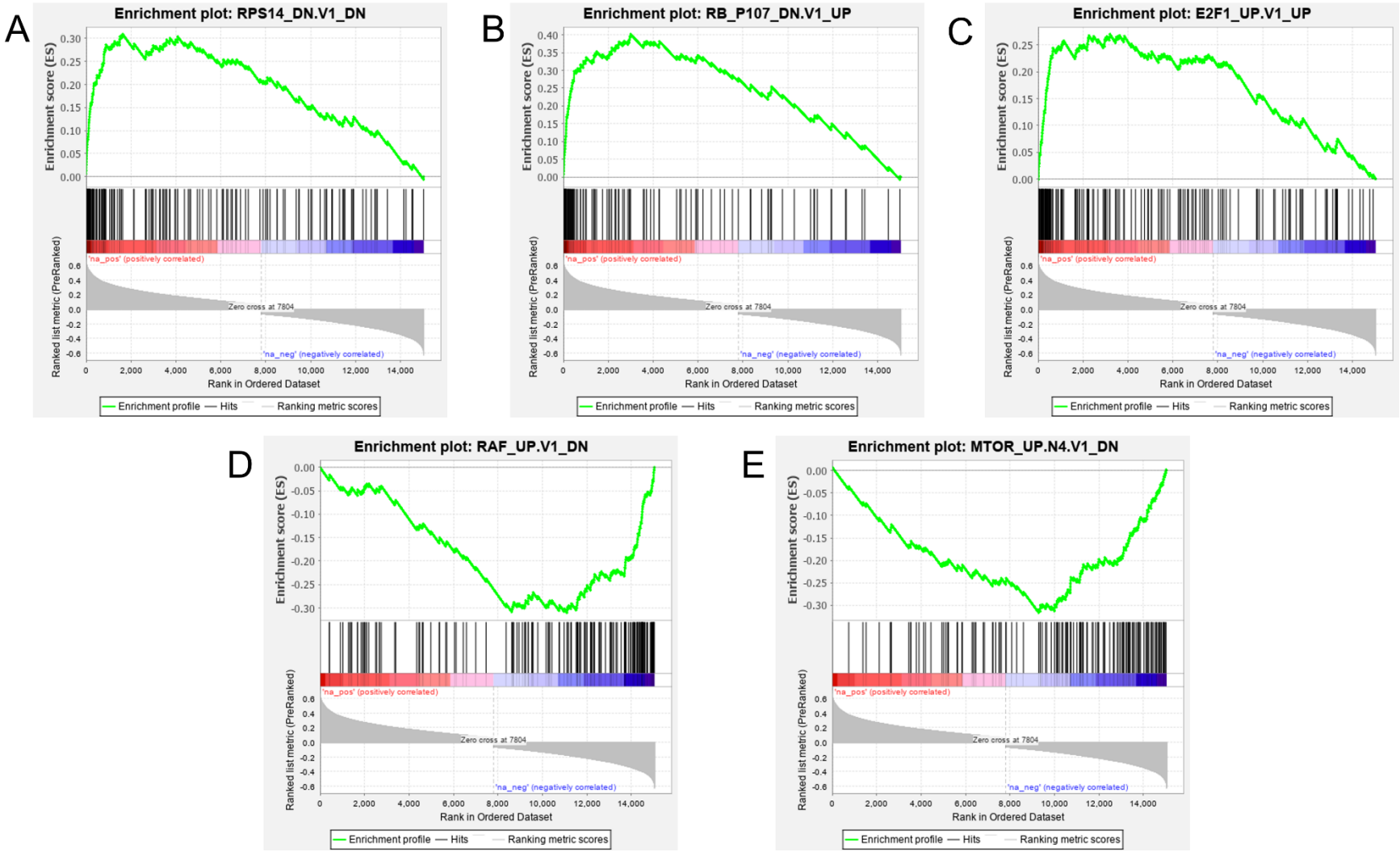

Figure 4.15. Positively and negatively correlated GSEA oncogenic gene signatures enrichment plots in breast cancer using TCGA, Pancancer database. The black lines indicate the different genes within the hallmark set. Lines within the red section indicate genes positively correlated with *CBX2* while the blue section show inversely correlated with *CBX2* expression. Genes are ranked by spearman's coefficient. A-C are oncogenic gene signature enrichment points that are positively correlated by CBX2, while D and E are negatively correlated. A-C are oncogenic gene signature enrichment points that are positively correlated by CBX2, while D and E are negatively correlated. A) FDR = <0, NES = 4.403209, Nominal P-value = <0. B) FDR = <0, NES = 4.863799, Nominal P-value = <0. C) FDR = <0, NES = 3.889796, Nominal P-value = <0. D) FDR = <0, NES = -4.3853335, Nominal P-value = <0. E) FDR = <0, NES = -4.2671657, Nominal P-value = <0. Genes are ranked by spearman's coefficient. A-C are oncogenic gene signature enrichment points that are positively correlated by CBX2, while D and E are negatively correlated.

GSEA identified several oncogenic gene signature sets, identified to be positively and inversely correlated with *CBX2* expression from the METABRICS database (Table 31 and Table 32). The same five oncogenic signature gene set identified in TCGA, Pancancer were identified in the METABRICS database. Positive NES are RB_P107_DN.V1_UP (NES = 2.8418398), RPS14_DN.V1_DN (NES = 2.990342) and E2F1_UP.V1_UP (NES = 2.5727172). Figure 4.16 also shows oncogenic gene signature sets which are enriched for genes inversely correlated with *CBX2* expression and associated with cancer development or progression. Gene sets include RAF_UP.V1_DN (NES = -2.5457542) and MTOR_UP.N4.V1_DN (NES = -2.3712335). All p-values were less than 0.01. Again, this demonstrates parity in identified gene sets across two different breast cancer patient databases. All q-values were less than 0.001.

The GSEA indicated hallmarks and oncogenic gene sets which are involved in in E2F an mTORC1 signalling e.g., E2F signalling, mTORC1 signalling, E2F1_UP.V1_UP and MTOR_UP.N4.V1_DN. This means CBX2 correlates with genes involved in the E2F signalling and mTORC1 activation, which suggest CBX2 may be necessary for activation of those pathways.

Table 31. Top 10 oncogenic signatures with a positive enrichment score in METABRICS database.

Name of Oncogenic Signature	Number of Genes	ES	NES	Nominal p-value	FDR q-value	FWER p-Value
RPS14_DN.V1_DN	135	0.564	2.990	<0.001	<0.001	<0.001
CSR_LATE_UP.V1_UP	128	0.556	2.928	<0.001	<0.001	<0.001
RB_P107_DN.V1_UP	106	0.555	2.842	<0.001	<0.001	<0.001
E2F1_UP.V1_UP	148	0.479	2.573	<0.001	<0.001	<0.001
GCNP_SHH_UP_LATE.V1_UP	138	0.446	2.373	<0.001	<0.001	<0.001
MTOR_UP.V1_UP	127	0.449	2.371	<0.001	<0.001	<0.001
CSR_EARLY_UP.V1_UP	115	0.454	2.330	<0.001	<0.001	<0.001
HOXA9_DN.V1_DN	132	0.426	2.235	<0.001	<0.001	<0.001
VEGF_A_UP.V1_DN	145	0.415	2.225	<0.001	<0.001	<0.001
PRC2_EZH2_UP.V1_DN	130	0.422	2.213	<0.001	<0.001	<0.001

Table 32. Top 10 oncogenic signatures with a negative enrichment score in METABRICS database.

Name of Oncogenic Signature	Number of Genes	ES	NES	Nominal p-value	FDR q-value	FWER p-Value
RAF_UP.V1_DN	147	-0.470	-2.546	<0.001	<0.001	<0.001
MTOR_UP.N4.V1_DN	117	-0.454	-2.456	<0.001	<0.001	<0.001
LEF1_UP.V1_DN	136	-0.377	-2.033	<0.001	0.001	0.004
ESC_J1_UP_EARLY.V1_UP	130	-0.347	-1.891	<0.001	0.004	0.017
MEK_UP.V1_DN	137	-0.337	-1.864	<0.001	0.004	0.022
KRAS.KIDNEY_UP.V1_UP	97	-0.347	-1.826	<0.001	0.007	0.043
CSR_EARLY_UP.V1_DN	102	-0.335	-1.785	0.002	0.011	0.072
EGFR_UP.V1_DN	135	-0.321	-1.778	<0.001	0.011	0.082
STK33_UP	187	-0.302	-1.733	<0.001	0.016	0.133
CSR_LATE_UP.V1_DN	101	-0.311	-1.633	0.007	0.035	0.298

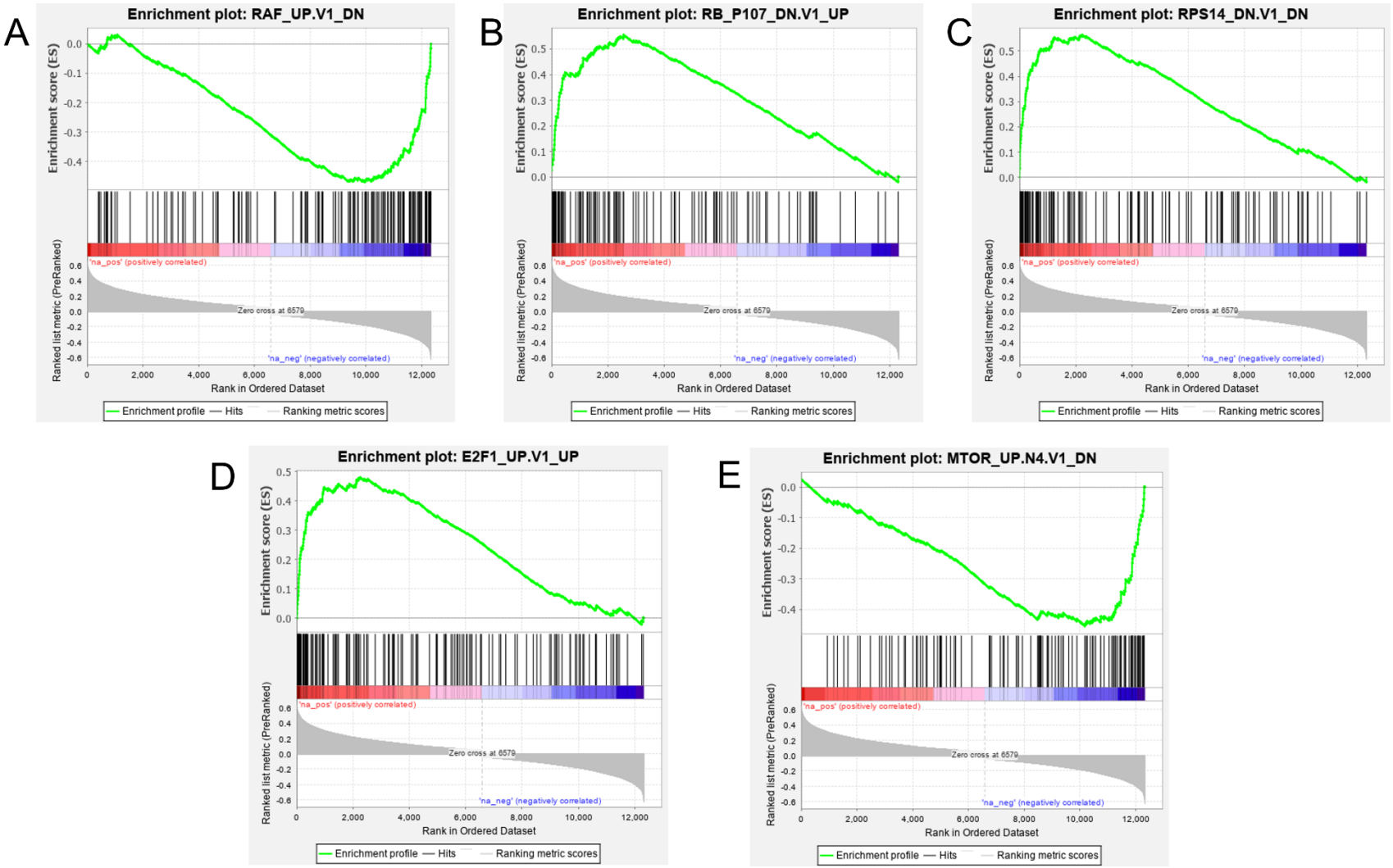

Figure 4.16. Positively and negatively correlated GSEA oncogenic gene signatures enrichment plots in breast cancer, using METABRICS database. The black lines indicate the different genes within the hallmark set. Lines within the red section indicate genes positively correlated with *CBX2* while the blue section show inversely correlated with *CBX2* expression. Black lines in the red area are genes positively correlated, while lines in the blue are genes negatively correlated. Genes are ranked by spearman's coefficient. B and C are oncogenic gene signature enrichment points that are positively correlated by CBX2, while A, D and E are negatively correlated. A) FDR = <0, NES = -2.5457542, Nominal P-value = <0. B) FDR = <0, NES = 2.8418398, Nominal P-value = <0. C) FDR = <0, NES = 2.990342, Nominal P-value = <0. D) FDR = <0, NES = 2.5727172, Nominal P-value = <0. E) FDR = <0, NES = -2.3712335, Nominal P-value= <0.

4.2.3. Hallmark and oncogenic signature enrichment plots for RNA-seq data

From the analysis done within the two different breast cancer databases, there is a clear indication of *CBX2* having a role in breast cancer development and progression, with *CBX2* being identified to have a role in G2/M checkpoint, and E2F and mTORC1 signalling, all of which are important for cancer cell growth. To further our research, an analysis using RNA-seq data (provided by Dr Mark Wade), was conducted. The gene lists used in the analysis were genes significantly differentially expressed in MDA-MB-231 cells (a TNBC cell line) which had been depleted of *CBX2* (via siRNA mediated *CBX2* knockdown) compared with control MDA-MB-231 cells (cells transfected with a non-silencing siRNA (siScr)). These genes were ranked by the relative log2 fold increase (positive number) or decrease (negative number) in gene expression following *CBX2* knockdown.

GSEA from differentially expressed genes is shown in Tables 33 and 34. Some gene sets enriched with genes that were downregulated following CBX2 knockdown, indicated by the negative NES (Table 33), were similar to those positively correlated with *CBX2* gene expression in patient datasets. These include MYC targets (NES = -1.7361276), MTORC1 signalling (NES = -1.484452), G2M checkpoint (NES = -1.476041) and E2F targets (NES = -1.2057885). All p-values were less than 0.05, except for E2F targets, which had a p-value of 0.16171618 (Figure 4.17). This indicates that gene sets putatively regulated by CBX2 in the patient datasets also appear to be regulated by CBX2 in a cell line model of TNBC.

Table 33. Top 9 hallmark gene sets for genes upregulated following CBX2 knockdown in MDA-MB-231 cells.

Name of Hallmark	Number of Genes	ES	NES	Nominal p-value	FDR q-value	FWER p-Value
KRAS_SIGNALING_DN	35	0.319	1.374	0.096	0.791	0.949
COAGULATION	41	0.286	1.234	0.176	0.819	0.995
UV_RESPONSE_DN	71	0.205	1.027	0.423	1.000	1.000
GLYCOLYSIS	93	0.190	1.013	0.409	0.997	1.000
APICAL_JUNCTION	74	0.188	0.952	0.544	0.963	1.000
P53_PATHWAY	93	0.160	0.849	0.717	1.000	1.000
TGF_BETA_SIGNALING	22	0.225	0.839	0.681	0.918	1.000
PEROXISOME	37	0.177	0.762	0.810	0.925	1.000
UNFOLDED_PROTEIN_RESPONSE	44	0.131	0.589	0.946	0.963	1.000

Table 34. Top 10 hallmark gene sets for genes downregulated following CBX2 knockdown in MDA-MB-231 cells.

Name of Hallmark	Number of Genes	ES	NES	Nominal p-value	FDR q-value	FWER p-Value
INTERFERON_GAMMA_RESPONSE	72	-0.531	-3.081	<0.001	<0.001	<0.001
INFLAMMATORY_RESPONSE	64	-0.523	-2.861	<0.001	<0.001	<0.001
IL6_JAK_STAT3_SIGNALING	37	-0.604	-2.851	<0.001	<0.001	<0.001
TNFA_SIGNALING_VIA_NFKB	96	-0.453	-2.750	<0.001	<0.001	<0.001
INTERFERON_ALPHA_RESPONSE	43	-0.517	-2.594	<0.001	<0.001	<0.001
ALLOGRAFT_REJECTION	54	-0.424	-2.218	<0.001	0.001	0.003
KRAS_SIGNALING_UP	69	-0.387	-2.158	<0.001	0.002	0.004
WNT_BETA_CATENIN_SIGNALING	19	-0.541	-2.092	<0.001	0.003	0.008
UV_RESPONSE_UP	61	-0.352	-1.918	0.006192	0.008	0.032
MYC_TARGETS_V1	76	-0.304	-1.736	0.006757	0.026	0.104

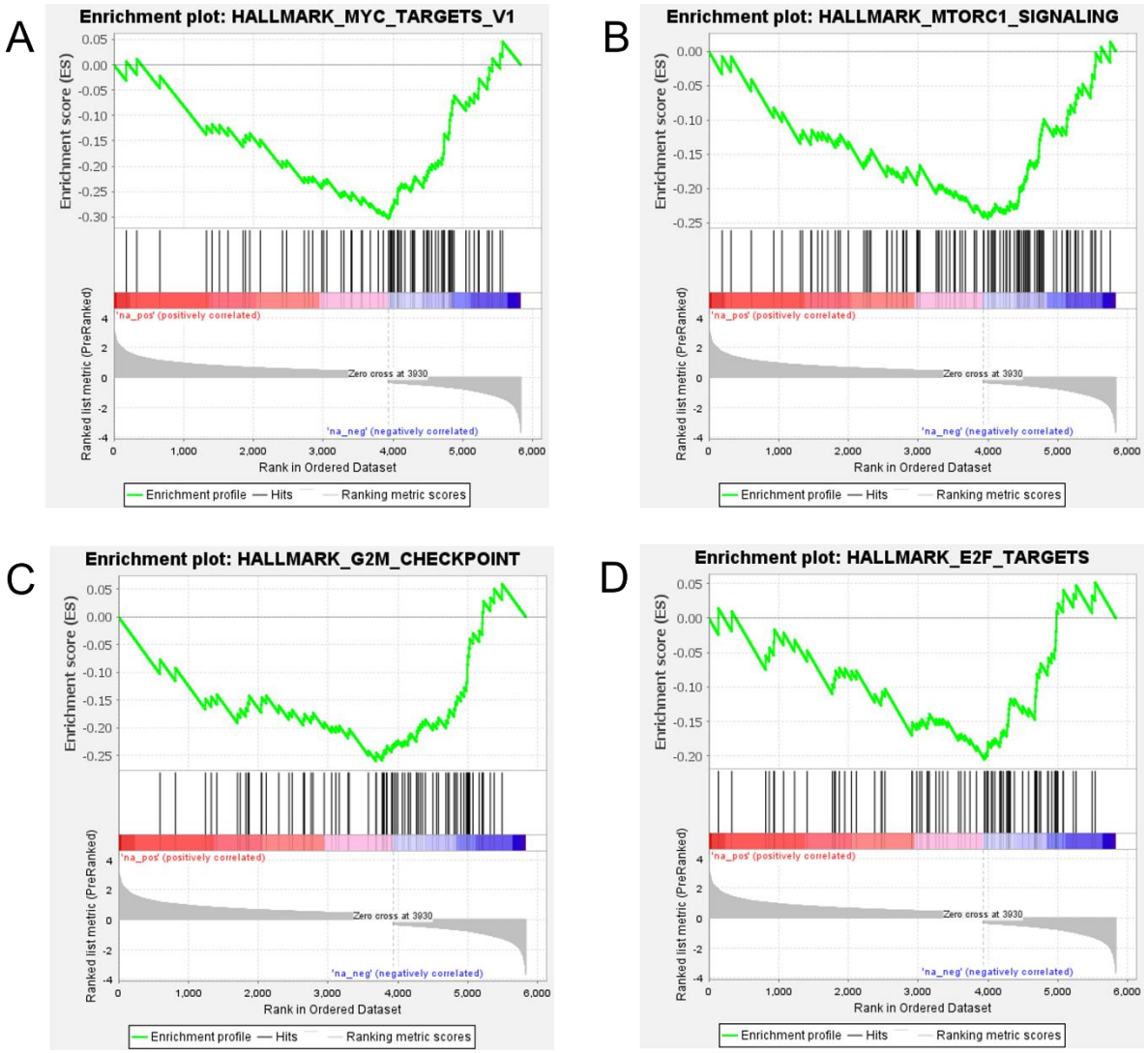

Figure 4.17. GSEA of differentially expressed genes following CBX2 knockdown in MDA-MB-231 cells. The black lines indicate the different genes within the hallmark set. Lines within the red section indicate genes upregulated following *CBX2* knockdown while the blue section shows genes downregulated following *CBX2* knockdown. A) FDR = 0.026387203, NES = -1.7361276, Nominal P-value = 0.006756757. B) FDR = 0.086029634, NES = -1.484452, Nominal P-value = 0.015209125. C) FDR = 0.08466947, NES = -1.476041, Nominal P-value = 0.027303753. D) FDR = 0.30009237, NES = -1.2057885, Nominal P-value = 0.16171618.

4.2.4. Hallmark and oncogenic signature enrichment plots for GBM and LUAD.

Our GSEA analysis of patient cancer datasets and CBX2 knockdown RNA-seq analysis shows that *CBX2* has a role in cancer development and progression, with MYC, E2F and mTORC1 pathways identified as potentially being regulated by CBX2. To expand our analysis, GSEA of genes positively and negatively correlated with CBX2 expression was undertaken in two other cancer datasets, glioblastoma and lung adenocarcinoma. Pre-ranked gene lists of genes positively and negatively correlated with *CBX2* expression identified by cBioPortal analysis (Section 4.2) were analysed against the hallmarks gene sets and oncogenic gene signature gene set via GSEA.

GSEA identified many hallmark gene sets enriched for genes identified to be positively and inversely correlated with CBX2 expression for GBM from the TCGA Pancancer database (Table 35 and Table 36). Gene sets were enriched with genes that positively correlate with CBX2 expression, indicated by the positive NES. Figure 4.18 shows chosen hallmark gene set plots that are associated with cancer development or progression, including E2F signalling (NES = 4.685634), G2M checkpoint (NES = 4.8801866) and MYC targets (NES = 2.483329). GSEA identified several oncogenic gene signature sets, which identified genes to be positively and inversely correlated with CBX2 expression for GBM, from the TCGA Pancancer database (Table 37 and Table 38). Figure 4.18 also shows chosen oncogenic gene signature plots including RPS14_DN.V1.DN (NES = 3.4131367) and RB_P107_DN.V1_UP (NES = 3.2788813). q-values were all less than 0.001.

Table 35. Top 10 hallmark gene sets with a positive enrichment score in GBM, TCGA Pancancer dataset.

Name of Hallmark	Number of Genes	ES	NES	Nominal p-value	FDR q-value	FWER p-Value
GSM_CHECKPOINT	168	0.664	4.880	<0.001	<0.001	<0.001
E2F_TARGETS	166	0.636	4.686	<0.001	<0.001	<0.001
MITOTIC_SPINDLE	136	0.526	3.725	<0.001	<0.001	<0.001
MYC_TARGETS_V1	123	0.359	2.483	<0.001	<0.001	0.001
MYC_TARGETS_V2	37	0.476	2.389	<0.001	<0.001	0.001
WNT_BETA_CATENIN_SIGNALING	28	0.476	2.215	<0.001	<0.001	0.003
SPERMATOGENESIS	61	0.299	1.752	0.008	0.011	0.173
DNA_REPAIR	106	0.244	1.622	0.014	0.029	0.451
HEDGEHOG_SIGNALING	19	0.360	1.456	0.080	0.083	0.854
PANCREAS_BETA_CELLS	29	0.219	1.007	0.442	0.630	1.000

Table 36. Top 10 hallmark gene sets with a negative enrichment score in GBM, TCGA Pancancer dataset.

Name of Hallmark	Number of Genes	ES	NES	Nominal p-value	FDR q-value	FWER p-Value
ALLOGRAFT_REJECTION	139	-0.569	-4.699	<0.001	<0.001	<0.001
INTERFERON_GAMMA_RESPONSE	155	-0.559	-4.672	<0.001	<0.001	<0.001
INFLAMMATORY_RESPONSE	154	-0.563	-4.617	<0.001	<0.001	<0.001
OXIDATIVE_PHOSPHORYLATION	140	-0.509	-4.106	<0.001	<0.001	<0.001
INTERFERON_ALPHA_RESPONSE	73	-0.582	-4.028	<0.001	<0.001	<0.001
COMPLEMENT	147	-0.481	-3.910	<0.001	<0.001	<0.001
IL6_JAK_STATS3_SIGNALING	66	-0.565	-3.866	<0.001	<0.001	<0.001
TNFA_SIGNALING_VIA_NFKB	132	-0.475	-3.787	<0.001	<0.001	<0.001
IL2_STATS_SIGNALING	132	-0.439	-3.558	<0.001	<0.001	<0.001
EPITHELIAL_MESENCHYMAL_TRANSITION	104	-0.455	-3.548	<0.001	<0.001	<0.001

Table 37. Top 10 oncogenic signatures positively correlated in GBM, TCGA Pancancer.

Name of Oncogenic Signature	Number of Genes	ES	NES	Nominal p-value	FDR q-value	FWER p-Value
RPS14_DN.V1_DN	113	0.503	3.413	<0.001	<0.001	<0.001
RB_P107_DN.V1_UP	90	0.511	3.279	<0.001	<0.001	<0.001
PRC2_EED_UP.V1_DN	124	0.457	3.218	<0.001	<0.001	<0.001
E2F1_UP.V1_UP	119	0.457	3.143	<0.001	<0.001	<0.001
PRC2_EZH2_UP.V1_DN	118	0.430	2.951	<0.001	<0.001	<0.001
GCNP_SHH_UP_LATE.V1_UP	118	0.401	2.711	<0.001	<0.001	<0.001
GCNP_SHH_UP_EARLY.V1_UP	117	0.368	2.495	<0.001	<0.001	<0.001
CSR_LATE_UP.V1_UP	103	0.373	2.462	<0.001	<0.001	0.001
RB_P130_DN.V1_UP	69	0.384	2.326	<0.001	<0.001	0.002
VEGF_A_UP.V1_DN	138	0.322	2.286	<0.001	<0.001	0.002

Table 38. Top 10 oncogenic signatures negatively correlated in GBM, TCGA Pancancer.

Name of Oncogenic Signature	Number of Genes	ES	NES	Nominal p-value	FDR q-value	FWER p-Value
RPS14_DN.V1_UP	136	-0.538	-4.402	<0.001	<0.001	<0.001
HOXA9_DN.V1_UP	135	-0.496	-4.189	<0.001	<0.001	<0.001
EGFR_UP.V1_UP	119	-0.483	-3.819	<0.001	<0.001	<0.001
STK33_NOMO_UP	179	-0.433	-3.752	<0.001	<0.001	<0.001
STK33_UP	175	-0.407	-3.490	<0.001	<0.001	<0.001
STK33_SKM_UP	170	-0.398	-3.343	<0.001	<0.001	<0.001
CAHOY_ASTROGLIAL	58	-0.484	-3.142	<0.001	<0.001	<0.001
ESC_J1_UP_LATE.V1_UP	91	-0.396	-2.897	<0.001	<0.001	<0.001
RAF_UP.V1_UP	128	-0.362	-2.884	<0.001	<0.001	<0.001
RB_DN.V1_DN	74	-0.396	-2.791	<0.001	<0.001	<0.001

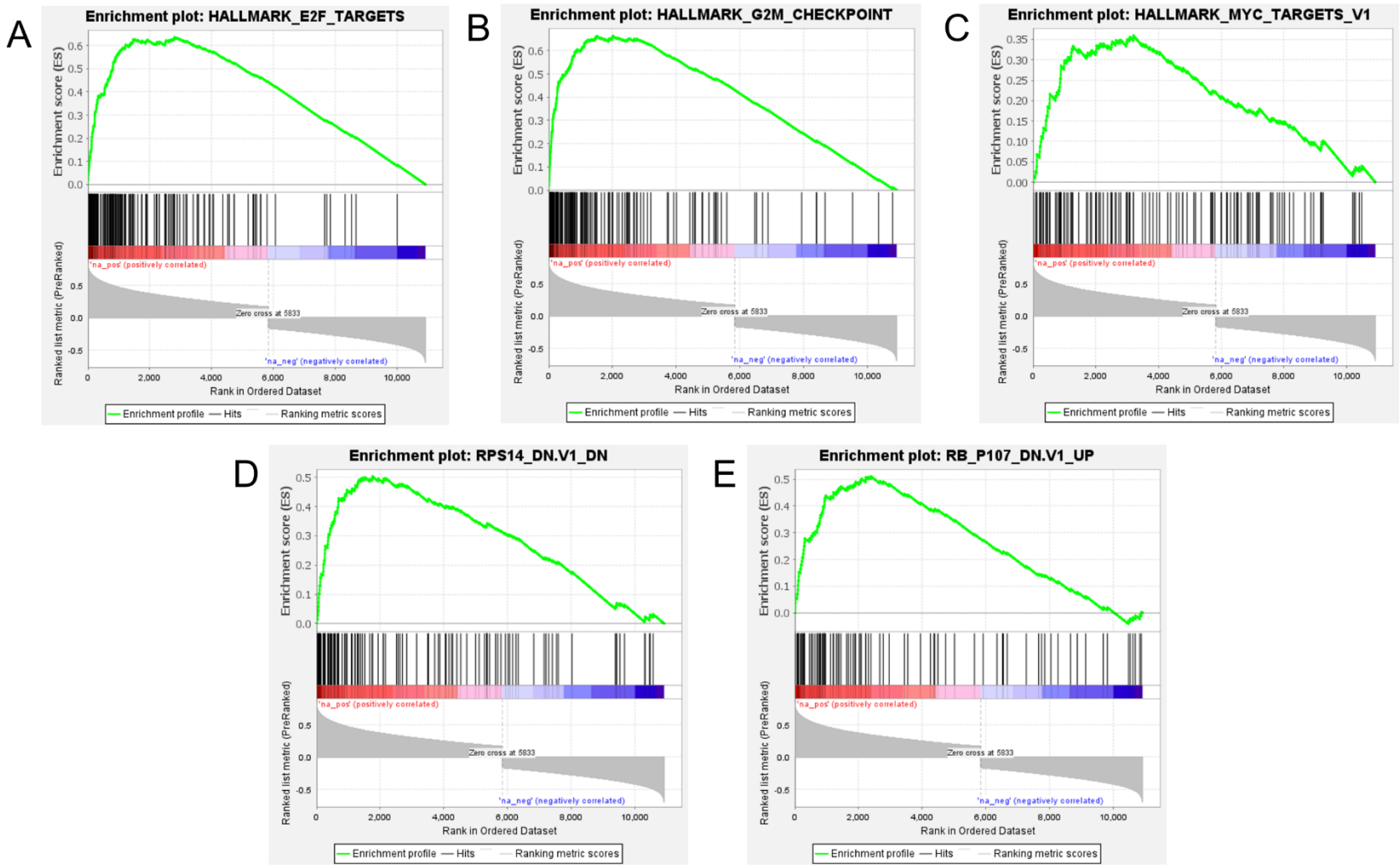

Figure 4.18. Positively correlated GSEA in Glioblastoma TCGA, Pancancer database.
The black lines indicate the different genes within the hallmark set. Lines within the red section indicate genes positively correlated with *CBX2* while the blue section show inversely correlated with *CBX2* expression. A) FDR = <0, NES = 4.685634, Nominal P-value = <0. B) FDR = <0, NES = 4.8801866, Nominal P-value = <0. C) FDR = 0.000094594594, NES = 2.483329, Nominal P-value = <0. D) FDR = <0, NES = 3.4131367, Nominal P-value = <0. E) FDR = <0, NES = 3.2788813, Nominal P-value = <0. Black lines in the red area are genes positively correlated, while lines in the blue are genes negatively correlated. Genes are ranked by spearman's coefficient. A-C are hallmark enrichment plots, while D-E are oncogenic gene signature enrichment plots. All enrichment points that are positively correlated by CBX2.

GSEA identified many hallmark gene sets enriched for genes identified to be positively and inversely correlated with CBX2 expression for LUAD from the TCGA Pancancer database (Table 39 and Table 40). Most of the gene sets were enriched with genes that positively correlate with CBX2 expression, indicated by the positive NES. Figure 4.19 shows chosen hallmark gene set plots that are associated with cancer development or progression, including E2F signalling (NES = 5.119062), G2M checkpoint (NES = 4.776148), MYC targets (NES = 4.440755) and MTORC1 signalling (NES = 3.6358624). GSEA identified several oncogenic gene signature sets, which identified genes to be positively and inversely correlated with CBX2 expression for GBM, from the TCGA Pancancer database (Table 41 and Table 42). The figure also shows chosen oncogenic gene signature plots including RPS14_DN.V1.DN (NES = 3.7759283), RB_P107_DN.V1_UP (NES = 3.3538637) and E2F1_UP.V1_UP (NES = 3.3812227). GSEA also identified several oncogenic gene signature plots that were negatively correlated with CBX2 expression, indicated by the negative NES. Figure 4.19 shows chosen negative oncogenic gene signature plots including: MTOR_UP.N4.V1_DN (NES = -2.6951063). All q-values were less than 0.01. This indicates that gene sets identified in GBM and LUAD were also identified in the breast cancer datasets, implying *CBX2* may have a role in these signatures across different cancer types.

Table 39. Top 10 hallmark gene sets positively correlated in LUAD, TCGA Pancancer.

Name of Hallmark	Number of Genes	ES	NES	Nominal p-value	FDR q-value	FWER p-Value
E2F_TARGETS	166	0.695	5.119	<0.001	<0.001	<0.001
G2M_CHECKPOINT	165	0.644	4.776	<0.001	<0.001	<0.001
MYC_TARGETS_V1	160	0.597	4.441	<0.001	<0.001	<0.001
MYC_TARGETS_V2	51	0.707	4.021	<0.001	<0.001	<0.001
MTORC1_SIGNALING	146	0.496	3.636	<0.001	<0.001	<0.001
UNFOLDED_PROTEIN_RESPONSE	71	0.474	2.938	<0.001	<0.001	<0.001
MITOTIC_SPINDLE	122	0.370	2.588	<0.001	<0.001	<0.001
REACTIVE_OXYGEN_SPECIES_PATHWAY	39	0.479	2.446	<0.001	<0.001	<0.001
OXIDATIVE_PHOSPHORYLATION	128	0.341	2.383	<0.001	<0.001	<0.001
SPERMATOGENESIS	69	0.392	2.376	<0.001	<0.001	<0.001

Table 40. Top 10 hallmark gene sets negatively correlated in LUAD, TCGA Pancancer.

Name of Hallmark	Number of Genes	ES	NES	Nominal p-value	FDR q-value	FWER p-Value
ALLOGRAFT_REJECTION	138	-0.501	-3.787	<0.001	<0.001	<0.001
INFLAMMATORY_RESPONSE	134	-0.475	-3.658	<0.001	<0.001	<0.001
INTERFERON_GAMMA_RESPONSE	128	-0.422	-3.240	<0.001	<0.001	<0.001
KRAS_SIGNALING_UP	129	-0.422	-3.157	<0.001	<0.001	<0.001
IL6_JAK_STAT3_SIGNALING	58	-0.494	-3.007	<0.001	<0.001	<0.001
UV_RESPONSE_DN	88	-0.432	-2.985	<0.001	<0.001	<0.001
INTERFERON_ALPHA_RESPONSE	64	-0.448	-2.830	<0.001	<0.001	<0.001
THFA_SIGNALING_VIA_NFKB	122	-0.382	-2.781	<0.001	<0.001	<0.001
IL2_STATS_SIGNALING	134	-0.350	-2.729	<0.001	<0.001	<0.001
COMPLEMENT	120	-0.323	-2.359	<0.001	<0.001	<0.001

Table 41. Top 10 oncogenic signatures positively correlated in LUAD, TCGA Pancancer.

Name of Oncogenic Signature	Number of Genes	ES	NES	Nominal p-value	FDR q-value	FWER p-Value
RPS14_DN.V1_DN	122	0.540	3.776	<0.001	<0.001	<0.001
NFE2L2.V2	290	0.441	3.620	<0.001	<0.001	<0.001
CSR_LATE_UP.V1_UP	125	0.498	3.512	<0.001	<0.001	<0.001
E2F1_UP.V1_UP	122	0.477	3.381	<0.001	<0.001	<0.001
RB_P107_DN.V1_UP	100	0.497	3.354	<0.001	<0.001	<0.001
PRC2_EZH2_UP.V1_DN	121	0.423	2.960	<0.001	<0.001	<0.001
GCNP_SHH_UP_LATE.V1_UP	120	0.404	2.826	<0.001	<0.001	<0.001
MYC_UP.V1_UP	107	0.405	2.740	<0.001	<0.001	<0.001
GCNP_SHH_UP_EARLY.V1_UP	101	0.393	2.644	<0.001	<0.001	<0.001
PRC2_EED_UP.V1_DN	115	0.377	2.588	<0.001	<0.001	<0.001

Table 42. Top 10 oncogenic signatures negatively correlated in LUAD, TCGA Pancancer.

Name of Oncogenic Signature	Number of Genes	ES	NES	Nominal p-value	FDR q-value	FWER p-Value
RPS14_DN.V1_UP	133	-0.574	-4.356	<0.001	<0.001	<0.001
STK33_SKM_UP	155	-0.418	-3.338	<0.001	<0.001	<0.001
STK33_UP	163	-0.404	-3.213	<0.001	<0.001	<0.001
HOXA9_DN.V1_UP	116	-0.434	-3.194	<0.001	<0.001	<0.001
STK33_NOMO_UP	167	-0.394	-3.120	<0.001	<0.001	<0.001
ESC_J1_UP_LATE.V1_UP	117	-0.410	-3.007	<0.001	<0.001	<0.001
MTOR_UP.N4.V1_DN	104	-0.379	-2.695	<0.001	<0.001	<0.001
EGFR_UP.V1_UP	111	-0.368	-2.668	<0.001	<0.001	<0.001
LEF1_UP.V1_DN	122	-0.354	-2.612	<0.001	<0.001	<0.001
RAF_UP.V1_UP	116	-0.349	-2.576	<0.001	<0.001	<0.001

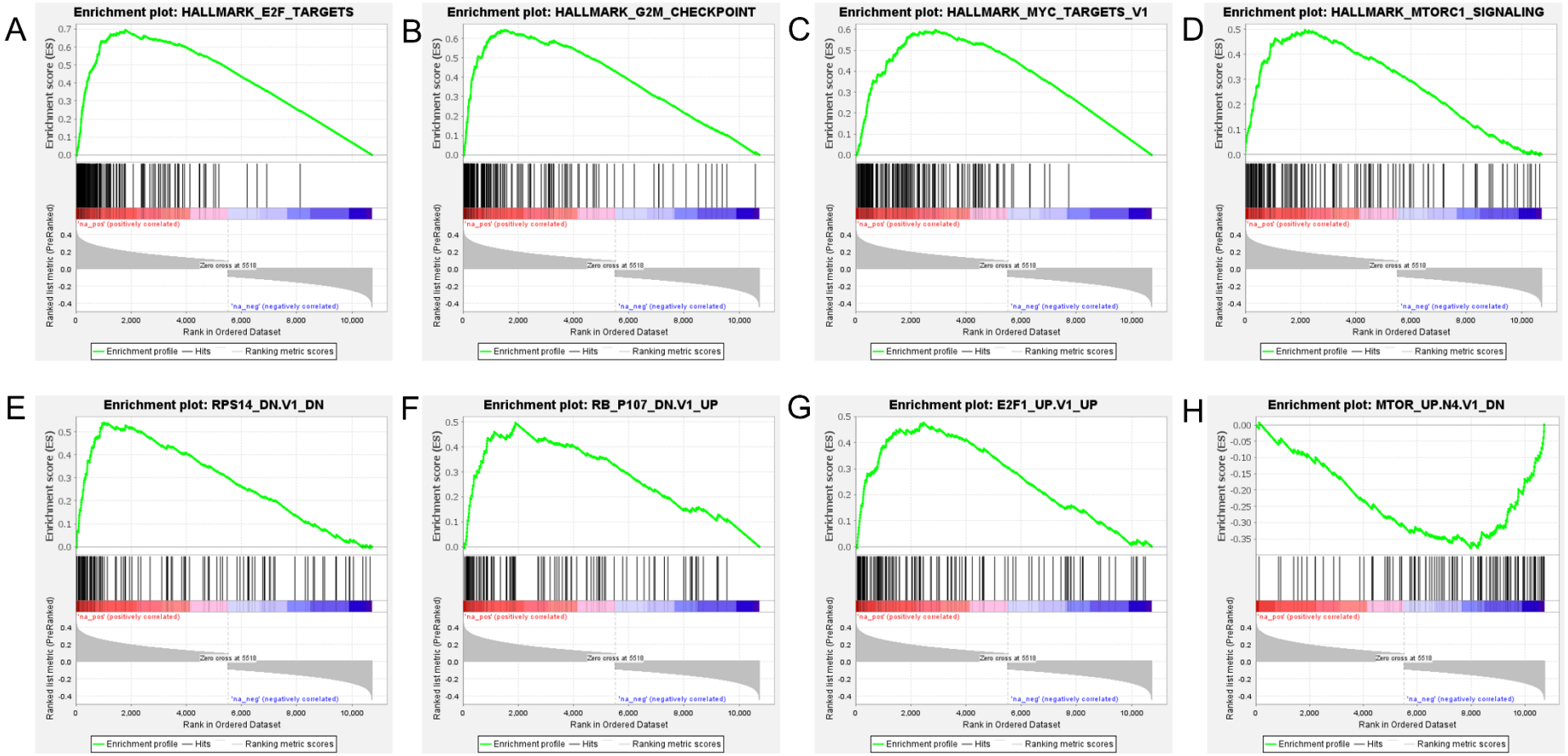

Figure 4.19. GSEA of Lung adenocarcinoma TCGA, Pancancer database.

Black lines in the red area are genes positively correlated, while lines in the blue are genes negatively correlated. Genes are ranked by spearman's coefficient. A-D are hallmark enrichment plots, while E-H are oncogenic gene signature enrichment plots. A-G enrichment points that are positively correlated by CBX2, while H is negatively correlated. A) FDR = <0, NES = 5.119062, Nominal P-value = <0. B) FDR = <0, NES = 4.776148, Nominal P-value = <0. C) FDR = <0, NES = 4.440755, Nominal P-value = <0. D) FDR = <0, NES = 3.6358624, Nominal P-value = <0. E) FDR = <0, NES = 3.7759283, Nominal P-value = <0. F) FDR = 0, NES = 3.3538637, Nominal P-value = <0. G) FDR = 0, NES = 3.3812227, Nominal P-value = <0. H) FDR = <0, NES = -2.6951063, Nominal P-value = <0.

4.3. In vitro analysis of CBX2 binding at putative tumour suppressor genes

From the GSEA analysis, both E2F and mTORC1 signalling has been identified as potential oncogenic signalling pathways regulated by CBX2 in multiple cancer types. While these pathways have been identified, an investigation was needed to see if CBX2 directly regulates them. From the RNA-seq data, tumour suppressor genes involved in regulating these pathways were identified to be upregulated following *CBX2* knockdown (figure 4.20). These include *RBL2* which is a negative regulator of E2F signalling and *TSC1* and *PRKAA2* which are negative regulators of mTORC1 signalling. CBX2 is known to be part of the PRC1 complex (Section 2.2) which classically is a transcriptional repressor complex. It is, therefore, possible that CBX2 directs PRC1 to the loci of these tumour suppressor genes to repress their expression in cancer and therefore promotes oncogenic E2F and mTORC1 signalling. To assess if CBX2 is directly affecting the expression of these tumour suppressor genes, and therefore to identify a potential mechanism of CBX2 regulation of E2F and mTORC1 signalling, ChIP and CUTandRUN techniques were used to assess if CBX2 binds to the promoter regions of these genes. To optimise and validate these techniques experiments were undertaken using cells which expressed CBX2 compared to cells depleted of CBX2. To do this, cells were depleted of CBX2 via siRNA mediated knockdown. Cell transfections were performed using 3 independent siRNA targeting CBX2 (siCBX2#1/#3/#4) and a non-targeting siRNA (siScr). To validate that knockdown of CBX2 was achieved by the CBX2 targeting siRNA cells were transfected with 25nM of siRNA and grown for 72 hours before harvesting for protein and analysis by western blot. Knockdowns were performed in MDA-MB-231 cells. Target antibodies were able to detect the respective proteins and to allow visualisation and subsequent semi-quantitative analysis of western blot membranes.

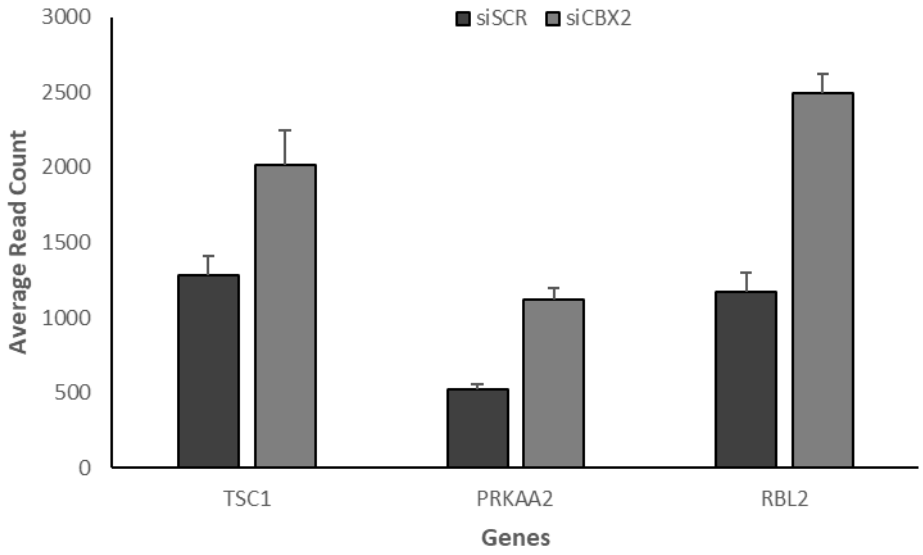

Figure 4.20. RNA-seq data.

MDA-MB-231 cells were transfected with siScr or a CBX2 targeting siRNA and incubated for 72 hours prior to RNA extraction and RNA-sequencing analysis. Average read counts for three genes, *TSC1, PRKAA2* and *RBL2* are shown. * = adjusted p-value < 0.01.

MDA-MB-231 cells transfected with siRNA targeting CBX2 shows a significant reduction in CBX2 protein expression (Figure 4.21). The molecular weight of CBX2 observed in our breast cancer cell lines is ~70 kDa. This molecular weight for CBX2 has been reported in literature (Kawaguchi *et al*., 2017), and in the Wade lab knockdown of CBX2 using multiple siRNAs, as well as overexpression of CBX2, is always observed ~70kDa (data not shown). The band observed at ~72 kDa is the siScr transfected cells. Probing with alpha-tubulin showed protein (52 kDa) was loaded into the wells, however, loading was not equal, with more protein loaded into the control sample and less protein in the CBX2 knockdowns. Despite this, it is apparent that knockdown of CBX2 was achieved by each CBX2 targeting siRNA and could be used for subsequent ChIP and CUTandRUN experiments.

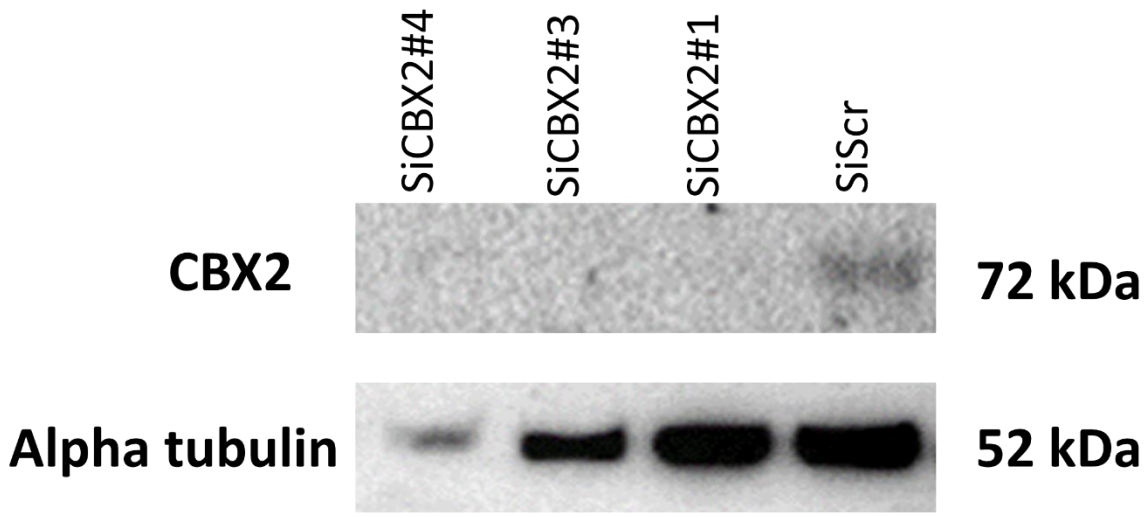

Figure 4.21. Western bot analysis of CBX2 knockdown by siRNA in MDA-MB-231 cells.
MDA-MB-231 cells were transfected with SiScr or SiCBX2 (#1,#3 and #4) for 72 hours. Lysates of cells transfected by the different siRNAs are indicated above the wells. Molecular weights of the proteins are shown in kDa to the right of the blot for reference. The bands for CBX2 and a-tubulin are indicated by labels on the left. (n=1, full blots will be in Appendix).

4.3.1. Chromatin immunoprecipitation to determine binding of CBX2 at putative tumour suppressor genes

It is hypothesised that CBX2 promotes E2F and mTORC1 oncogenic gene signatures via transcriptional repression of inhibitors of these pathways (e.g. *RBL2* for E2F signalling and *TSC1/PRKAA2* for mTORC1 signalling). If this was the case it may be expected that CBX2 would be bound to these promotor regions, thereby directing PRC1 to silence gene expression. To investigate this, preliminary analysis was conducted to see if CBX2 binds to the promoter regions of E2F and mTORC1 inhibitors.

ChIP was initially undertaken in MDA-MB-231 cells to see if *CBX2* is bound to the promoters of *RBL2, TSC1* and *PRKAA2* in TNBC. Further preliminary investigations was also done to investigate if *CBX2* influences the repressive mark deposited by PRC1 (H2AK119Ub), using an H2AK119Ub antibody. This was done to determine if regulation via PRC1 may be active at these promoter regions. The regions analysed were the promoter regions of TSC1 (66 bp upstream of the transcriptional start site), PRKAA2 (475 bp upstream of the transcriptional start site) and RBL2 (120 bp upstream of the transcriptional start site).

To try and identify if CBX2 was bound to these promotor regions and whether the presence of CBX2 affects H2AK119Ub at these sites, MDA-MB-231 cells were transfected with either siRNA targeting

CBX2 or siScr followed by ChIP analysis (figure 4.22 and 4.23). This was done to confirm that the CBX2 antibody being used was genuinely detecting CBX2 at these sites.

In figure 4.22, the promoter of *TSC1* shows a slightly higher relative amount of DNA bound for the *CBX2* knockdown than the control, which was not expected. This indicates an elevated amount of *CBX2* was bound to these sites in the si*CBX2* cells. At the promoters of *PRKAA2* and *RBL2,* there is a higher amount of DNA bound for the siScr than the *siCBX2* cells. This suggests that CBX2 was bound to these sites in the siScr MDA-MB-231 cells, whereas in the siCBX2 arm of the experiment, the amount of CBX2 protein present has been reduced. Enrichment when using the isotype control IgG antibody should be lower than the enrichment identified in the siScr arm of the experiment and potentially equivalent to the si*CBX2* arm of the experiment, which is the case for all promoter regions. This data, therefore, suggests that the CBX2 antibody is binding to CBX2 and that CBX2 is present at the promoter region of these genes. The caveat being that this is a single experiment.

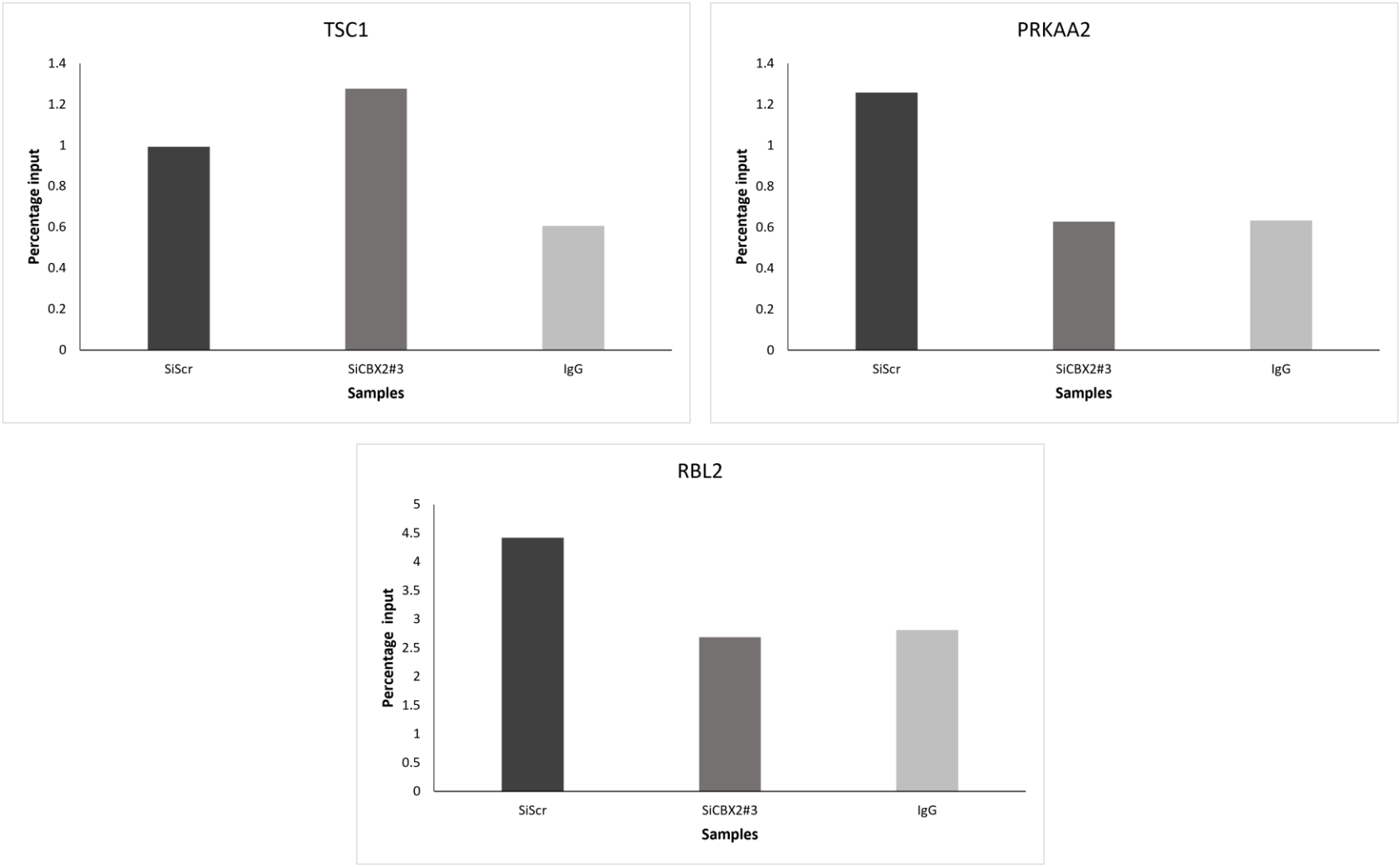

Figure 4.22. ChIP analysis of CBX2 knockdown by siRNA in MDA-MB-231 cells, indicate CBX2 may be bound at tumour suppressor regions.
MDA-MB-231 cells were transfected with siScr or SiCBX2#3 and incubated for 72 hours prior to ChIP using an antibody specific for CBX2 and in isotype control antibody (IgG) using chromatin from the siScr transfected cells. Resultant DNA was analysed by qPCR using primers specific for the promotors of TSC1, PRKAA2 and RBL2. n=1.

In a repeat experiment (Figure 4.23), a higher amount of DNA bound by CBX2 was identified in the siCBX2 cells rather than the siScr cells. This was the opposite of what was hypothesised. In addition, for both the PRKAA2 and *RBL2* promotors, more DNA was precipitated out with the IgG isotype control antibody, compared with the CBX2 antibody. The IgG isotype control is an indicator of background/non-specific immunoprecipitation of DNA. Therefore, the fact the amount of DNA immunoprecipitated when using IgG control is similar to, if not higher than, when using the CBX2 antibody, suggests that the immunoprecipitation for the CBX2 antibody may not have been successful in this experiment.

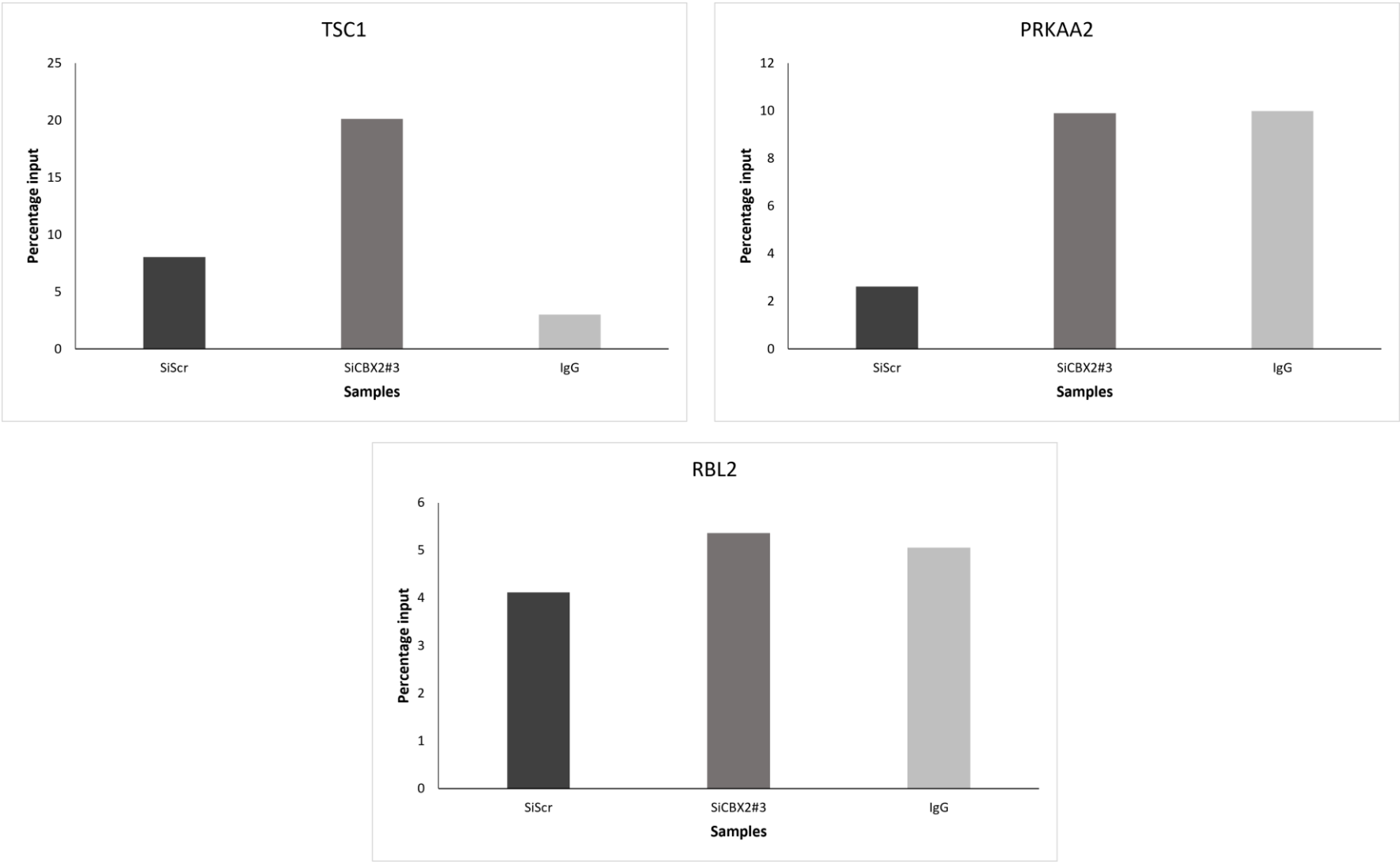

Figure 4.23. ChIP analysis of CBX2 knockdown by siRNA in MDA-MB-231 cells.
MDA-MB-231 cells were transfected with siScr or SiCBX2#3 and incubated for 72 hours prior to ChIP using an antibody specific for CBX2 and in isotype control antibody (IgG) using cells transfected with siScr. Resultant DNA was analysed by qPCR using primers specific for tumour suppressor genes. n=1.

Further analysis was done using an additional antibody H2AK119Ub, which is specific for the ubiquitination of H2A and lysine 199, to investigate whether CBX2 may be influencing the PRC1 deposited repressive mark at these promoter regions. MDA-MB-231 cells were transfected as in previous experiments - prior to ChIP, with the additional H2AK119Ub antibody (Figure 4.24). If CBX2 is directing PRC1 to these sites to ubiquitinate H2AK119, it may be expected, that when CBX2 is not present the amount of H2AK119Ub at these sites would reduce. The amount of ubiquitination in the siScr cells was higher than the amount of ubiquitination for *CBX2* knockdown cells, which suggests CBX2 may be involved in deposition of this repressive mark by PRC1, as if CBX2 was playing an active role in PRC1-mediated deposition of this repressive marks these promoter regions then loss of CBX2 would reduce H2AK119Ub at these sites. Enrichment of the IgG control antibody was low or similar to the SiCBX2 cells for all promotor regions.

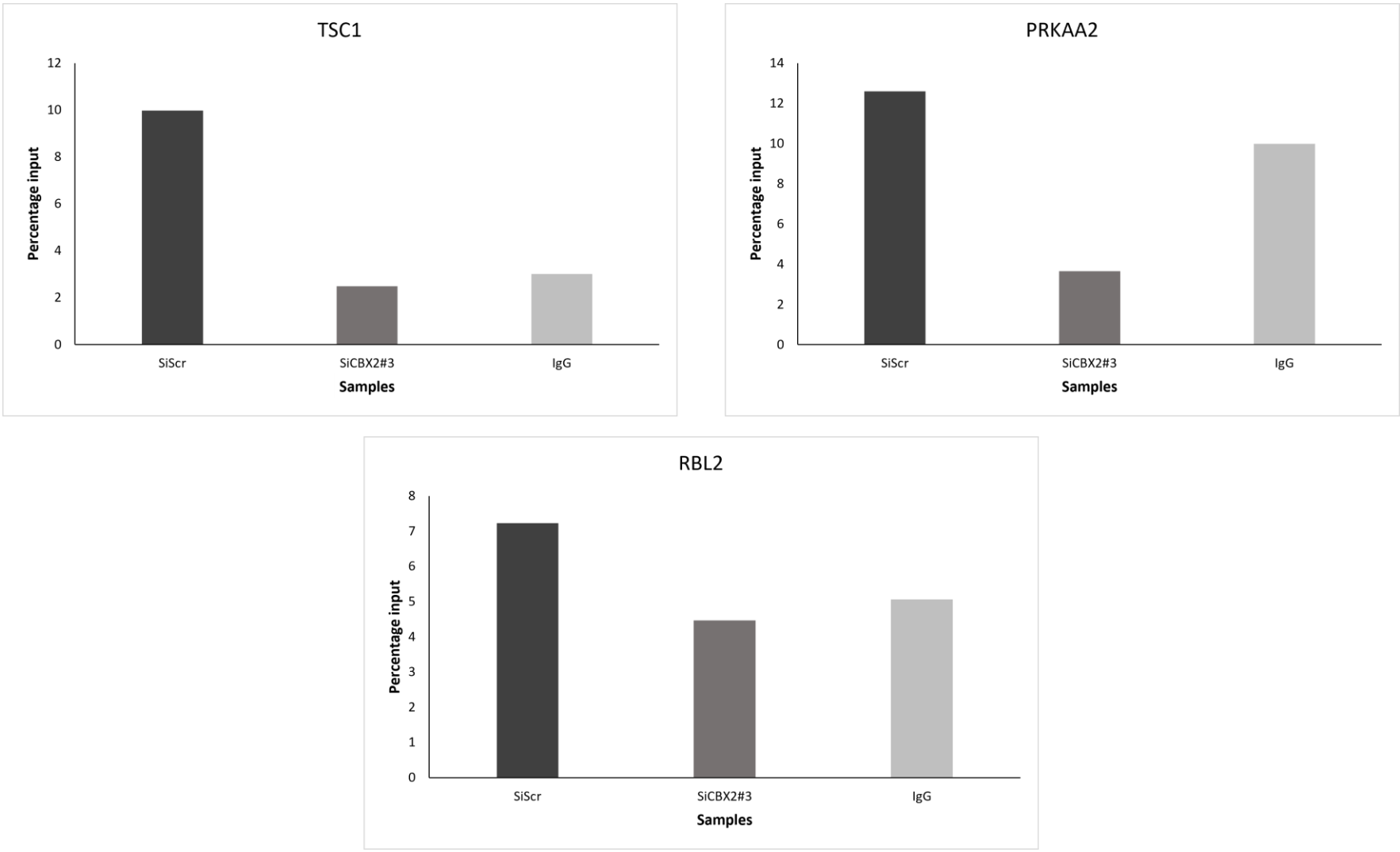

Figure 4.24. ChIP analysis indicates CBX2 may play a role in H2AK119 mono-ubiquitination, in CBX2 knockdown of MDA-MB-231 cells.
MDA-MB-231 cells were transfected with siScr or SiCBX2#3 and incubated for 72 hours prior to ChIP using an antibody specific for H2AK119 mono-ubiquitination and in isotype control antibody (IgG) using chromatin from the siScr transfected sample. Resultant DNA was analysed by qPCR using primers specific for tumour suppressor genes. n=1.

4.3.2. CUTandRUN analysis to determine binding of CBX2 at putative tumour suppressor genes

Preliminary ChIP analysis suggested that CBX2 is potentially bound to the promoter regions of *TSC1, PRKAA2* and *RBL2*. To investigate this further, analysis was conducted using the cutting edge CUTandRUN technique. CUTandRUN is a relatively new technique that has several potential advantages over traditional ChIP protocols, including a lower input number of cells required, lower background binding of antibodies and therefore reduced depth of sequencing required if samples were sequenced as there is a much lower signal to noise ratio. Less DNA needs to be sequenced in order to identify genuine binding sites above the background noise. It was hoped CUTandRUN-sequencing could be used to determine the global binding profile of CBX2 in TNBC cells, therefore, to ensure the new technique would work CUTandRUN was validated using the CBX2 antibody at the *TSC1, PRKAA2* and *RBL2* promoter regions.

The CUTandRUN technique was initially validated in MDA-MB-231 cells by using a positive control antibody (H3K4me3), which is a mark at transcriptionally active regions of the genome, and positive control primers of the promoter region of the constitutively expressed *RPLPO* gene which is supplied within the CUTandRUN kit (Figure 4.25). An IgG antibody was also used as a negative control. The enrichment for the H3K4me3 antibody was high compared to the IgG, indicating a positive enrichment, and that the CUTandRUN had worked.

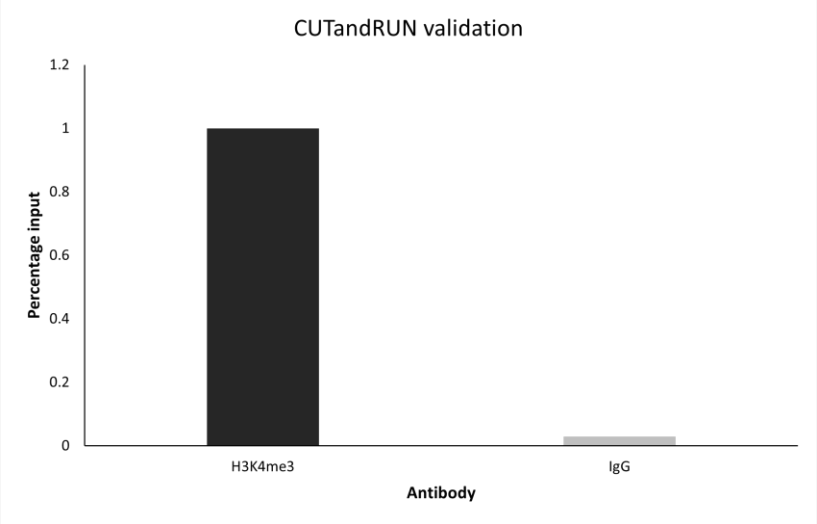

Figure 4.25. Validating antibody for the CUTandRUN in MDA-MB-231 cells.
MDA-MB-231 cells were transfected and incubated for 72 hours prior to CUTandRUN using a positive control antibody (H3K4me3) and a control antibody (IgG) to validate if the CUTandRUN worked in the cells. Resultant DNA was analysed by qPCR using primers for the promoter of the RPLPO gene. n=1.

Next, MDA-MB-231 cells were transfected with siRNA targeting *CBX2* and analysed by CUTandRUN using an antibody specific for CBX2 and primers for the promoter regions of *TSC1, PRKAA2* and *RBL2* (Figure 4.26). *TSC1* and *PRKAA2* promoter regions show an elevated relative amount of DNA bound for the siScr arm of the experiment than the *CBX2* knockdown arm of the experiment. This indicates an elevated amount of *CBX2* was bound to these sites in the siScr cells, which is what was expected as CBX2 should be depleted in siCBX2 transfected cells. This was also done for RBL2, but the qPCR was unsuccessful. The qPCR for the IgG control arm of the experiment was also unsuccessful, with no amplification curves produced.

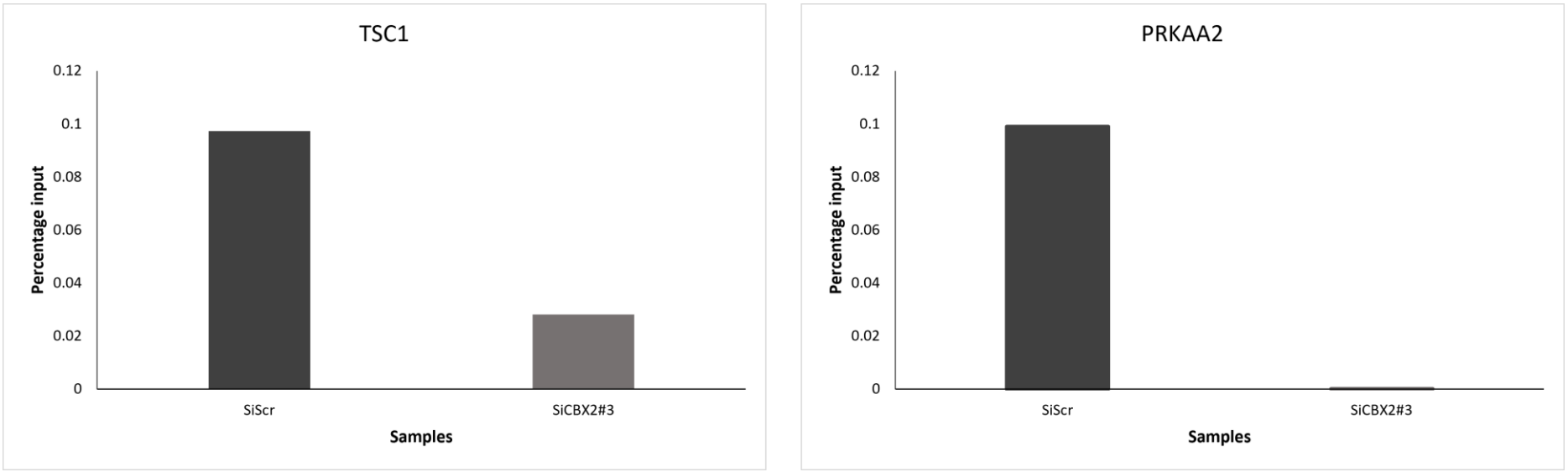

Figure 4.26. CUTandRUN analysis indicates CBX2 may be present at tumour suppressor regions.
MDA-MB-231 cells were transfected with siScr or SiCBX2#4 and incubated for 72 hours prior to CUTandRUN using an antibody specific for CBX2 and in isotype control antibody (IgG) using the siScr sample. Resultant DNA was analysed by qPCR using primers specific for tumour suppressor genes. n=1.

A repeat of this experiment was conducted (Figure 4.27). For each promoter region, a high relative amount of DNA bound was bound by CBX2 for the siScr cells compared to the Si*CBX2* cells, which was expected. IgG enrichment was extremely low in all samples. This analysis, coupled with the preliminary data from the ChIP analysis (figure 4.22) indicates CBX2 could be bound to the promotor regions of these putative tumour suppressor genes.

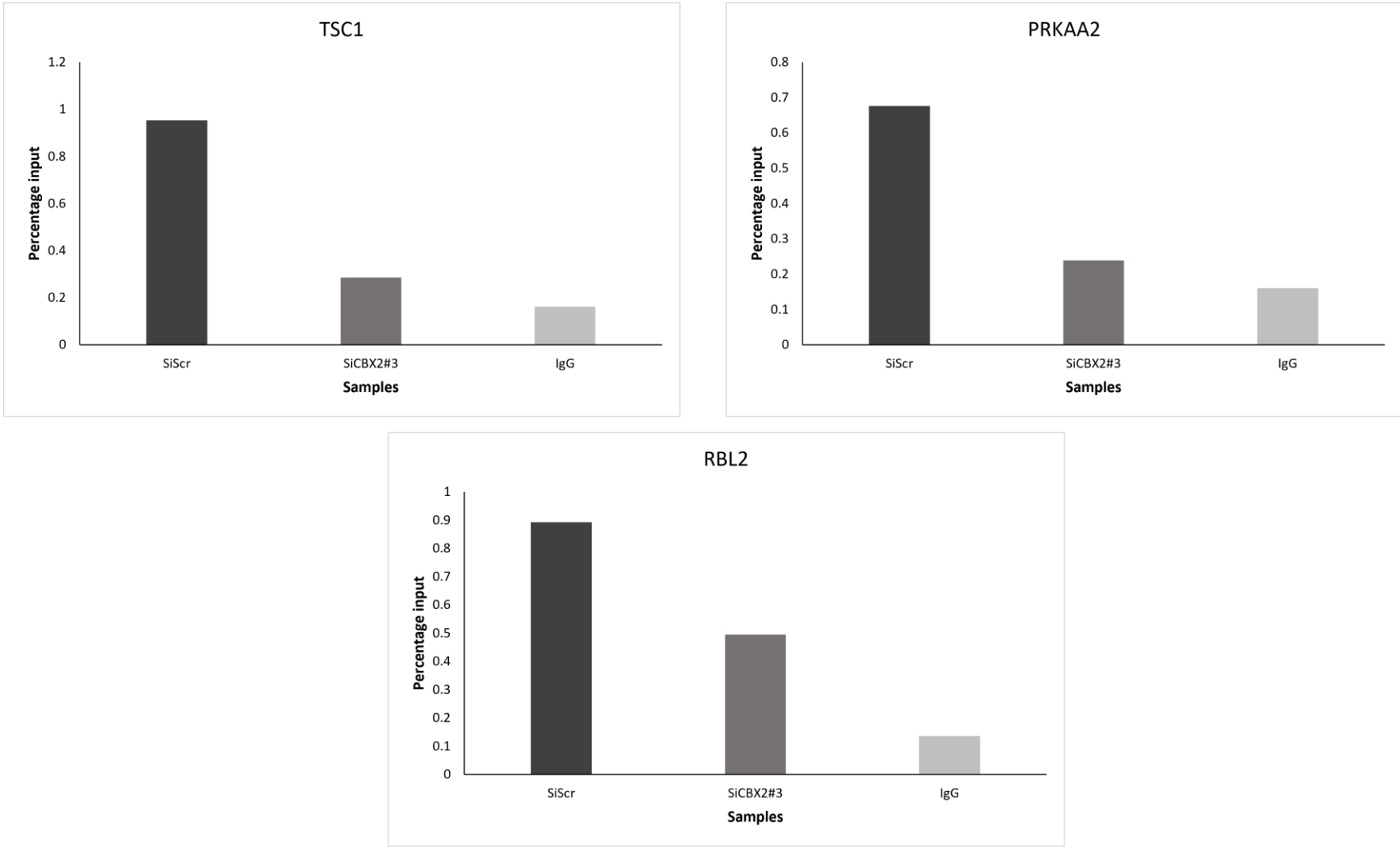

Figure 4.27. CUTandRUN analysis of CBX2 knockdown by siRNA in MDA-MB-231 cells indicate CBX2 is bound to the tumour suppressor regions.
MDA-MB-231 cells were with siScr or SiCBX2#4 and incubated for 72 hours prior to CUTandRUN using an antibody specific for CBX2 and in isotype control antibody (IgG) using the siScr sample. Resultant DNA was analysed by qPCR using primers specific for tumour suppressor genes. n=1.

Further analysis was also done using an additional antibody H2AK119Ub, which is specific for the ubiquitination of H2A, to confirm if CBX2 influences the repressive mark at these genomic regions. MDA-MB-231 cells were transfected before CUTandRUN with the H2AK119Ub antibody (Figure 4.28). At the *TSC1* there was a slightly elevated level of ubiquitination in the siScr cells, compared to the siCBX2 cells, but this was not significant. Both the *RBL2* promotor and *PRKAA2* did not follow this trend and contradicted the ChIP result (figure 4.24). IgG level was low in all samples, however not as low, relatively, as the CUTandRUN experiments using CBX2 antibody (indicated by a higher % input value). Unfortunately, at this stage of the work, the COVID-19 pandemic prevented any further repeats, optimisation, or validation experiments to be completed.

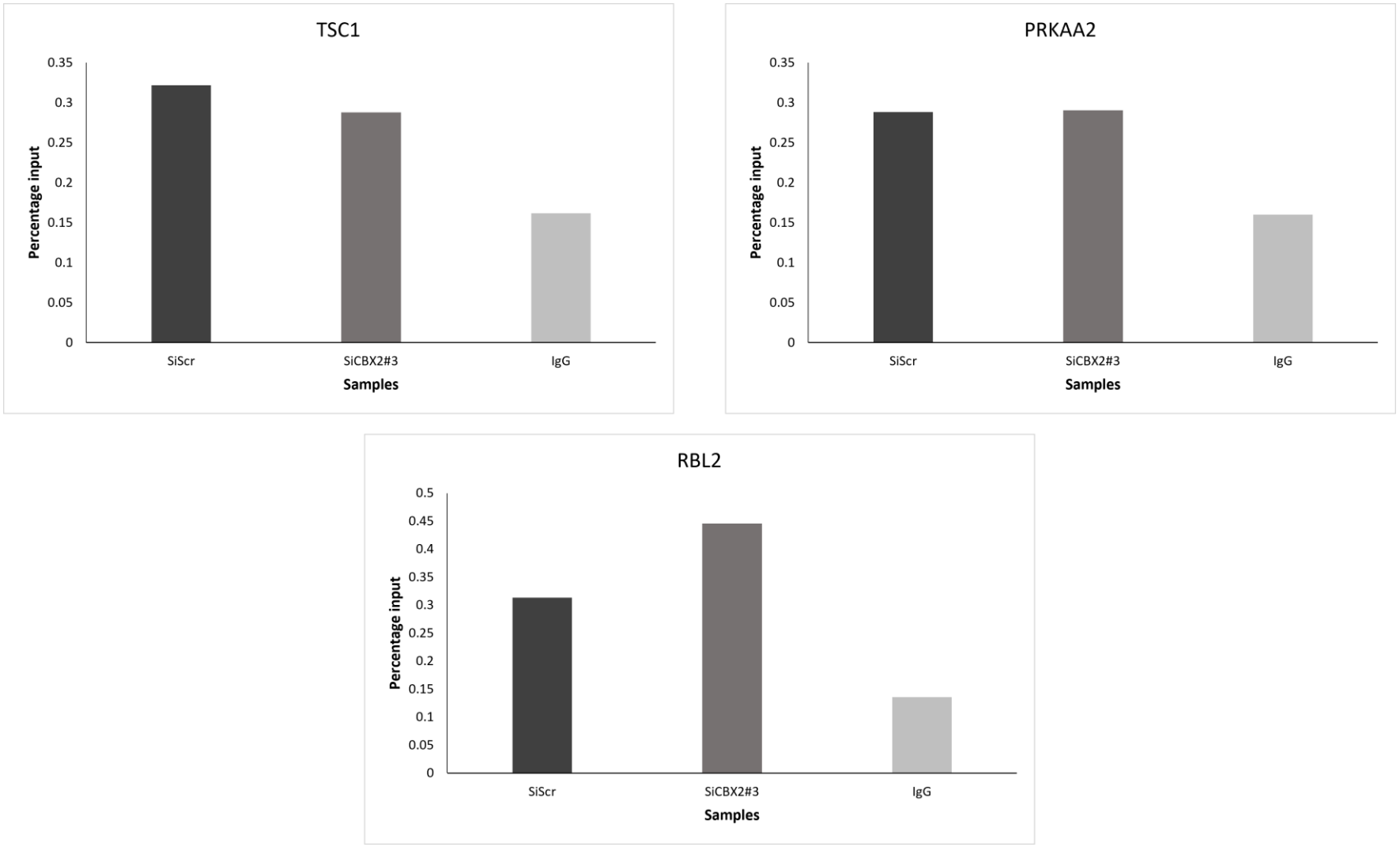

Figure 4.28. CUTandRUN analysis of H2AK119 mono-ubiquitination, in CBX2 knockdown MDA-MB-231 cells.

MDA-MB-231 cells were transfected with siScr or SiCBX2#3 and incubated for 72 hours prior to CUTandRUN using an antibody specific for H2AK119 mono-ubiquitination and an isotype control antibody (IgG) using chromatin from the siScr transfected sample. Resultant DNA was analysed by qPCR using primers specific for tumour suppressor genes. n=1.

4.4. Effect of CBX2 knockdown on MDA-MB-468 cells.

Knockdown of *CBX2* has been seen by the Wade group and others (Clermont *et al*., 2016; Zheng *et al*., 2019) in MDA-MB-231 cells and has shown a reduction in cell growth. However, this has not been seen in another TNBC cell line. These experiments were to see if effects could be replicated in a different TNBC cell line, MDA-MB-468. Initially, MDA-MB-468 were transfected with three individual siRNA sequences, targeting CBX2 (Figure 4.29) and the siScr control to see if CBX2 knockdown affected MDA-MB-468 cell growth as previously shown for MDA-MB-231 cells. Following transfection cells were grown for 96 hours and then cell counts were performed. CBX2 knockdown caused a significant reduction in the number of MDA-MB-468 cells after 96 hours (rather than 72 hours due to the slower growth in this cell line) indicating that CBX2 is required for cell growth in these cells. (Student T-test p-values in comparison with siScr control from 2 data points: SiCBX2#1 p= 0.0001, SiCBX2#3 = 0.08916 and SiCBX2#4 = 0.00941). Due to COVID-19, further investigations into *CBX2* could not be conducted.

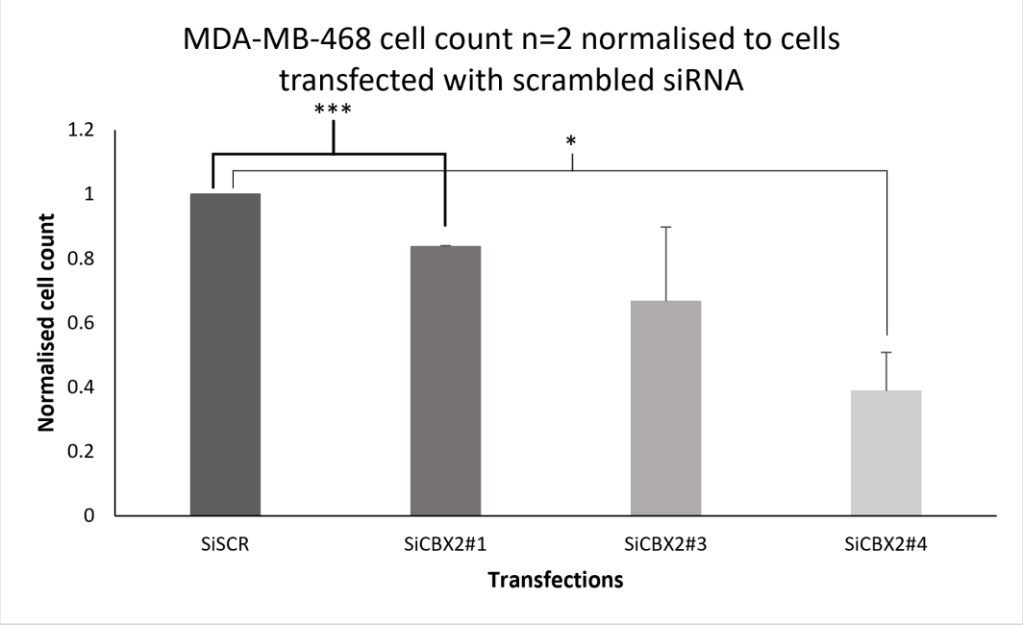

Figure 4.29. Cell count of MDA-MB-468 after CBX2 knockdown.
MDA-MB-468 cell were transfected for 96 hours, with siScr, SiCBX2#1, SiCBX2#3 and SiCBX2#4 and then cell counts were performed. Data is the average of 2 repeats. Cell counts are relative to siScr transfected cells. P-values determined by student T-test (P-values; * = <0.05, ** = <0.005 and *** = <0.0005).

5. Discussion

TNBC is an aggressive form of breast cancer that has a lack of therapeutic target treatments, meaning identifying novel targets is a necessity. CBX2 has been shown to be upregulated in a variety of cancers, such as prostate cancer, and therefore has the potential to be a possible therapeutic target (Clermont *et al*., 2014; Clermont *et al*., 2016) . CBX2 has also been shown to be upregulated in basal and HER2 positive breast cancers, indicating a potential regulatory role (Chan *et al*., 2018). This study aimed to determine possible pathways and processes regulated by CBX2 and the specific role CBX2 may play in regulating these pathways. Ultimately, this information will add further avenues for investigation into CBX2 as a genuine therapeutic target and add to the knowledge about the role of CBX2 in cancer progression and development.

5.1. CBX2 is upregulated in cancer tissue compared to normal

In-silico analysis identified that *CBX2* has an elevated mRNA expression in tumour tissue compared to normal tissue, in a variety of cancers, with a significant difference observed in glioblastoma and breast cancer, which supports previous findings from different studies (Chen *et al*., 2017; Zheng *et al*., 2019). When we further investigated breast cancer, we saw there was a significant increase in expression in all molecular subtypes apart from the luminal A subtype of breast cancer. When looking at OS, patients with elevated CBX2 expression had a poorer prognosis in luminal A and B breast cancer. However, despite the p-values not being statistically significant, in DFS there was an indication in basal-like breast cancer that those in the "high" *CBX2* group had a higher relapse rate. This is supported by Zheng *et al*., (2019) who showed that OS and PFS for breast cancer patients with high CBX2 expression was noticeably worse than those of the low CBX2 expression.

Overall, survival analysis showed that there was a significant difference in "high" *CBX2* expression groups for glioblastoma and lung adenocarcinoma, indicating those with a high *CBX2* expression had a poorer prognosis. DFS, showed a significant difference in prostate adenocarcinoma, indicating those in the "high" *CBX2* expression group had an increased chance of relapse. These results support earlier findings by Mao *et al.*, (2019) who used Kaplan-Meier plots to determine CBX2 effect in the OS of

patients with hepatocellular carcinoma. Patients with elevated expression of CBX2 had a worse prognosis and lower survival compared to the patients with a lower CBX2 expression (Mao *et al*., 2019). CBX2 was also seen as highly expressed in gastric cancer cell lines compared to normal gastric cells, via RT-PCR, with it being particularly high in MFC cells (Zeng *et al*., 2021). Together, these data indicate that CBX2 may perform a role in the growth and advancement of different cancers and may also be a prognostic indicator in various cancer types.

5.2. CBX2 has a role in E2F and mTORC1 signalling

Our GSEA analysis of patient datasets indicated that *CBX2* may influence a variety of different oncogenic cellular processes, in glioblastoma, lung adenocarcinoma and breast cancer, such as progression through the cell cycle. This has been supported by murine studies as disruption of S phase progression occurred due to the loss of M33 (mouse CBX2 orthologue), causing cells to enter senescence (Core *et al*., 2004). Through our analysis, several hallmark gene sets were repeated throughout the datasets, indicating *CBX2* may have a particular role or effect with regulating genes within those gene sets. For example, genes that positively correlated with CBX2 expression were enriched in the hallmark gene sets for E2F and mTORC1 signalling, which indicates *CBX2* may have a role in promoting these oncogenic pathways. The identification of these pathways from patient tumour datasets also agreed with analysis from our RNA-seq data, as the *CBX2* knockdown in MDA-MB-231 cells caused genes in both the mTORC1 and E2F hallmark gene set to be downregulated, supporting a direct role for CBX2 in promoting these pathways. Our analysis also revealed oncogenic gene sets which were noticeable throughout the different datasets, such as RB_P107_DN.V1_UP(genes upregulated in keratinocytes from RB1 and RBL1 knockout mice), E2F1_UP.V1_UP (genes upregulated in fibroblasts over-expressing E2F1) and MTOR_UP.N4.V1_DN (genes downregulated in CEM-C1 cells treated with rapamycin, mTOR pathway inhibitor), indicating *CBX2* association with these gene signatures. Due to this finding, we investigated the potential role of CBX2 in regulating both E2F and mTORC1 signalling.

5.2.1. mTORC1 signalling

mTORC1 is a kinase complex that is bound to Rag isoforms which are required for its action. mTORC1 is also activated by the GTP binding protein, Rheb (figure 5.1). Activated mTORC1 phosphorylates P70S6K which binds to CAD allowing pyrimidine synthesis. P70S6K phosphorylation also activates eIF4B, allowing it to directly phosphorylate the S6 subunit of the ribosome (which P70S6K can also do directly). Additionally, P70S6K also inhibits eEF2K (via phosphorylation) allowing eEF2 to activate and bind to the S6 subunit. Other proteins inhibited by mTORC1 mediated phosphorylation are 4EBP1, ULK1 and TFEB. When 4EBP1 is inhibited, it allows the eIF4E-eIF4G complex to activate the S6 subunit. ULK1 also triggers autophagy, and TFEB causes lysosome biogenesis, which in turn contributes to autophagy. This activity of the mTORC1 pathway ultimately increases protein synthesis but decreases autophagy and protein degradation which in turn can lead to cell growth. mTORC1 signalling has been associated with tumour growth and proliferation in several cancers, such as breast, colorectal, lung, pancreatic cancer and perivascular epithelioid cell tumours (Zoncu *et al*., 2011; Kwiatkowski & Wagle, 2015; Ullah *et al*., 2015; Knight *et al*., 2021).

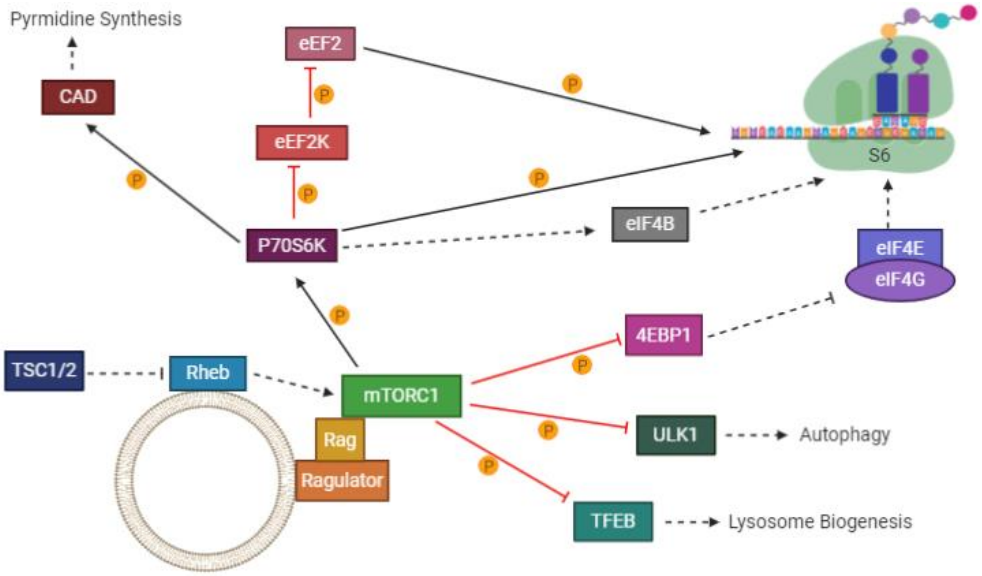

Figure 5.1. mTORC1 signalling.
Mechanism of mTORC1 signalling. Figure made in Biorender.com.

Recently, Iqbal *et al*., (2021) published evidence of CBX2 having a role within mTORC1 signalling in breast cancer. In correlation analysis, not only was *CBX2* confirmed to have a role via PDS score (analysis done by Pathifier which calculates in every sample the extent the pathway is deregulated), but *CBX2* expression also correlated with a marker of mTOR signalling (phosphorylated ribosomal S6 protein) in patient tumours within the TCGA database. The link between CBX2 and mTORC1 signalling was further supported by a reduction of phosphorylation of ribosomal s6 protein (identified by immunoblotting) in MDA-MB-231 and MCF-7 breast cancer cell lines following silencing of *CBX2* by siRNA. To further investigate the role of CBX2 in cancer, a geno-transcriptomic analysis (using the METABRIC and TCGA database) was performed. This analysis showed *CBX2* to be mostly upregulated in breast tumours, compared to normal, in both METABRIC and TCGA datasets. This mirrored the GEPIA2 results reported in this thesis. The findings additionally showed CBX2 correlated with aggressive subtypes. As high proliferation rates are linked with aggressive tumours, CBX2 was found to correlate with the protein levels of Ki67 (tumour proliferation marker), further indicating CBX2 has a role in cancer progression. Kaplan-Meier plots were also conducted to determine CBX2 effect on disease-specific survival which showed that those in the CBX2 high group has a poorer prognosis in breast cancer, even when basal-like and HER2 subtype samples were removed. This indicates that CBX2 influences survival. This also authenticates our results, as we found those in the CBX2 high group tend to have a relapse

Despite the identification in our study and the study by Iqbal *et al.,* (2021) that CBX2 plays a role in promoting mTORC1 signalling it has not been determined how CBX2 does this. To try and elucidate this we looked at the expression of mTORC1 inhibitors in our RNA-seq dataset. Classically CBX2 is an epigenetic repressor, we, therefore, looked to see if the expression of mTORC1 inhibitors increased following CBX2 knockdown. We identified upregulation of *TSC1*, *TSC2* and *PRKAA2* in our ChIP RNA-seq data following *CBX2* knockdown.

Tuberous sclerosis complex 1 and 2 (TSC1 and TSC2) is a heterodimeric complex that are also tumour suppressors (Huang & Manning, 2008; Dibble *et al*., 2012; Saxton & Sabatini, 2017). The TSC1-TSC2 heterodimer dephosphorylates GTP-bound Rheb (which is active), to form GDP-bound Rheb (which is inactive). The active form of Rheb in turn activates mTORC1 signalling (Figure 5.1). PRKAA2 is the catalytic subunit of AMP-activated protein kinase (AMPK). AMPK can phosphorylate TSC2 to activate TSC1-TSC2 hetrerodimerisation which in turn would inhibit mTORC1 (via dephosphorylation of Rheb). Upregulation of *PRKAA2* may therefore also enhance inhibition of mTORC1. The hypothesis is therefore that high expression of CBX2 represses the expression of *TSC1/TSC2* and *PRKAA2* in order to enhance pro-oncogenic mTORC1 signalling.

To determine what effect CBX2 was having on the expression of these genes we performed ChIP and CUTandRUN experiments using an anti-CBX2 antibody and analysed the enrichment of CBX2, at the promotor regions of *TSC1* and *PRKAA2*. The analysis indicated that CBX2 was bound to promotors of these genes and that knockdown of CBX2 reduced the amount of repressive H2AK119Ub marks at the promotors. This would suggest that CBX2 directly represses the expression of these genes via PRC1 mediated H2AK119Ub, therefore promoting mTORC1 signalling. Unfortunately, due to the COVID-19 pandemic subsequent repeats of these experiments could not be performed therefore definitive conclusions cannot be drawn. This is however the first evidence of a direct regulatory role of CBX2 on mTORC1 signalling.

5.2.2. E2F signalling

E2Fs are transcription factors that coordinate transition through different stages of the cell cycle to S phase (Nevins, 2001). E2Fs are separated into three groups, activators which peak in the G1-S phase (E2F1, E2F2 & E2F3A), atypical repressor occurring in the late S phase (E2F7 & E2F8) and canonical repressor which are present throughout the cell cycle (E2F3B, E2F4, E2F5 & E2F6) (Chen *et al*., 2009; Kent *et al*., 2017; Liban *et al*., 2016). E2Fs are bound to retinoblastoma tumour suppressor protein (Rb) which prevents transcriptional activity. CDK and cyclin complexes phosphorylate Rb to interrupt

the Rb/E2F interaction, thereby activating E2F signalling and movement through the cell cycle (Henley & Dick, 2012; Bertoli *et al*., 2013; Dyson, 2016; Liban *et al*., 2016).

E2Fs have a defined role within the normal cellular process. However, dysregulation can contribute to cancer progression. Kent *et al*., 2016 explored E2F7 and E2F8 genetic and epigenetic alterations within three different databases (COSMIC, TCGA and Gene expression Omnibus). They concluded while genetic alterations were infrequent, mRNA expression was elevated within various stages of hepatocellular carcinoma (HCC). Another analysis, using ChIP, showed a correlation between E2F7 and E2F8 mRNA expression and MKI67 (encodes proliferation marker Ki67) mRNA expression, with elevated expressions of EFs associating with proliferation in HCC (Kent *et al*., 2016). Further evidence, using TCGA data sets, showed E2F7 and E2F8 elevated in a variety of cancers, with high E2F8 expression found in human liver samples, alongside MK167 expression. Further studies showed that an extra copy of *E2F1* or *E2F3B* increases those proteins levels and induces HCC (Kent *et al*., 2017). Overall, we can determine from these studies that E2F has a role in cancer progression.

Despite the discovery in our study that CBX2 plays a part in promoting E2F signalling, it has not been determined how CBX2 does this. To attempt and explain this, we analysed the expression of E2F inhibitors in our RNA-seq dataset, to determine if the inhibitors increased after CBX2 knockdown. The gene RBL2 was identified to be upregulated following CBX2 knockdown. RBL2/p130 is a tumour suppressor which has been linked to a variety of cancers, including breast and prostate cancer (Sadasivam & DeCaprio, 2013; Farman *et al*., 2018; Zhu *et al*., 2018). Phosphorylation of RBL2 prevents binding to E2F transcriptional factors, facilitating the G1-S phase of the cell cycle, with dysregulated phosphorylation leading to hyperproliferation and tumour development (Farman *et al*., 2018). RBL2/p130 has been shown to bind to E2F4 or E2F5 which forms part of a complex, known as the DREAM complex (Dimerisation partner (DP), RB-like, E2F and multi-vulval class B (MuVB)); the DREAM complex also includes the MuVB complex, with BMYB and FOXM1 (Forkhead box M1) (Sadasivam & DeCaprio, 2013; Guiley *et al*., 2015; Fischer *et al*., 2016; Engeland, 2018). The RBL2 and E2F4/E2F5

component has been associated with transcriptional repression. p21/CDKN1A (and other inhibitors such as p27/Kip1 and p57/Kip2) inhibits cyclin-dependent kinases (CDKs), allowing hyperphosphorylation of p130 (and p107), triggering the DREAM complex to form. The RBL2-E2F4/E2F5 component can bind to E2F targets genes preventing transcription and therefore halting cell cycle progression and cell growth (Engeland, 2018; Iness & Litovchick, 2018; Pentimalli *et al*., 2018; Zhu *et al*., 2018b). The DREAM complex is apparent in quiescent cells but disassembles during G1/S transition, when CDK4 and CDK2 phosphorylate RBL2, triggering MuVB and E2F4 dissociation from RBL2, thereby progressing the cell cycle (Iness & Litovchick, 2018; Schade *et al*., 2019). This can also be disrupted by the human papilloma virus (HPV), which uses E6 and E7 to target RBL2 and dysregulate the DREAM complex (Fischer *et al*., 2014)**.** In two studies, qRT-PCR determined RBL2 expression was quantified in breast tumours and adjacent normal tissue (control), and RBL2 was found to be significantly reduced in tumour tissues, then the control, confirming RBL2 tumour suppressor role (Ullah *et al*., 2015; Farman *et al*., 2018). The hypothesis is therefore that high expression of CBX2 in cancer represses the expression of RBL2 to prevent DREAM complex formation and therefore promote transition through the cell cycle.

To establish what direct effect CBX2 had on RBL2 expression we performed ChIP and CUTandRUN experiments using an anti-CBX2 antibody and an anti-H2AK119Ub antibody and analysed the promotor region of *RBL2.* Our collected evidence suggests that CBX2 is bound to the promoters of *RBL2* and that knockdown of CBX2 reduces repressive H2AK119Ub marks. This data, therefore, indicates that CBX2 may regulate the expression of RBL2 and could have an inhibitory effect, which would promote E2F signalling and progression through the cell cycle. Unfortunately, due to the COVID-19 pandemic subsequent repeats of these experiments could not be performed therefore definitive conclusions cannot be drawn.

5.3. Conclusion

The TNBC subtype is a severe problem due to the lack of therapeutic targets and resistance to endocrine therapy. In total, our research has identified through analysis of patient datasets and gene

sets that CBX2 contributes to cancer progression, and possibly regulates pro-oncogenic E2F and mTORC1 signalling. Our RNA-seq data showed that genes that contribute to E2F and mTORC1 signalling are downregulated when CBX2 is knockdown, which indicates that CBX2 has a direct regulatory role. We also revealed four tumour suppressor genes involved with these pathways (TSC1, TSC2, PRKAA2 and RBL2) which are upregulated when CBX2 is knocked down. This supports that CBX2 has a prospective role in regulating not only these genes but the pathways they affect. Our ChIP and CUTandRUN data indicated that CBX2 is bound to the promoter regions of the four tumour suppressor genes, suggesting that CBX2 may repress expression of these genes and therefore promote E2F and mTORC1 pathways in cancer. Inhibition of CBX2 in TNBC may therefore be a potential therapeutic option and CBX2 merits further investigation as a potential therapeutic target.

5.4. Further directions

In general, our research has shown that CBX2 has a potential role in cancer progression and may also be a therapeutic target to investigate. CBX2 has been shown to have an increased expression in a variety of cancers, particularly breast cancer. From our research, we can see that CBX2 may have a possible role or effect within numerous signalling pathways, associated with breast cancer progression and we have identified potential ways that CBX2 could directly regulate these pathways via repression of tumour suppressor genes. Further research to determine CBX2 role within breast cancer is needed, as well as further investigations on CBX2 being used as a prognostic target. Additional repeats from our experiments are needed, to determine the validity of our results and investigation in additional cell lines as well as more translationally relevant models such as *ex vivo* patient cultures of patient derived xenograft models. Identification that CBX2 binds to, and has an effect on H2AK119Ub, at the promotors of proposed targets genes in additional models would help validate the role of CBX2 in regulating these pathways. In addition, the global DNA binding profile of CBX2 in TNBC should be investigated via ChIP-sequencing or CUTandRUN-sequencing to determine the full CBX2 regulatory signature. In this study conditions for CBX2 ChIP and CUTandRUN have begun to be optimised. ChIP/CUTandRUN-sequencing of TNBC models would identify where on the chromatin CBX2 is bound.

Through further research of CBX2, we can investigate its potential as a therapeutic target for TNBC, which is an aggressive form and lacks treatments. This could potentially lead to the development of treatments for a wider range of cancers, as this preliminary data has shown CBX2 is upregulated in cancer tissues compared to normal tissues.

6. References

Abd El-Rehim, D. M., Ball, G., Pinder, S. E., Rakha, E., Paish, C., Robertson, J. F. R., Macmillan, D., Blamey, R. W. & Ellis, I. O. (2005) High-throughput protein expression analysis using tissue microarray technology of a large well-characterised series identifies biologically distinct classes of breast cancer confirming recent cDNA expression analyses. *International Journal of Cancer,* 116 (3), 340-350.

Abe, O., Abe, R., Enomoto, K., Kikuchi, K., Koyama, H., Masuda, H., Nomura, Y., Sakai, K., Sugimachi, K., Tominaga, T., Uchino, J., Yoshida, M., Haybittle, J. L., Davies, C., Harvey, V. J., Holdaway, T. M., Kay, R. G., Mason, B. H., Forbes, J. F., Wilcken, N., Gnant, M., Jakesz, R., Ploner, M., Yosef, H., Focan, C., Lobelle, J. P., Peek, U., Oates, G. D., Powell, J., Durand, M., Mauriac, L., Di Leo, A., Dolci, S., Piccart, M. J., Masood, M. B., Parker, D., Price, J. J., Hupperets, P., Jackson, S., Ragaz, J., Berry, D., Dubois, J. B., Gray, R., de Vries, E., Klijn, J., Kaufmann, M., Zhao, D. B., Nilsson, J. & Buchanan, R. B. (2005) Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: An overview of the randomised trials. *The Lancet (British Edition),* 365 (9472), 1687-1717.

Akram, M., Iqbal, M., Daniyal, M. & Khan, A. U. (2017) Awareness and current knowledge of breast cancer. *Biological Research,* 50 (1), 33.

Alaskhar Alhamwe, B., Khalaila, R., Wolf, J., von Bülow, V., Harb, H., Alhamdan, F., Hii, C. S., Prescott, S. L., Ferrante, A., Renz, H., Garn, H. & Potaczek, D. P. (2018) Histone modifications and their role in epigenetics of atopy and allergic diseases. *Allergy, Asthma, and Clinical Immunology,* 14 (1), 39.

Anderson, W., Chatterjee, N., Ershler, W. & Brawley, O. (2002) Estrogen receptor breast cancer phenotypes in the surveillance, epidemiology, and end results database. *Breast Cancer Research and Treatment,* 76 (1), 27-36.

Arteaga, C. L., Sliwkowski, M. X., Osborne, C. K., Perez, E. A., Puglisi, F. & Gianni, L. (2012) Treatment of HER2-positive breast cancer: Current status and future perspectives. *Nature Reviews Clinical Oncology,* 9 (1), 16-32.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., Sherlock, G. & The Gene Ontology Consortium (2000) Gene ontology: Tool for the unification of biology. *Nature Genetics,* 25 (1), 25-29.

Audia, J. E. & Campbell, R. M. (2016) Histone modifications and cancer. *Cold Spring Harbor Perspectives in Biology,* 8 (4), a019521.

Badeaux, A. I. & Shi, Y. (2013) Emerging roles for chromatin as a signal integration and storage platform. *Nature Reviews. Molecular Cell Biology,* 14 (4), 211-224.

Badowska-Kozakiewicz, A. M. & Budzik, M. P. (2016) Immunohistochemical characteristics of basal-like breast cancer. *Contemporary Oncology (Poznan, Poland),* 20 (6), 436-443.

Badve, S., Dabbs, D. J., Schnitt, S. J., Baehner, F. L., Decker, T., Eusebi, V., Fox, S. B., Ichihara, S., Jacquemier, J., Lakhani, S. R., Palacios, J., Rakha, E. A., Richardson, A. L., Schmitt, F. C., Tan, P., Tse, G. M., Weigelt, B., Ellis, I. O. & Reis-Filho, J. S. (2011) Basal-like and triple-negative breast cancers: A

critical review with an emphasis on the implications for pathologists and oncologists. *Modern Pathology,* 24 (2), 157-167.

Bannister, A. J. & Kouzarides, T. (2011) Regulation of chromatin by histone modifications. *Cell Res,* 21 (3), 381-395.

Barbour, H., Daou, S., Hendzel, M. & Affar, E. B. (2020) Polycomb group-mediated histone H2A monoubiquitination in epigenome regulation and nuclear processes. *Nature Communications,* 11 (1), 5947.

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C. L., Serova, N., Davis, S. & Soboleva, A. (2012) NCBI GEO: Archive for functional genomics data sets—update. *Nucleic Acids Research,* 41 (D1), D991-D995.

Bertoli, C., Skotheim, J. M. & de Bruin, Robertus A M (2013) Control of cell cycle transcription during G1 and S phases. *Nature Reviews. Molecular Cell Biology,* 14 (8), 518-528.

Bianchini, G., Balko, J. M., Mayer, I. A., Sanders, M. E. & Gianni, L. (2016) Triple-negative breast cancer: Challenges and opportunities of a heterogeneous disease. *Nature Reviews. Clinical Oncology,* 13 (11), 674-690.

Blackledge, N. P., Fursova, N. A., Kelley, J. R., Huseyin, M. K., Feldmann, A. & Klose, R. J. (2019) PRC1 catalytic activity is central to polycomb system function. Cold Spring Harbor Laboratory.

Burstein, H. J. (2005) The distinctive nature of HER2-positive breast cancers. *The New England Journal of Medicine,* 353 (16), 1652-1654.

Cancer Research, U. K. (2017) *Cancer statistics.* Available online: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer/incidence-invasive#heading-Zero [Accessed 15/04/ 2020].

Carey, L., Winer, E., Viale, G., Cameron, D. & Gianni, L. (2010) Triple-negative breast cancer: Disease entity or title of convenience? *Nature Reviews. Clinical Oncology,* 7 (12), 683-692.

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., Jacobsen, A., Byrne, C. J., Heuer, M. L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A. P., Sander, C. & Schultz, N. (2012) The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discovery,* 2 (5), 401.

Chan, H. L., Beckedorff, F., Zhang, Y., Garcia-Huidobro, J., Jiang, H., Colaprico, A., Bilbao, D., Figueroa, M. E., LaCava, J., Shiekhattar, R. & Morey, L. (2018) Polycomb complexes associate with enhancers and promote oncogenic transcriptional programs in cancer through multiple mechanisms. *Nature Communications,* 9 (1), 3377-16.

Chatterjee, N. & Walker, G. C. (2017) Mechanisms of DNA damage, repair, and mutagenesis. *Environmental and Molecular Mutagenesis,* 58 (5), 235-263.

Cheang, M. C. U., Voduc, D., Bajdik, C., Leung, S., McKinney, S., Chia, S. K., Perou, C. M. & Nielsen, T. O. (2008) Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype. *Clinical Cancer Research,* 14 (5), 1368-1376.

Cheang, M. C. U., Chia, S. K., Voduc, D., Gao, D., Leung, S., Snider, J., Watson, M., Davies, S., Bernard, P. S., Parker, J. S., Perou, C. M., Ellis, M. J. & Nielsen, T. O. (2009) Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. *Journal of the National Cancer Institute,* 101 (10), 736-750.

Chen, H., Tsai, S. & Leone, G. (2009) Emerging roles of E2Fs in cancer: An exit from cell cycle control. *Nature Reviews. Cancer,* 9 (11), 785-797.

Chen, W. Y., Zhang, X. Y., Liu, T., Liu, Y., Zhao, Y. S. & Pang, D. (2017) Chromobox homolog 2 protein: A novel biomarker for predicting prognosis and taxol sensitivity in patients with breast cancer. *Oncology Letters,* 13 (3), 1149-1156.

Clermont, P., Crea, F., Chiang, Y. T., Lin, D., Zhang, A., Wang, J. Z. L., Parolia, A., Wu, R., Xue, H., Wang, Y., Ding, J., Thu, K. L., Lam, W. L., Shah, S. P., Collins, C. C., Wang, Y. & Helgason, C. D. (2016) Identification of the epigenetic reader CBX2 as a potential drug target in advanced prostate cancer. BioMed Central.

Clermont, P., Sun, L., Crea, F., Thu, K. L., Zhang, A., Parolia, A., Lam, W. L. & Helgason, C. D. (2014) Genotranscriptomic meta-analysis of the polycomb gene CBX2 in human cancers: Initial evidence of an oncogenic role. *British Journal of Cancer,* 111 (8), 1663-1672.

Core, N., Joly, F., Boned, A. & Djabali, M. (2004) Disruption of E2F signaling suppresses the INK4a-induced proliferative defect in M33-deficient mice. *Oncogene,* 23 (46), 7660-7668.

Dai, X., Li, T., Bai, Z., Yang, Y., Liu, X., Zhan, J. & Shi, B. (2015) Breast cancer intrinsic subtype classification, clinical use and future trends. *American Journal of Cancer Research,* 5 (10), 2929-2943.

Dent, R., Trudeau, M., Pritchard, K. I., Hanna, W. M., Kahn, H. K., Sawka, C. A., Lickley, L. A., Rawlinson, E., Sun, P. & Narod, S. A. (2007) Triple-negative breast cancer: Clinical features and patterns of recurrence. *Clinical Cancer Research,* 13 (15), 4429-4434.

Dhar, S., Gursoy-Yuzugullu, O., Parasuram, R. & Price, B. D. (2017) The tale of a tail: Histone H4 acetylation and the repair of DNA breaks. *Philosophical Transactions. Biological Sciences,* 372 (1731), 20160284.

Di Costanzo, A., Del Gaudio, N., Conte, L., Dell'Aversana, C., Vermeulen, M., de The, H., Nebbioso, A. & Altucci, L. (2018) The HDAC inhibitor SAHA regulates CBX2 stability via a SUMO-triggered ubiquitin-mediated pathway in leukemia. *Oncogene,* 37 (19), 2559-2572.

Dibble, C., Elis, W., Menon, S., Qin, W., Klekota, J., Asara, J., Finan, P., Kwiatkowski, D., Murphy, L. & Manning, B. (2012) TBC1D7 is a third subunit of the TSC1-TSC2 complex upstream of mTORC1. *Molecular Cell,* 47 (4), 535-546.

Dyson, N. J. (2016) RB1: A prototype tumor suppressor and an enigma. *Genes & Development,* 30 (13), 1492-1502.

Edgar, R., Domrachev, M. & Lash, A. E. (2002) Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research,* 30 (1), 207-210.

Engeland, K. (2018) Cell cycle arrest through indirect transcriptional repression by p53: I have a DREAM. *Cell Death and Differentiation,* 25 (1), 114-132.

Farman, F., Iqbal, M., Azam, M. & Saeed, M. (2018) Nucleosomes positioning around transcriptional start site of tumor suppressor (Rbl2/p130) gene in breast cancer. *Molecular Biology Reports,* 45 (2), 185-194.

Feng, Y., Spezia, M., Huang, S., Yuan, C., Zeng, Z., Zhang, L., Ji, X., Liu, W., Huang, B., Luo, W., Liu, B., Lei, Y., Du, S., Vuppalapati, A., Luu, H. H., Haydon, R. C., He, T. & Ren, G. (2018) Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis. *Genes & Diseases,* 5 (2), 77-106.

Fischer, M., Quaas, M., Steiner, L. & Engeland, K. (2016) The p53-p21-DREAM-CDE/CHR pathway regulates G2/M cell cycle genes. *Nucleic Acids Research,* 44 (1), 164-174.

Fischer, M., Steiner, L. & Engeland, K. (2014) The transcription factor p53: Not a repressor, solely an activator. *Cell Cycle (Georgetown, Tex.),* 13 (19), 3037-3058.

Foulkes, W. D., Smith, I. E. & Reis-Filho, J. S. (2010) Triple-negative breast cancer. *The New England Journal of Medicine,* 363 (20), 1938-1948.

Fraga, M. F., Ballestar, E., Gopalakrishna Iyer, N., Perez-Rosado, A., Calvo, E., Lopez, J. A., Cano, A., Calasanz, M. J., Colomer, D., Piris, M. A., Ahn, N., Imhof, A., Villar-Garea, A., Caldas, C., Jenuwein, T., Esteller, M., Boix-Chornet, M., Espada, J., Schotta, G., Bonaldi, T., Haydon, C., Ropero, S. & Petrie, K. (2005) Loss of acetylation at Lysl6 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nature Genetics,* 37 (4), 391-400.

Fuchs, E. & Chen, T. (2013) A matter of life and death: Self-renewal in stem cells. *EMBO Reports,* 14 (1), 39-48.

Fursova, N. A., Blackledge, N. P., Nakayama, M., Ito, S., Koseki, Y., Farcas, A. M., King, H. W., Koseki, H. & Klose, R. J. (2019) Synergy between variant PRC1 complexes defines polycomb-mediated gene repression. *Molecular Cell,* 74 (5), 1020-1036.e8.

Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., Cerami, E., Sander, C. & Schultz, N. (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science Signaling,* 6 (269), pl1.

Gaughan, L., Stockley, J., Coffey, K., O'Neill, D., Jones, D. L., Wade, M., Wright, J., Moore, M., Tse, S., Rogerson, L. & Robson, C. N. (2013) KDM4B is a master regulator of the estrogen receptor signalling cascade. *Nucleic Acids Research,* 41 (14), 6892-6904.

Gil, J. & O'Loghlen, A. (2014) PRC1 complex diversity: Where is it taking us? *Trends in Cell Biology,* 24 (11), 632-641.

Gnant, M., Filipits, M., Greil, R., Stoeger, H., Rudas, M., Bago-Horvath, Z., Mlineritsch, B., Kwasny, W., Knauer, M., Singer, C., Jakesz, R., Dubsky, P., Fitzal, F., Bartsch, R., Steger, G., Balic, M., Ressler, S., Cowens, J. W., Storhoff, J., Ferree, S., Schaper, C., Liu, S., Fesl, C. & Nielsen, T. O. (2014) Predicting distant recurrence in receptor-positive breast cancer patients with limited clinicopathological risk: Using the PAM50 risk of recurrence score in 1478 postmenopausal patients of the ABCSG-8 trial treated with adjuvant endocrine therapy alone. *Annals of Oncology : Official Journal of the European Society for Medical Oncology,* 25 (2), 339-345.

Guiley, K. Z., Liban, T. J., Felthousen, J. G., Ramanan, P., Litovchick, L. & Rubin, S. M. (2015) Structural mechanisms of DREAM complex assembly and regulation. *Genes & Development,* 29 (9), 961-974.

Hanahan, D. & Weinberg, R. (2011) Hallmarks of cancer: The next generation. *Cell,* 144 (5), 646-674.

Henley, S. A. & Dick, F. A. (2012) The retinoblastoma family of proteins and their regulatory functions in the mammalian cell division cycle. *Cell Division,* 7 (1), 10.

Herceg, Z. & Wang, Z. (2001) Functions of poly(ADP-ribose) polymerase (PARP) in DNA repair, genomic integrity and cell death. *Mutation Research,* 477 (1-2), 97-110.

Hoadley, K. A., Yau, C., Hinoue, T., Wolf, D. M., Lazar, A. J., Drill, E., Shen, R., Taylor, A. M., Cherniack, A. D., Thorsson, V., Akbani, R., Bowlby, R., Wong, C. K., Wiznerowicz, M., Sanchez-Vega, F., Robertson, A. G., Schneider, B. G., Lawrence, M. S., Noushmehr, H., Malta, T. M., Cancer Genome Atlas Network, Stuart, J. M., Benz, C. C. & Laird, P. W. (2018) Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer.

Hu, Z., Fan, C., Oh, D. S., Marron, J. S., He, X., Qaqish, B. F., Livasy, C., Carey, L. A., Reynolds, E., Dressler, L., Nobel, A., Parker, J., Ewend, M. G., Sawyer, L. R., Wu, J., Liu, Y., Nanda, R., Tretiakova, M., Ruiz Orrico, A., Dreher, D., Palazzo, J. P., Perreard, L., Nelson, E., Mone, M., Hansen, H., Mullins, M., Quackenbush, J. F., Ellis, M. J., Olopade, O. I., Bernard, P. S. & Perou, C. M. (2006) The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics,* 7 (1), 96.

Huang, B., Warner, M. & Gustafsson, J. (2015) Estrogen receptors in breast carcinogenesis and endocrine therapy. *Molecular and Cellular Endocrinology,* 418 240-244.

Huang, J. M. & Hornyak, T. J. (2015) Polycomb group proteins – epigenetic repressors with emerging roles in melanocytes and melanoma. *Pigment Cell and Melanoma Research,* 28 (3), 330-339.

Huang, J. & Manning, B. D. (2008) The TSC1–TSC2 complex: A molecular switchboard controlling cell growth. *Biochemical Journal,* 412 (2), 179-190.

Hudis, C. A. & Gianni, L. (2011) Triple-Negative breast cancer: An unmet medical need. *The Oncologist (Dayton, Ohio),* 16 (S1), 1-11.

Hurtado, A., Holmes, K. A., Ross-Innes, C. S., Schmidt, D. & Carroll, J. S. (2011) FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nature Genetics,* 43 (1), 27-33.

Iness, A. N. & Litovchick, L. (2018) MuvB: A key to cell cycle control in ovarian cancer. *Frontiers in Oncology,* 8 223.

Iqbal, M. A., Siddiqui, S., Ur Rehman, A., Siddiqui, F. A., Singh, P., Kumar, B. & Saluja, D. (2021) Multiomics integrative analysis reveals antagonistic roles of CBX2 and CBX7 in metabolic reprogramming of breast cancer. *Molecular Oncology,* .

Jones, D., Wilson, L., Thomas, H., Gaughan, L. & Wade, M. A. (2019) The histone demethylase enzymes KDM3A and KDM4B co-operatively regulate chromatin transactions of the estrogen receptor in breast cancer. *Cancers,* 11 (8), 1122.

Kent, L. N., Bae, S., Tsai, S., Tang, X., Srivastava, A., Koivisto, C., Martin, C. K., Ridolfi, E., Miller, G. C., Zorko, S. M., Plevris, E., Hadjiyannis, Y., Perez, M., Nolan, E., Kladney, R., Westendorp, B., de Bruin,

A., Fernandez, S., Rosol, T. J., Pohar, K. S., Pipas, J. M. & Leone, G. (2017) Dosage-dependent copy number gains in E2f1 and E2f3 drive hepatocellular carcinoma. *The Journal of Clinical Investigation,* 127 (3), 830-842.

Kent, L. N., Rakijas, J. B., Pandit, S. K., Westendorp, B., Chen, H., Huntington, J. T., Tang, X., Bae, S., Srivastava, A., Senapati, S., Koivisto, C., Martin, C. K., Cuitino, M. C., Perez, M., Clouse, J. M., Chokshi, V., Shinde, N., Kladney, R., Sun, D., Perez-Castro, A., Matondo, R. B., Nantasanti, S., Mokry, M., Huang, K., Machiraju, R., Fernandez, S., Rosol, T. J., Coppola, V., Pohar, K. S., Pipas, J. M., Schmidt, C. R., de Bruin, A. & Leone, G. (2016) E2f8 mediates tumor suppression in postnatal liver development. *The Journal of Clinical Investigation,* 126 (8), 2955-2969.

Khan, S. A., Reddy, D. & Gupta, S. (2015) Global histone post-translational modifications and cancer: Biomarkers for diagnosis, prognosis and treatment? *World Journal of Biological Chemistry,* 6 (4), 333-345.

Klauke, K., Radulovic, V., Broekhuis, M., Weersing, E., Zwart, E., Olthof, S., Ritsema, M., Bruggeman, S., Wu, X., Helin, K., Bystrykh, L. & de Haan, G. (2013) Polycomb cbx family members mediate the balance between haematopoietic stem cell self-renewal and differentiation. *Nature Cell Biology,* 15 (4), 353-+.

Knight, J. R. P., Alexandrou, C., Skalka, G. L., Vlahov, N., Pennel, K., Officer, L., Teodosio, A., Kanellos, G., Gay, D. M., May-Wilson, S., Smith, E. M., Najumudeen, A. K., Gilroy, K., Ridgway, R. A., Flanagan, D. J., Smith, R. C. L., McDonald, L., MacKay, C., Cheasty, A., McArthur, K., Stanway, E., Leach, J. D., Jackstadt, R., Waldron, J. A., Campbell, A. D., Vlachogiannis, G., Valeri, N., Haigis, K. M., Sonenberg, N., Proud, C. G., Jones, N. P., Swarbrick, M. E., McKinnon, H. J., Faller, W. J., Le Quesne, J., Edwards, J., Willis, A. E., Bushell, M. & Sansom, O. J. (2021) MNK inhibition sensitizes KRAS -mutant colorectal cancer to mTORC1 inhibition by reducing eIF4E phosphorylation and c-MYC expression. *Cancer Discovery,* 11 (5), 1228-1247.

Kundu, S., Ji, F., Sunwoo, H., Jain, G., Lee, J. T., Sadreyev, R. I., Dekker, J. & Kingston, R. E. (2017) Polycomb repressive complex 1 generates discrete compacted domains that change during differentiation. *Molecular Cell,* 65 (3), 432-446.e5.

Kwiatkowski, D. J. & Wagle, N. (2015) mTOR inhibitors in cancer: What can we learn from exceptional responses? *EBioMedicine,* 2 (1), 2-4.

Lawrence, M., Daujat, S. & Schneider, R. (2015) Lateral thinking: How histone modifications regulate gene expression. *Trends in Genetics,* 32 (1), 42-56.

Liang, Y., Lin, H., Chen, C. & Zeng, D. (2017) Prognostic values of distinct CBX family members in breast cancer. *Oncotarget,* 8 (54), 92375-92387.

Liban, T. J., Thwaites, M. J., Dick, F. A. & Rubin, S. M. (2016) Structural conservation and E2F binding specificity within the retinoblastoma pocket protein family. *Journal of Molecular Biology,* 428 (20), 3960-3971.

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. & Tamayo, P. (2015) The molecular signatures database hallmark gene set collection. *Cell Systems,* 1 (6), 417-425.

Liu, M., Miller, C. L. & Eaves, C. J. (2012) Human long-term culture initiating cell assay. In Anonymous *Basic cell culture protocols.* Totowa, NJ: Humana Press, 241-256.

Long, X. & Nephew, K. P. (2006) Fulvestrant (ICI 182,780)-dependent interacting proteins mediate immobilization and degradation of estrogen receptor-α. *The Journal of Biological Chemistry,* 281 (14), 9607-9615.

Macek, B., Forchhammer, K., Hardouin, J., Weber-Ban, E., Grangeasse, C. & Mijakovic, I. (2019) Protein post-translational modifications in bacteria. *Nature Reviews. Microbiology,* 17 (11), 651-664.

Mao, J., Tian, Y., Wang, C., Jiang, K., Li, R., Yao, Y., Zhang, R., Sun, D., Liang, R., Gao, Z., Wang, Q. & Wang, L. (2019) CBX2 regulates proliferation and apoptosis via the phosphorylation of YAP in hepatocellular carcinoma. *Journal of Cancer,* 10 (12), 2706-2719.

Milioli, H. H., Tishchenko, I., Riveros, C., Berretta, R. & Moscato, P. (2017) Basal-like breast cancer: Molecular profiles, clinical features and survival outcomes. *BMC Medical Genomics,* 10 (1), 19.

Mootha, V. K., Lindgren, C. M., Eriksson, K., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., Houstis, N., Daly, M. J., Patterson, N., Mesirov, J. P., Golub, T. R., Tamayo, P., Spiegelman, B., Lander, E. S., Hirschhorn, J. N., Altshuler, D. & Groop, L. C. (2003) PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genetics,* 34 (3), 267-273.

Morey, L., Santanach, A., Blanco, E., Aloia, L., Nora, E., Bruneau, B. & Di Croce, L. (2015) Polycomb regulates mesoderm cell fate-specification in embryonic stem cells through activation and repression mechanisms. *Cell Stem Cell,* 17 (3), 300-315.

Narui, K., Ishikawa, T., Shimizu, D., Yamada, A., Tanabe, M., Sasaki, T., Oba, M. S., Morita, S., Nawata, S., Kida, K., Mogaki, M., Doi, T., Tsugawa, K., Ogata, H., Ota, T., Kosaka, Y., Sengoku, N., Kuranami, M., Niikura, N., Saito, Y., Suzuki, Y., Suto, A., Arioka, H., Chishima, T., Ichikawa, Y., Endo, I. & Tokuda, Y. (2019) Anthracycline could be essential for triple-negative breast cancer: A randomised phase II study by the kanagawa breast oncology group (KBOG) 1101. *Breast (Edinburgh),* 47 1-9.

Nevins, J. R. (2001) The rb/E2F pathway and cancer. *Human Molecular Genetics,* 10 (7), 699-703.

Nichol, J. N., Dupéré-Richer, D., Ezponda, T., Licht, J. D. & Miller, J., W H (2016) H3K27 methylation: A focal point of epigenetic deregulation in cancer. *Advances in Cancer Research,* 131 59-95.

Normanno, N., Di Maio, M., De Maio, E., De Luca, A., de Matteis, A., Giordano, A. & Perrone, F. (2005) Mechanisms of endocrine resistance and novel therapeutic strategies in breast cancer. *Endocrine-Related Cancer,* 12 (4), 721-747.

Osborne, C. K. & Schiff, R. (2011) Mechanisms of endocrine resistance in breast cancer. *Annual Review of Medicine,* 62 (1), 233-247.

O'Shaughnessy, J., Osborne, C., Pippen, J. E., Yoffe, M., Patt, D., Rocha, C., Koo, I. C., Sherman, B. M. & Bradley, C. (2011) Iniparib plus chemotherapy in metastatic triple-negative breast cancer. *The New England Journal of Medicine,* 364 (3), 205-214.

Paterni, I., Granchi, C., Katzenellenbogen, J. A. & Minutolo, F. (2014) Estrogen receptors alpha (ERα) and beta (ERβ): Subtype-selective ligands and clinical potential. *Steroids,* 90 13-29.

Pentimalli, F., Forte, I. M., Esposito, L., Indovina, P., Iannuzzi, C. A., Alfano, L., Costa, C., Barone, D., Rocco, G. & Giordano, A. (2018) RBL2/p130 is a direct AKT target and is required to induce apoptosis upon AKT inhibition in lung cancer and mesothelioma cell lines. *Oncogene,* 37 (27), 3657-3671.

Pereira, B., Chin, S., Rueda, O. M., Vollan, H. M., Provenzano, E., Bardwell, H. A., Pugh, M., Jones, L., Russell, R., Sammut, S., Tsui, D. W. Y., Liu, B., Dawson, S., Abraham, J., Northen, H., Peden, J. F., Mukherjee, A., Turashvili, G., Green, A. R., McKinney, S., Oloumi, A., Shah, S., Rosenfeld, N., Murphy, L., Bentley, D. R., Ellis, I. O., Purushotham, A., Pinder, S. E., Børresen-Dale, A., Earl, H. M., Pharoah, P. D., Ross, M. T., Aparicio, S. & Caldas, C. (2016) The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nature Communications,* 7 11479.

Perou, C. M., Sørlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., Fluge, Ø, Pergamenschikov, A., Williams, C., Zhu, S. X., Lønning, P. E., Børresen-Dale, A., Brown, P. O. & Botstein, D. (2000) Molecular portraits of human breast tumours. *Nature (London),* 406 (6797), 747-752.

Pherson, M., Misulovin, Z., Gause, M., Mihindukulasuriya, K., Swain, A. & Dorsett, D. (2017) Polycomb repressive complex 1 modifies transcription of active genes. *Science Advances,* 3 (8), e1700944.

Piqué, D. G., Montagna, C., Greally, J. M. & Mar, J. C. (2019) A novel approach to modelling transcriptional heterogeneity identifies the oncogene candidate CBX2 in invasive breast carcinoma. *British Journal of Cancer,* 120 (7), 746-753.

Rakha, E. A., El-Sayed, M. E., Green, A. R., Paish, E. C., Powe, D. G., Gee, J., Nicholson, R. I., Lee, A. H. S., Robertson, J. F. R. & Ellis, I. O. (2007) Biologic and clinical characteristics of breast cancer with single hormone Receptor–Positive phenotype. *Journal of Clinical Oncology,* 25 (30), 4772-4778.

Sadasivam, S. & DeCaprio, J. A. (2013) The DREAM complex: Master coordinator of cell cycle-dependent gene expression. *Nature Reviews. Cancer,* 13 (8), 585-595.

Saxton, R. A. & Sabatini, D. M. (2017) mTOR signaling in growth, metabolism, and disease. *Cell,* 168 (6), 960-976.

Scelfo, A., Piunti, A. & Pasini, D. (2015) The controversial role of the polycomb group proteins in transcription and cancer: How much do we not understand polycomb proteins? *The FEBS Journal,* 282 (9), 1703-1722.

Schade, A. E., Fischer, M. & DeCaprio, J. A. (2019) RB, p130 and p107 differentially repress G1/S and G2/M genes after p53 activation. *Nucleic Acids Research,* 47 (21), 11197-11208.

Schiff, R., Massarweh, S., Shou, J. & Osborne, C. K. (2003) Breast cancer endocrine resistance. *Clinical Cancer Research,* 9 (1), .

Schmid, P., Adams, S., Rugo, H. S., Schneeweiss, A., Barrios, C. H., Iwata, H., Diéras, V., Hegg, R., Im, S., Shaw Wright, G., Henschel, V., Molinero, L., Chui, S. Y., Funke, R., Husain, A., Winer, E. P., Loi, S. & Emens, L. A. (2018) Atezolizumab and nab-paclitaxel in advanced triple-negative breast cancer. *The New England Journal of Medicine,* 379 (22), 2108-2121.

Shanmugam, M. K., Arfuso, F., Arumugam, S., Chinnathambi, A., Jinsong, B., Warrier, S., Wang, L. Z., Kumar, A. P., Ahn, K. S., Sethi, G. & Lakshmanan, M. (2018) Role of novel histone modifications in cancer. *Oncotarget,* 9 (13), 11414-11426.

Sharma, G. N., Dave, R., Sanadya, J., Sharma, P. & Sharma, K. K. (2010) Various types and management of breast cancer: An overview. *Journal of Advanced Pharmaceutical Technology & Research,* 1 (2), 109-126.

Shenghui, H., Nakada, D. & Morrison, S. J. (2009) Mechanisms of stem cell self-renewal. *Annual Review of Cell and Developmental,* 25 (1), 377-406.

Simó-Riudalbas, L. & Esteller, M. (2015) Targeting the histone orthography of cancer: Drugs for writers, erasers and readers. *British Journal of Pharmacology,* 172 (11), 2716-2732.

Slamon, D., Eiermann, W., Robert, N., Pienkowski, T., Martin, M., Press, M., Mackey, J., Glaspy, J., Chan, A., Pawlicki, M., Pinter, T., Valero, V., Liu, M., Sauter, G., von Minckwitz, G., Visco, F., Bee, V., Buyse, M., Bendahmane, B., Tabah-Fisch, I., Lindsay, M., Riva, A. & Crown, J. (2011) Adjuvant trastuzumab in HER2-positive breast cancer. *The New England Journal of Medicine,* 365 (14), 1273-1283.

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. *Proceedings of the National Academy of Sciences - PNAS,* 98 (19), 10869-10874.

Stice, J. P. & Knowlton, A. A. (2008) Estrogen, NFκB, and the heat shock response. *Molecular Medicine (Cambridge, Mass.),* 14 (7-8), 517-527.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S. & Mesirov, J. P. (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences - PNAS,* 102 (43), 15545-15550.

Sun, Y., Zhao, Z., Yang, Z., Xu, F., Lu, H., Zhu, Z., Shi, W., Jiang, J., Yao, P. & Zhu, H. (2017) Risk factors and preventions of breast cancer. *International Journal of Biological Sciences,* 13 (11), 1387-1397.

Tang, Z., Kang, B., Li, C., Chen, T. & Zhang, Z. (2019) GEPIA2: An enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Research,* 47 (W1), W556-W560.

The Cancer Genome Atlas Network (2012) Comprehensive molecular portraits of human breast tumors. *Nature,* 490 (7418), 61-70.

The Gene Ontology Consortium, Carbon, S., Thomas, P. D., Albou, L. P., Hill, D. P., Gaudet, P., Van Auken, K. & Among Others (2019) The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Research,* 47 (D1), D330-D338.

Tropberger, P. & Schneider, R. (2013) Scratching the (lateral) surface of chromatin regulation by histone modifications. *Nature Structural & Molecular Biology,* 20 (6), 657-661.

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., Djureinovic, D., Micke, P., Lindskog, C., Mardinoglu, A. & Ponten, F. (2017) A pathology atlas of the human cancer transcriptome. *Science,* 357 (6352), eaan2507.

Ullah, F., Khan, T., Ali, N., Malik, F. A., Kayani, M. A., Shah, S. T. A. & Saeed, M. (2015) Promoter methylation status modulate the expression of tumor suppressor (RbL2/p130) gene in breast cancer. *PloS One,* 10 (8), e0134687.

van den Boom, V., Rozenveld-Geugien, M., Bonardi, F., Malanga, D., van Gosliga, D., Heyink, A. M., Viglietto, G., Morrone, G., Fusetti, F., Vellenga, E. & Schuringa, J. J. (2013) Nonredundant and locus-specific gene repression functions of PRC1 paralog family members in human hematopoietic stem/progenitor cells. *Blood,* 121 (13), 2452-2461.

Waddington, C. H. (2012) The epigenotype. (1942). *International Journal of Epidemiology,* 41 (1), 10-13.

Wade, M. A., Jones, D., Wilson, L., Stockley, J., Coffey, K., Robson, C. N. & Gaughan, L. (2015) The histone demethylase enzyme KDM3A is a key estrogen receptor regulator in breast cancer. *Nucleic Acids Research,* 43 (1), 196-207.

Wheeler, L. J., Watson, Z. L., Qamar, L., Yamamoto, T. M., Post, M. D., Berning, A. A., Spillman, M. A., Behbakht, K. & Bitler, B. G. (2018) CBX2 identified as driver of anoikis escape and dissemination in high grade serous ovarian cancer. *Oncogenesis (New York, NY),* 7 (11), 92-14.

Williams, C. & Lin, C. (2013) Oestrogen receptors in breast cancer: Basic mechanisms and clinical implications. *Ecancermedicalscience,* 7 (1), 370.

Yin, L., Duan, J., Bian, X. & Yu, S. (2020) Triple-negative breast cancer molecular subtyping and treatment progress. *Breast Cancer Research : BCR,* 22 (1), 1-61.

Yu, M., Mazor, T., Huang, H., Huang, H., Kathrein, K., Woo, A., Chouinard, C., Labadorf, A., Akie, T., Moran, T., Xie, H., Zacharek, S., Taniuchi, I., Roeder, R., Kim, C., Zon, L., Fraenkel, E. & Cantor, A. (2012) Direct recruitment of polycomb repressive complex 1 to chromatin by core binding transcription factors. *Molecular Cell,* 45 (3), 330-343.

Zeng, M., Li, B., Yang, L. & Guan, Q. (2021) CBX2 depletion inhibits the proliferation, invasion and migration of gastric cancer cells by inactivating the YAP/β-catenin pathway. *Molecular Medicine Reports,* 23 (2), .

Zhang, W., Yan, W., You, G., Bao, Z., Wang, Y., Liu, Y., You, Y. & Jiang, T. (2012) Genome-wide DNA methylation profiling identifies ALDH1A3 promoter methylation as a prognostic predictor in G-CIMP− primary glioblastoma. *Cancer Letters,* 328 (1), 120-125.

Zhao, Y. & Garcia, B. A. (2015) Comprehensive catalog of currently documented histone modifications. *Cold Spring Harbor Perspectives in Biology,* 7 (9), a025064.

Zhao, Z. & Shilatifard, A. (2019) Epigenetic modifications of histones in cancer. *Genome Biology,* 20 (1), 245.

Zheng, S., Lv, P., Su, J., Miao, K., Xu, H. & Li, M. (2019) Overexpression of CBX2 in breast cancer promotes tumor progression through the PI3K/AKT signaling pathway. *American Journal of Translational Research,* 11 (3), 1668-1682.

Zhu, S., Zhao, D., Yan, L., Jiang, W., Kim, J., Gu, B., Liu, Q., Wang, R., Xia, B., Zhao, J. C., Song, G., Mi, W., Wang, R., Shi, X., Lam, H., Dong, X., Yu, J., Chen, K. & Cao, Q. (2018a) BMI1 regulates androgen receptor in prostate cancer independently of the polycomb repressive complex 1. *Nature Communications,* 9 (1), 500-13.

Zhu, W. & Nelson, C. M. (2013) Adipose and mammary epithelial tissue engineering. *Biomatter (Austin, TX),* 3 (3), e24630.

Zhu, Y., Gu, J., Li, Y., Peng, C., Shi, M., Wang, X., Wei, G., Ge, O., Wang, D., Zhang, B., Wu, J., Zhong, Y., Shen, B. & Chen, H. (2018b) MiR-17-5p enhances pancreatic cancer proliferation by altering cell cycle profiles via disruption of RBL2/E2F4-repressing complexes. *Cancer Letters,* 412 59-68.

Zoncu, R., Sabatini, D. M. & Efeyan, A. (2011) mTOR: From growth signal integration to cancer, diabetes and ageing. *Nature Reviews. Molecular Cell Biology,* 12 (1), 21-35.

7. Appendix

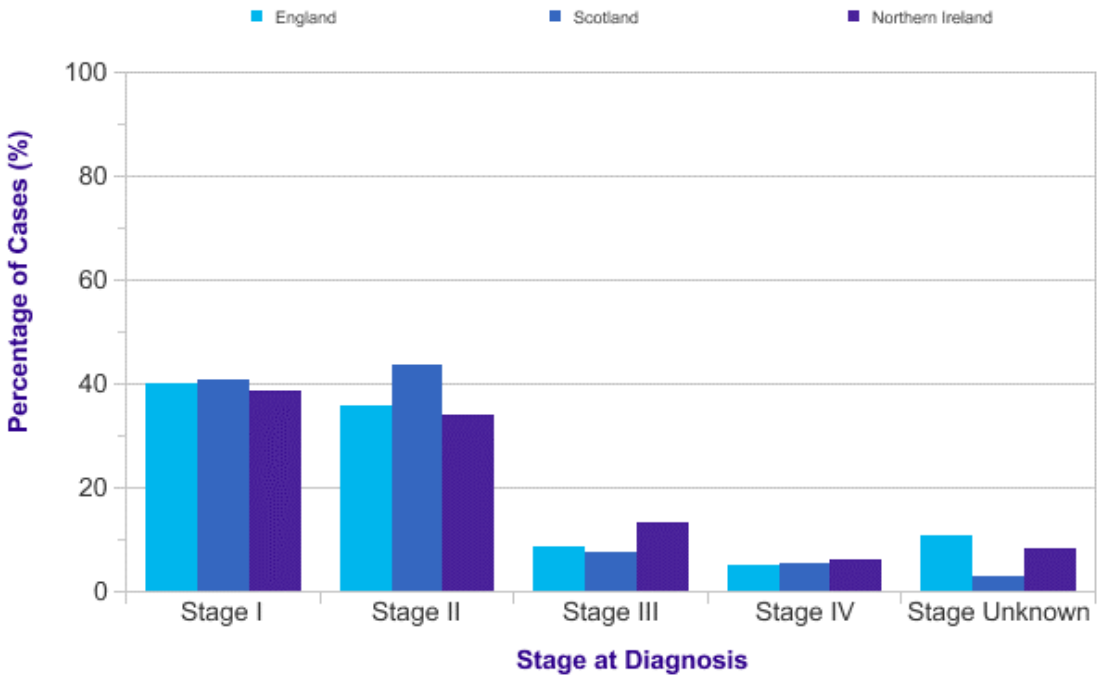

Figure 7.1. Overall stage of breast cancer at diagnosis.

Proportion of cases, of breast cancer stage at diagnosis, in England (2012-2014), Scotland (2014-2015) and Northern Ireland (2010-2014) for all ages. Taken from Cancer Research UK (2017).

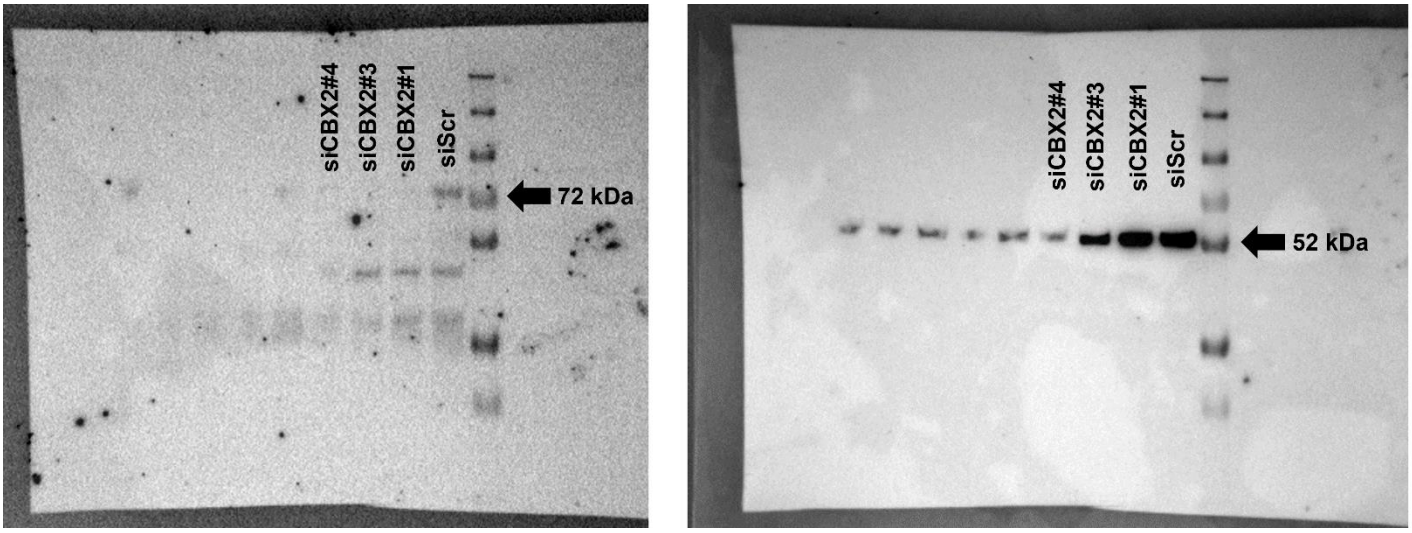

Figure 7.2. Full western blot from figure 4.21.
MDA-MB-231 cells were transfected with siScr or SiCBX2 (#1,#3 and #4) for 72 hours. Lysates of cells transfected by the different siRNAs are indicated above the wells. Arrows indicate the molecular weight. Western blot on the left is the membrane after incubation with anti-CBX2 primary antibody and shows that no band is present for the knockdown cells. Western blot on the right is the same membrane after incubation with alpha tubulin, showing clear bands for all samples.