Adaptation to chronic hypoxia in triple-negative breast cancer

Fiona Hartley

Michaelmas 2023

A thesis presented for the degree of Doctor of Philosophy

Acknowledgements

I would like to express my gratitude to my primary supervisor Prof Francesca Buffa, my secondary supervisor Prof Adrian Harris, and my bonus supervisor Dr Helen Sheldon for their guidance, support, and academic critique during this project. I would also like to thank the Medical Research Council, the Department of Oncology, and Exeter College who together funded this research.

I would also like to acknowledge my previous supervisors here in Oxford, Dr Karl Morten and Prof Joanna Poulton. Karl, for being incredibly supportive supervisor during a challenging project, and Jo, for hiring me as a research assistant with zero experience! Without your combined support I would never have started this DPhil.

Thank you to members of Buffa lab, particularly Anjali, Helen, and May for always being friendly faces around the lab. To Helen and Anjali for answering my never-ending questions and making sure everything is organised in the lab - you are stars! May, thank you for being a friend, for listening to my many, many rants, and for sharing in the agony that is the hypoxia chamber breaking down mid experiment; you made it bearable (just). Claudia, I will miss our lunchtime catch ups and library days. Thank you for meeting me in the mornings so I would get out of bed at a reasonable hour!

To my family for their support throughout my academic life. You'll be glad to know this will be my last degree.

My time in Oxford has been wonderful and that is in no small part due to Exeter College, who have provided great formals and, most importantly, great friends. To all those who spent many a late night in the MCR finishing off the post-dinner port, thank you. Those times hold a special place in my heart. I couldn't mention my life in Oxford without giving a shout out to Maddy, Hamish, and Nick, who lived through the craziness of lockdowns with me and made our house(s) in Cowley a home. A special thank you goes to Maddy, for being my dinner party co-host, joint inventor of the wine label tasting game, and always up for an adventure. You've been the best friend to go through my DPhil journey with.

Darren, I guess I have to write something nice here since you acknowledged me in your thesis... I know I can say anything to you. Thank you for being there for the last nine years.

And finally to Berklee, I'm a better person with you. Thank you for being by my side, always believing in me, and rescuing the pages of my thesis that blew into a hotel pool. You are truly the best.

Abstract

Low levels of oxygen (i.e. hypoxia) are often observed in solid tumours. Hypoxia is able to drive several of the cancer hallmarks and promotes resistance to both radio and chemotherapy, resulting in its association with poor prognosis. Triple-negative breast cancer has the worst prognosis of all breast cancer subtypes with a 5 year survival rate of only 76.7%, and only 10.8% if the cancer has already metastasised. In line with this, triple-negative tumours often have higher levels of hypoxia than their hormone receptor expressing counterparts.

The length and severity of hypoxia is heterogeneous both within and between tumours, meaning cells within the same tumour may experience both acute and chronic hypoxia. Acute hypoxia ($\leqslant$ 24 hours) is relatively well studied, whereas research into chronic hypoxia ($\geqslant$ 7 days) is lacking. This thesis therefore aimed to characterise the transcriptional response to chronic hypoxia using RNA sequencing. It also aimed to identify proteins or microRNAs which are necessary for adaptation to chronic hypoxia using genome-wide CRISPR screening. This method enables specific vulnerabilities in hypoxic cells to be identified, meaning drugs can be designed to target these “Achilles heels”.

The major pathways altered under chronic hypoxia were related to epithelial-mesenchymal transition (EMT) and cell adhesion, with triple-negative breast cancer cells undergoing an atypical form of EMT. However, none of the genes which were upregulated in response to chronic hypoxia were found to be essential for their survival. Instead, hypoxic cells had a greater dependence on general control nonderepressible 2 (GCN2), an instigator of the integrated stress response usually activated under amino acid starvation. GCN2 was found to be active under chronic hypoxia, and chemical inhibition of GCN2 reduced cell growth *in vitro*. Previous work has shown that GCN2 knockout mice are healthy unless fed a diet restricted in essential amino acids, meaning targeting GCN2 is unlikely to result in severe toxicities. Together, this means GCN2 constitutes a promising therapeutic target for hypoxic tumours. No microRNAs essential for survival under hypoxia were identified in this study.

Contents

List of tables

List of figures

Abbreviations

Akt	protein kinase B
ANOVA	analysis of variance
AP-1	activator protein 1
ASNS	asparagine synthetase
ATF4	activating transcription factor 4
ATF6	activating transcription factor 6
ATG7	autophagy related 7
ATP	adenosine triphosphate
BCL-2	B-cell lymphoma 2
Bcl-xL	B-cell lymphoma-extra large
Bcl-xS	B-cell lymphoma-extra small
BNIP3	Bcl-2 interacting protein 3
bp	base pairs
BSA	bovine serum albumin
CCK-8	Cell Counting Kit-8
CEG	core essential gene
ChIP	chromatin immunoprecipitation
CHOP	C/EBP homologous protein
CRISPR	clustered regularly interspaced short palidromic repeats
CTL	control
DEPC	diethyl pyrocarbonate
DepMap	Dependency Map
DMEM	Dulbecco's Modified Eagle Medium
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
dsDNA	double stranded DNA
DTT	dithiothreitol
E/M	epithelial/mesenchymal
ECL	enhanced chemiluminescence
ECM	extracellular matrix

EDTA	ethylenediaminetetraacetic acid
eIF2	eukaryotic translation initiation factor 2
EIF2AK1	eukaryotic translation initiation factor 2 alpha kinase 1
EIF2AK2	eukaryotic translation initiation factor 2 alpha kinase 2
EIF2AK3	eukaryotic translation initiation factor 2 alpha kinase 3
EIF2AK4	eukaryotic translation initiation factor 2 alpha kinase 4
ELISA	enzyme-linked immunosorbent assay
EMT	epithelial to mesenchymal transition
EMT-TF	EMT transcription factor
ER	endoplasmic reticulum
FBS	foetal bovine serum
FIH	factor inhibiting HIF
GADD34	growth arrest and DNA damage-inducible protein
GCN2	general control nonderepressible 2
GCN2i	GCN2 inhibitor
GDP	guanosine diphosphate
GPI	glucose-6-phosphate isomerase
GSEA	gene set enrichment analysis
GTP	guanosine triphosphate
HCL	hydrochloric acid
HIF	hypoxia inducible factor
HRE	hypoxia response element
HRI	heme-regulated inhibitor
HRP	horseradish peroxidase
IAP-2	inhibitor-of-apoptosis protein 2
ICAM1	intercellular adhesion molecule 1
ID	identification
IGF2	insulin-like growth factor 2
IgG	immunoglobulin G
INHBA	inhibin beta-A
IRE	inositol-requiring enzyme 1
ISR	integrated stress response

KEGG	Kyoto Encyclopedia of Genes and Genomes
LB	Luria-Bertani
MAGeCK	Model-based analysis of genome-wide CRISPR/Cas9 knockout
MAP1LC3B	microtubule-associated protein 1A/1B light chain 3B
MES	2-(N-morpholino)ethanesulfonic acid
MET	mesenchymal to epithelial transition
miRNA	microRNA
MOI	multiplicity of infection
MOPS	3-(N-morpholino)propanesulfonic acid
mRNA	messenger RNA
mTOR	mammalian target of rapamycin
NES	normalised enrichment score
P/S	penicillin streptomycin
PAM	protospacer adjacent motif
PBS	phosphate buffered saline
PC1	principal component 1
PC2	principal component 2
PCA	principal component analysis
PCR	polymerase chain reaction
PERK	protein kinase R-like endoplasmic reticulum kinase
PRK	protein kinase RNA-activated
p-GCN2	phosphorylated GCN2
PI3K	phosphoinositide 3-kinase
PLOD2	procollagen-lysine 2-oxoglutarate 5-dioxygenase 2
pLX-miR	Lin Human miRNA CRISPR Knockout Library
pre-miRNA	precursor miRNA
pri-miRNA	primary miRNA
PTEN	phosphatase and tensin homolog
PTPMT1	protein tyrosine phosphatase mitochondrial 1
PVDF	polyvinylidene difluoride
qPCR	quantitative PCR
QSOX1	quiescin sulfhydryl oxidase 1

RIPA	radio-immunoprecipitation assay
RISC	RNA-induced silencing complex
RNA	ribonucleic acid
RNA-seq	RNA sequencing
ROS	reactive oxygen species
rpm	revolutions per minute
RRA	robust ranking algorithm
rRNA	ribosomal RNA
S100A2	S100 calcium-binding protein A2
SAM	sequence alignment map
SDS	sodium dodecyl sulphate
sgRNA	single guide RNA
shRNA	short hairpin RNA
SOC	super optimal broth with catabolite repression
STAT3	signal transducer and activator of transcription 3
TAE	tris-acetate-EDTA
TBS	tris buffered saline
TBS-T	tris buffered saline Tween 20
TCA	tricarboxylic acid
TCGA	the cancer genome atlas
TDI	Target Discovery Institute
TE	tris-EDTA
TET	ten-eleven translocation
TGF-β	transforming growth factor beta
TKOv3	Toronto Knockout Library version 3
Tm	melting temperature
TNBC	triple-negative breast cancer
TNFSF13	tumour necrosis factor ligand superfamily member 13
tRNA	transfer RNA
TU	transducing units
uORF	upstream open reading frame
UPR	unfolded protein response

UTR	untranslated region
VEGF	vascular endothelial growth factor
VEGFR	vascular endothelial growth factor receptor
VHL	von Hippel–Lindau
xCT	cystine/glutamate antiporter
ZEB1	zinc finger E-box binding homeobox 1
ZEB2	zinc finger E-box binding homeobox 2

1 Introduction

1.1 Cancer

Cancer is the cause of almost 10 million deaths per year worldwide, and the incidence rates are predicted to increase by 47% over the next two decades [1]. The disease is characterised by uncontrolled cell growth and a propensity for cells to migrate and grow in other parts of the body (metastasise). Alterations of cellular processes which usually keep cell growth under control allow cancers to develop. These processes are typically altered through the acquisition of DNA mutations which affect the control of key cell signalling cascades.

1.1.1 Breast cancer

Breast cancer is the most commonly diagnosed cancer worldwide and is the leading cause of cancer related deaths in women [1]. In the UK alone, there are over 55,000 cases per year [2]. Although the 10 year survival rate for breast cancer is 76%, improving from only 40% in the 1970s, the survival rate is highly dependent on the subtype of breast cancer. There are four main subtypes of breast cancer which are categorised by the expression of nuclear and cell surface receptors (Table 1). The three receptors are the oestrogen hormone receptor, the progesterone hormone receptor, and the human epidermal growth factor receptor 2 (HER2). Expression of the hormone receptors is associated with better prognosis with a 5 year survival rate of over 90% [3]. Conversely, a lack of hormone receptor expression along with a lack of HER2 amplification results in the worst prognosis, with a 5 year survival rate of only 76.7% [3]. This subtype is named triple-negative breast cancer (TNBC). TNBC has poor prognosis even when diagnosed at an early stage [3]. This is in part due to TNBC being more invasive and having greater metastatic potential, but also due to the lack of expression of these receptors which means specialised treatments are more difficult to develop [4]. Even in metastatic cancers, TNBC has a poorer 5 year survival rate than non-TNBC (10.81% vs 33.46%) [5], and a higher rate of disease recurrence [6].

Six subtypes of TNBC have been described [6]. The six subtypes are basal-like 1 (BL1), basal-like 2 (BL2), immunomodulatory (IM), mesenchymal (M), mesenchymal stem-like (MSL), and luminal androgen receptor positive (LAR). The two basal subtypes are separated by expression

of cell cycle and DNA damage response genes (BL1) and expression of growth factor signalling genes and myoepithelial markers (BL2). Though the two mesenchymal subtypes (M and MSL) both have high expression of genes involved in epithelial-mesenchymal-transition (EMT), the MSL subtype is distinguished through lower levels of genes involved in proliferation. The IM subtype expresses genes involved in cytokine and immune signal transduction pathways while LAR is charcterised by the androgen response pathway and has a luminal phenotype [7]. This heterogeneity means that specialised treatments which target all TNBCs are difficult to develop and have had limited success in clinical trials [6].

Table 1: Breast cancer subtypes. Data are based on U.S. cancer statistics [3]. The subtype of 8% of cancers was unknown.

Subtype	Hormone receptor	HER2 enriched	Incidence rate	5 year survival rate
Luminal A	Positive	No	68%	94.1%
Luminal B	Positive	Yes	10%	90.4%
Triple-negative	Negative	No	10%	76.7%
HER2	Negative	Yes	4%	83.6%

1.2 Hypoxia

Oxygen concentration varies naturally in the body, with some organs having a greater partial pressure of oxygen than others under normal conditions (Figure 1) [8]. This variation in physiological oxygen levels means that oxygen insufficiency, termed hypoxia, does not have a set point, and what constitutes hypoxia in one tissue may not in another [9]. As such, hypoxia may be defined as an inadequacy in oxygen supply which is sufficient to trigger a biological response. This can occur during normal development (e.g. blood vessel formation [10]) or under pathological conditions such as cardiac arrest, ischaemic stroke, and tumour development [11, 12].

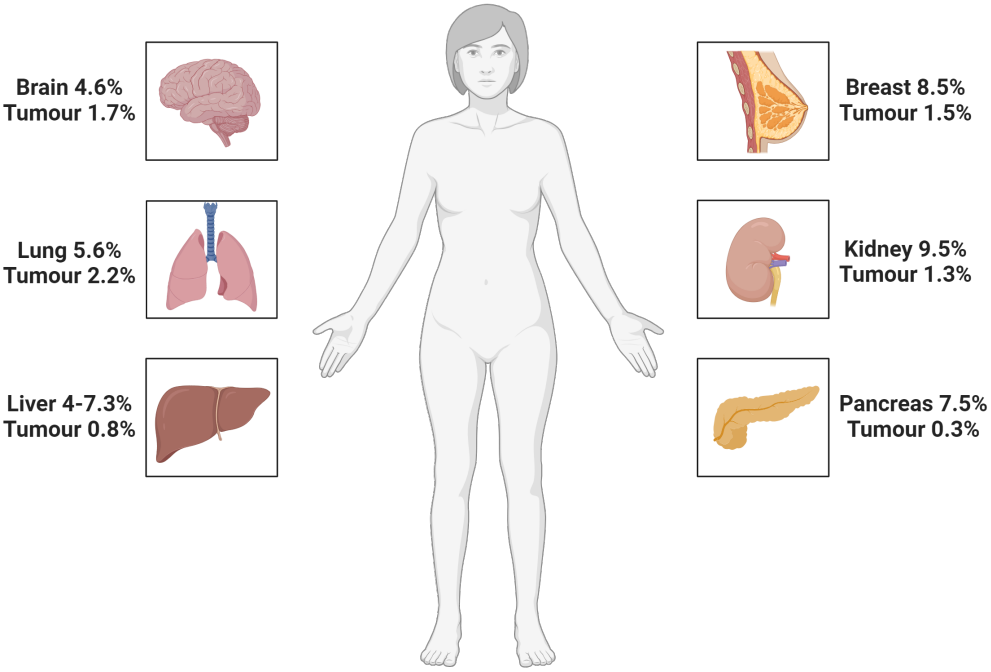

Figure 1: Tissue and tumour oxygenation. Physiological oxygen levels in organs and their associated tumours. Figure created using BioRender using oxygen concentration values from Muz et al. [8].

1.2.1 Tumour hypoxia

Hypoxia is found in 50-60% of solid tumours [13] and occurs when the oxygen demand of a tumour outweighs its supply [14]. This can lead to oxygen tensions as low as 0-20 mmHg (approximately 0-2.5% O_2) [14–16]. Tumour hypoxia occurs via four main mechanisms. The first is diffusion restricted hypoxia which occurs when the distance between cells within the

tumour and the closest blood vessel reaches 70-100 µm (Figure 2) [14, 15]. If the distance to the closest vessel exceeds 150-180 µm the lack of oxygen and nutrients can result in tumour cells becoming necrotic [14, 15]. The second mechanism leading to tumour hypoxia is perfusion restriction [14]. This is caused by poorly developed tumour vasculature which leads to local and temporary disturbances in blood flow [13–15]. This results in intermittent interruptions to oxygen supply. Irregularities in the vasculature can also lead to arteriovenous shunting of blood away from the tumour [17]. Finally, anaemia in the patient can lead to tumour hypoxia through a decrease in oxygen carrying capacity of blood in the whole body [14, 18]. Tumours are more vulnerable to anaemia induced hypoxia due to the aforementioned issues which reduce spare capacity. These differing mechanisms mean there is heterogeneity in both the length and severity of hypoxia experienced within a single tumour.

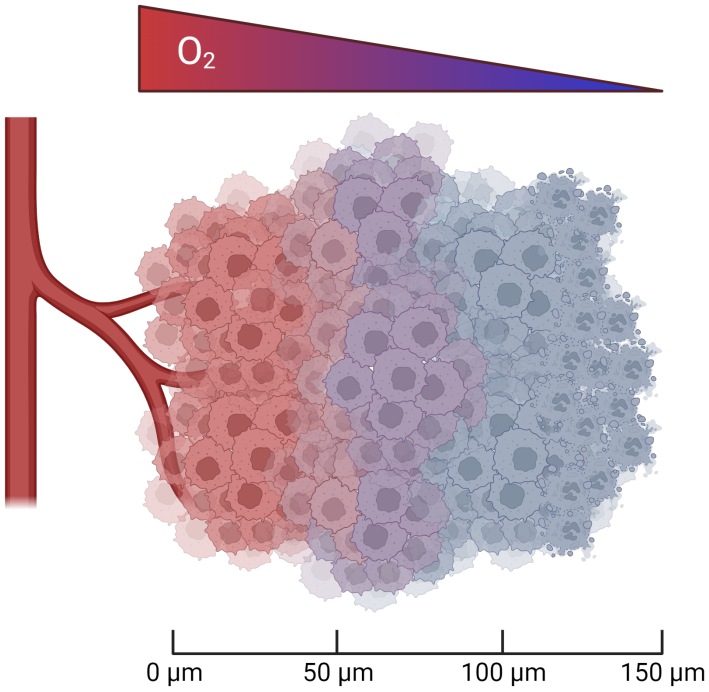

Figure 2: Hypoxia in tumours. Tumour hypoxia arises 70-100 µm from the closest vessel, and necrosis occurs at a distance of 150-180 µm [14, 15]. Figure created using BioRender.

The types of hypoxia experienced by a tumour can be broadly categorised into three subtypes: acute, chronic, and intermittent [8, 13]. Intermittent, or cycling, hypoxia requires a period of hypoxia to be followed by a period of reoxygenation which is typically repeated in subsequent

cycles. *In vivo*, this type of hypoxia occurs in cells that are experiencing perfusion limited hypoxia and cycles can last from minutes to days [13, 14]. In contrast, acute and chronic hypoxia both require a consistent low oxygen level to be maintained and are distinguished only by the amount of time spent under hypoxia. However, the amount of time required to distinguish between the two is not clearly defined. Many researchers use opposing terminology to refer to hypoxia exposures of the same length of time [9]. For example, exposure times for acute hypoxia start at as little as 1 hour [19] and reach up 72 hours [20]. Meanwhile, exposure times for chronic hypoxia range from as low as 16 hours [21] up to 8 weeks [22]. The overlap in these time spans demonstrates a lack of consensus in the field over what constitutes acute or chronic hypoxia. The lack of a clear definition hampers differences in their downstream effects from being identified in the literature.

In vivo, chronic hypoxia is likely to arise as a result of restrictions in diffusion, rather than perfusion, since cells remain at a distance from the vessel for prolonged periods of time [13, 14, 18]. On the other hand, short exposures to hypoxia are more likely to be caused by perfusion limitations, meaning acute hypoxia is likely to exist as part of a cycling hypoxia regime. Despite this, there is an abundance of research into the acute (not intermittent) hypoxia response, whereas the response to chronic hypoxia is not well characterised [12]. Few studies characterise the hypoxia response beyond 24 hours [12], and a systematic review conducted in 2020 showed there were only four transcriptomic studies which exposed cells to hypoxia for $\geqslant$ 7 days (Di Giovannantonio, M. 2020, unpublished). These datasets will be discussed further in Section 4.

1.3 Control of gene expression under hypoxia

Hypoxia drives major changes in the cell and regulates hundreds of genes across a myriad of functions [12, 23]. These changes allow the cell to adapt to the hypoxic microenvironment and are driven by a combination of mechanisms acting at the transcriptional, post-transcriptional, and translational levels.

1.3.1 Hypoxia driven transcriptional regulation

The response to hypoxia is predominantly mediated by the hypoxia inducible factor (HIF) (Figure 3) [12]. HIFα is constitutively expressed but under normal oxygen concentrations is quickly degraded. The prolyl hydroxylase (PHD) enzyme hydroxylates HIFα in an oxygen-dependent manner. This modification is recognised by von Hippel–Lindau (VHL) which ubiquitinates HIFα, tagging it for proteasomal degradation [15, 24, 25]. When oxygen levels drop, PHD is no longer able to hydroxylate HIFα, meaning HIFα is not tagged for degradation and instead accumulates in the cell. HIFα translocates to the nucleus and forms a heterodimer with HIF1β. The HIF heterodimer functions as a transcription factor and recognises hypoxia response elements (HRE) in the DNA, driving transcription of its target genes. Factor inhibiting HIF (FIH) mediates a second HIF regulating mechanism. FIH is an asparaginyl hydroxylase enzyme which, like PHD, functions in an oxygen-dependent manner. It hydroxylates HIFα and prevents it from binding to its transcriptional coactivator, p300, thus blocking its transcriptional activity [15, 24, 25]. Drastic changes in the transcriptome of the cell occur within the first few hours of hypoxia exposure, predominantly due to HIF activity [26].

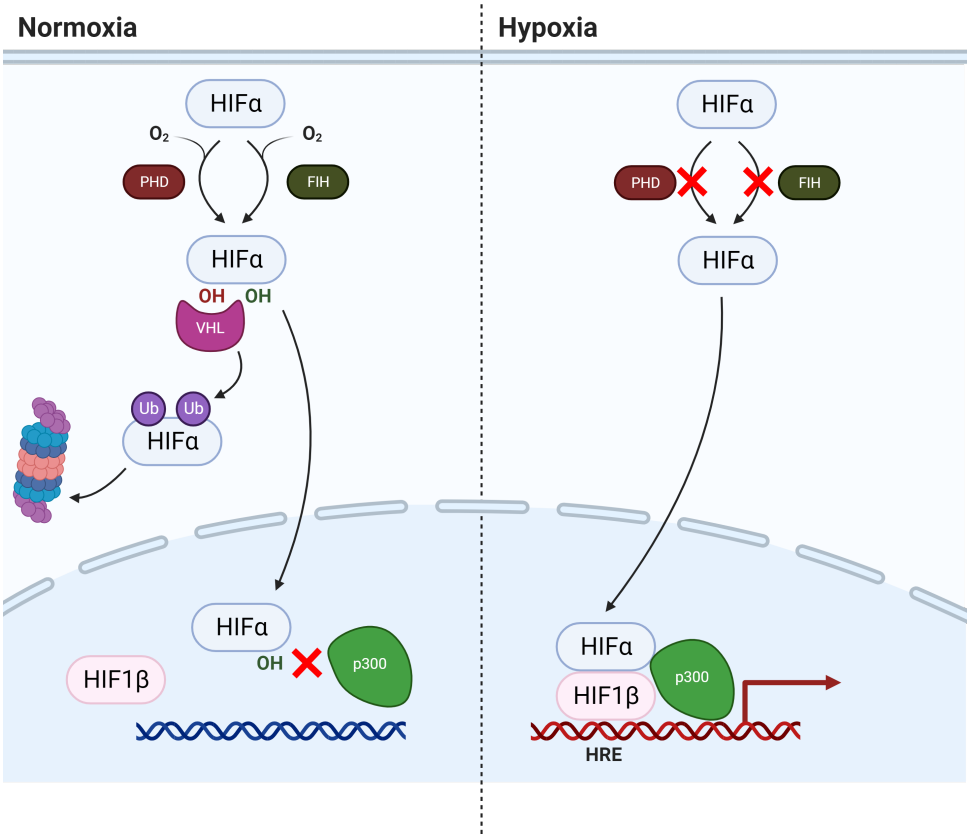

Figure 3: HIF signalling pathway. Figure created using BioRender.

There are three HIFα subtypes each encoded by different genes: HIF1α, HIF2α, and HIF3α. These all interact with the same HIF1β protein to form heterodimers. Less is known about HIF3α [27], and HIF1α and HIF2α are known to be the main propagators of the hypoxic response. Though there is a great deal of overlap in the function of the two HIF subtypes, with both able to upregulate many of the key hypoxia response genes (e.g. glucose transporter, vascular endothelial growth factor (VEGF) [24, 28]), their functions are non-redundant [25]. For example, some genes are only activated by one of the two subtypes (e.g. carbonic anhydrase IX is activated by HIF1 exclusively [29, 30]). Additionally, the HIF subtypes have been shown to respond differently to both the level of oxygen and the length of time spent under hypoxia [24, 25]. Typically, HIF1 is stabilised sooner after hypoxia exposure but degrades after less than 24 hours of hypoxia exposure, while HIF2 accumulates later and has been shown to stabilise under milder hypoxia than HIF1 (1-5% O_2) [12, 24, 25, 31–33].

Aside from HIF driven transcription of target genes, hypoxia can also regulate transcription through epigenetic modifications [23]. Jumonji-C histone demethylases use oxygen as a cofactor and are therefore less active under hypoxia. This results in a HIF-independent increase in histone methylation under hypoxia [34, 35], allowing transcriptional repression to occur under hypoxia as well as activation.

It is important to note that the transcriptional response to hypoxia is also influenced by the existing epigenetic landscape of the cell [23]. There are millions of HREs present in the human genome, but the HIF proteins in a given cell bind to only a small proportion of these [35]. This means that chromatin accessibility is a driver of the HIF response and results in tissue specific responses to hypoxia [23, 36].

1.3.2 Hypoxia driven post-transcriptional regulation

Hypoxia can also alter gene expression at the post-transcriptional level. Alternate splicing events are common in hypoxia, affecting hundreds or thousands of genes depending on the cell type [37, 38]. This means that a gene whose expression level does not change under hypoxia can still behave in a different manner due to the presence of new isoforms. For example, the anti-apoptotic protein tumour necrosis factor ligand superfamily member 13 (TNFSF13) undergoes alternative splicing under hypoxia that results in intron retention and a reduction in

the functional protein [39].

In addition to this, hypoxia can affect mRNA stability through regulation of microRNAs (miRNAs) [40]. miRNAs are small (approximately 22 nucleotide) RNA molecules which are loaded into the RNA-induced silencing complex (RISC). The RISC binds to specific mRNA molecules depending on base pairing between the miRNA and the 3' untranslated region (UTR) of the target mRNA [41]. When the RISC binds to an mRNA molecule in this manner, it either directly suppresses translation and prevents the protein product being formed, or it destabilises the mRNA through deadenylation and decapping resulting in mRNA degradation [42]. In this way, miRNAs attenuate expression of their target genes and affect the phenotype of the cell.

Hypoxia is known to trigger extensive changes in miRNA expression profiles and thus impacts the cell phenotype through miRNA regulation [40]. For example, miR-210 is universally detected under hypoxia and is the best characterised hypoxia response miRNA [43]. There is evidence supporting the involvement of miR-210 in various aspects of tumour biology (e.g. apoptosis, angiogenesis, DNA damage response, cell cycle) [44]. Downregulation of miRNAs can also occur under hypoxia (e.g. miR-125a, miR-150), allowing their mRNA targets to be upregulated [45, 46].

1.3.3 Regulation of protein translation under hypoxia

Hypoxia also exerts control over the cell through regulation of protein translation. Translation of mRNA can be controlled by miRNAs as described above. Additionally, as with other stress responses, a global decrease in protein translation is triggered under hypoxia. This is mediated through inhibition of mammalian target of rapamycin (mTOR) signalling and, under more extreme cases of hypoxia, initiation of the unfolded protein response (UPR) [47, 48]. The UPR is comprised of three major signalling cascades: one mediated by activating transcription factor 6 (ATF6), one mediated by inositol-requiring enzyme 1 (IRE1), and another mediated by protein kinase R-like endoplasmic reticulum kinase (PERK) which triggers the integrated stress response (ISR) [48]. The ISR is particularly relevant for this project and is discussed in detail in the following section.

1.4 The integrated stress response

The ISR is an intracellular system which senses different types of stress, both intrinsic and extrinsic, and responds by enacting a coordinated cellular response. There are four eukaryotic initiation factor 2α (eIF2α) kinases which detect stress [49]. They are activated in response to different types of stress, but they each phosphorylate eIF2α at serine 51 to trigger the ISR (Figure 4). eIF2α is a key component of the translational machinery. When phosphorylated, eIF2α is no longer able to assemble into the ternary complex with tRNA-methionine [49]. This results in a global reduction in cap-dependent protein synthesis. This in itself alleviates stress by conserving energy, preserving amino acids, preventing viral RNA translation, and slowing protein entry to the endoplasmic reticulum (ER) to prevent the accumulation of more unfolded proteins. In addition, certain stress response proteins (e.g. activating transcription factor 4 (ATF4)) are selectively translated during this time. These proteins are found on mRNA molecules which contain inhibitory upstream open reading frames (uORFs) [50]. The mRNA for ATF4 contains two uORFs (Figure 5). Under normal conditions, the ribosome translates uORF1, then immediately re-initiates translation at uORF2. uORF2 overlaps with the ATF4 ORF out of frame, so translation of uORF2 prevents translation of ATF4. Under stress conditions, lower concentrations of the ternary complex mean that ribosomes are more likely to scan through uORF2 and initiate at ATF4. In this manner, cells can upregulate stress response genes even though global protein translation is reduced.

ATF4 is able to dimerise with other proteins and drive the transcription of genes which alleviate cell stress [49]. For example, the ISR can trigger G1 arrest [51] which preserves nutrients/energy and allows time for the stress to be resolved. ATF4 can also increase autophagy through transcription of microtubule-associated protein 1A/1B light chain 3B (MAP1LC3B) and autophagy related 7 (ATG7), replenishing cellular building blocks [52]. ATF4 mitigates amino acid starvation by increasing expression of amino acid synthesis/transport systems such as asparagine synthetase (ASNS) and the cystine/glutamate antiporter (xCT) [52]. Like HIF, ATF4 upregulates VEGF, thus promoting angiogenesis to increase the supply of amino acids, glucose, and oxygen [52]. In severe cases of cell stress, apoptosis may be triggered through C/EBP homologous protein (CHOP) [49]. ATF4 also drives the transcription of growth arrest and DNA damage-inducible protein 34 (GADD34), a protein involved in eIF2α dephosphorylation [49]. This means that a negative feedback loop exists so that when the

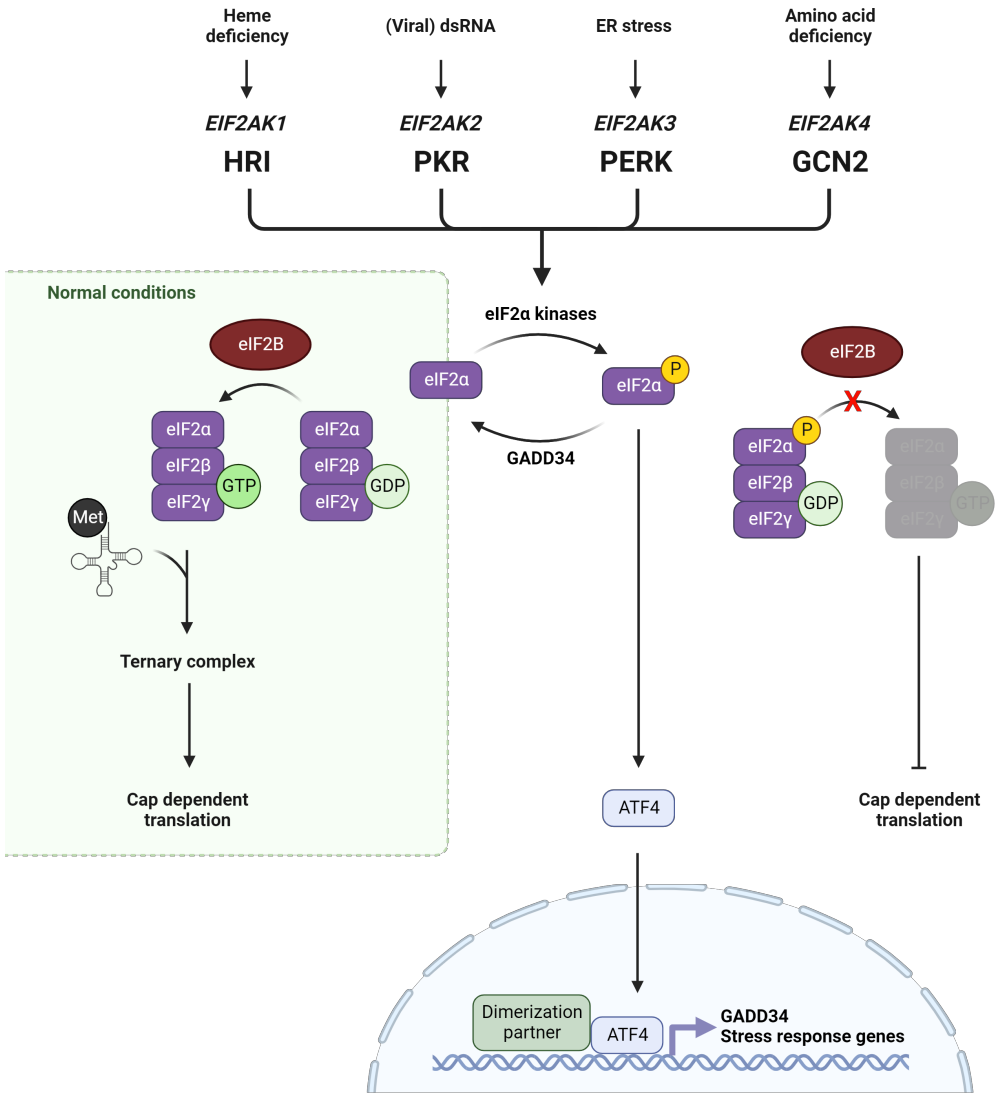

Figure 4: The integrated stress response. Under normal conditions, eIF2α is not phosphorylated at serine 51. It assembles into the eIF2 complex, and the guanine exchange factor eIF2B exchanges GDP for GTP. eIF2 with its bound GTP assembles with methionine-tRNA to form the ternary complex and cap-dependent translation ensues. Under stress conditions, this process is inhibited. Different stress conditions activate each of the four eIF2α kinases. The kinases autophosphorylate upon detection of their stress signal and go on to phosphorylate eIF2α. Phosphorylation of eIF2α prevents the guanine exchange factor eIF2B from exchanging GDP for GTP in the eIF2 complex. This inhibits formation of the ternary complex and therefore protein translation. ATF4 is selectively translated, and translocates to the nucleus where it binds to dimerisation partners and drives expression of stress response genes. GADD34 is also upregulated by ATF4. This protein functions with protein phosphatase 1 to dephosphorylate eIF2α, forming a negative feedback loop to restore normal protein translation. HRI = heme-regulated inhibitor. PKR = protein kinase RNA-activated. PERK = protein kinase R-like endoplasmic reticulum kinase. GCN2 = general control nonderepressible 2. Figure created using BioRender.

stress is resolved, normal protein synthesis may resume.

As hypoxia is a type of environmental stress, it is able to activate the ISR. Hypoxia causes an accumulation of unfolded proteins in the ER and stimulates the UPR [53]. This activates PERK, one of the four drivers of the ISR, which undergoes autophosphorylation when bound to unfolded proteins [54]. PERK is the canonical activator of the ISR under hypoxia [55, 56].

General control derepressible 2 (GCN2) is activated under amino acid starvation through binding of uncharged tRNAs [57]. These activate GCN2 by causing autophosphorylation of threonine 899 and threonine 904 (p-GCN2) [57]. There are other less well-documented activators of GCN2. For example, both PERK and GCN2 have been shown to respond to glucose starvation [55], although in the case of GCN2 this may be because glucose deprivation leads to rewired metabolism and subsequent amino acid depletion, as the phenotype was rescued by excess glutamine. Ultraviolet light [58, 59], ribosome stalling [60, 61], and reactive oxygen species (ROS) [62] have also been documented as GCN2 activators.

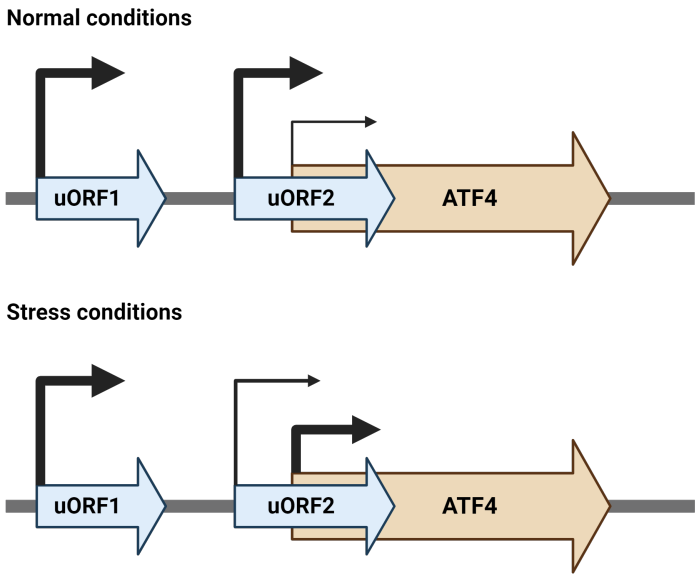

Figure 5: ATF4 translation under normal and stress conditions. Under normal conditions, translation of uORF2 precludes translation of ATF4. Under stress conditions, depletion of the ternary complex means ribosomes are more likely to scan through uORF2 and initiate at the ATF4 start codon. Figure created using BioRender.

There is a small amount of conflicting evidence regarding the activation of GCN2 under hypoxia [51, 55]. Liu et al. [51] show phosphorylation of eIF2α in hypoxia remains in both GCN2 and PERK knockout cells, but at lower levels than in wild-type cells. This indicates that both of these eIF2α kinases are involved in triggering the ISR under hypoxia. These researchers also showed that both GCN2 and PERK knockout cells had reduced survival under hypoxic conditions. This is in contrast to Ye et al. [55] who showed that eIF2α phosphorylation under hypoxia was eliminated in PERK knockout cells, but remained in GCN2 knockout cells. In accordance, their GCN2 knockouts did not show reduced survival under hypoxia. This group had previously shown that PERK and ATF4 facilitated cell survival under hypoxia, in accordance with a model of PERK activation of the ISR in these conditions [63].

Tumours are exposed to many stressors, with cells often experiencing hypoxia and nutrient deprivation due to insufficient vasculature. In accordance with this, there is increased expression of ATF4 in cancer tissues which often co-localises with hypoxic regions [63]. Knockout/knockdown of ATF4, GCN2, and PERK have all been shown to reduce tumour growth in animal models [55, 63]. Activity of the ISR has also been linked to treatment resistance in cancer. For example, inhibition of the ISR re-sensitised breast cancer cells to paclitaxel [57]. Together, these show the importance of the cell's ability to respond to stress in cancer progression.

1.5 Hypoxia and the hallmarks of cancer

The presence of hypoxia in a tumour is associated with poor prognosis in many cancers, including breast [28]. In part, this is conferred through increased resistance to radio and chemotherapy [15, 28]. Under hypoxia, the production of radiotherapy induced ROS is limited, meaning higher doses of radiation are required to kill hypoxic cells [64]. Meanwhile, the effectiveness of many chemotherapy drugs is reduced [65]. This is partly due to the lack of vasculature which reduces delivery of drugs to the hypoxic cells [15, 65]. However, even if drug delivery is effective, hypoxia triggers changes in the cell which render them chemoresistant. Hypoxia slows cell proliferation meaning drugs that target rapidly dividing cells are less effective. Additionally, HIF driven expression of chemoresistance proteins such as the p-glycoprotein drug efflux pump (*MDR1*) and the multidrug resistance associated protein 1 (*MRP1*) can lead to multi-drug resistance [15, 66, 67]. Furthermore, HIF induced autophagy was found to drive resistance to radiotherapy and the chemotherapeutic agent N-(4-Hydroxyphenyl)retinamide [68, 69].

The prognostic value of hypoxia can also be explained through its interaction with the hallmarks of cancer. The hallmarks required for the development and maintenance of cancer were defined in a landmark paper by Hanahan and Weinberg in 2000 [70]. These are:

1. Sustaining proliferative signalling
2. Evading growth suppressors
3. Resisting cell death
4. Enabling replicative immortality
5. Inducing angiogenesis
6. Activating invasion and metastasis

Their updated reviews published in 2011 and 2022 list eight further enabling hallmarks and emerging characteristics [71, 72]:

1. Deregulating cellular energetics
2. Avoiding immune destruction
3. Genome instability and mutation
4. Tumour promoting inflammation
5. Unlocking phenotypic plasticity
6. Nonmutational epigenetic reprogramming

7. Polymorphic microbiomes
8. Senescent cells

The influence of hypoxia on selected key hallmarks is discussed below.

1.5.1 Resisting cell death

Cancer cells are exposed to extrinsic stress from the tumour microenvironment and intrinsic stress from factors such as DNA damage and imbalanced oncogenic signalling. These stressors drive apoptotic signalling which must be resisted by cancer cells if tumour formation is to occur [71]. Hypoxia can have both a positive and negative effect on apoptotic signalling depending on the cell type, intensity of hypoxia, and presence of other stressors [73, 74]. For example, known HIF target B-cell lymphoma 2 interacting protein 3 (BNIP3) is a pro-apoptotic protein which can stimulate cell death under hypoxia [73]. However, hypoxia is also able to increase expression of anti-apoptotic proteins such as inhibitor-of-apoptosis protein 2 (IAP-2) [75]. Additionally, alternate splicing driven by hypoxia increases the amount of the fully spliced anti-apoptotic Bcl-xL protein compared to the alternatively spliced pro-apoptotic Bcl-xS [38]. Finally, hypoxic breast tumours have been shown to have higher rates of *TP53* mutation, suggesting an association between perturbation of apoptosis and hypoxia [9, 76].

1.5.2 Inducing angiogenesis

Angiogenesis is the process by which new blood vessels sprout from the existing vasculature. The primary mediator of angiogenesis is VEGF, which binds to VEGF receptors (VEGFR) to trigger a pro-angiogenic signalling cascade [71]. VEGF is a direct transcriptional target of HIF [14], and VEGFR1 is upregulated under hypoxic conditions [77], linking hypoxic signalling to the stimulation of angiogenesis. In addition, the expression of matrix metalloproteinases is upregulated by HIF signalling [8, 10]. These enzymes degrade the extracellular matrix (ECM), allowing endothelial cells to migrate and sprout from the blood vessels. Together, these factors enable tumours to grow their own vasculature and increase their access to oxygen and nutrients. However, the vessels produced during tumourigenesis are often poorly matured, resulting in leakiness, and poorly organised with excessive convolution and branching [8, 71, 78]. This leaves areas of the tumour without access to a regular oxygen supply, contributing to cycling hypoxia, and results in further hypoxia driven pro-angiogenic signalling.

1.5.3 Invasion and metastasis

The activation of invasion and metastasis is another key hallmark of cancer [71]. Invasion and metastasis describes a process whereby cancer cells migrate into surrounding tissue, intravasate to enter the circulatory system, then extravasate to colonise distant tissues forming micro- and macro-metastases [71]. The expression of matrix metalloproteinases by hypoxic cells degrades the ECM and basement membrane, allowing cells to invade and metastasise [79]. Furthermore, the poorly developed vasculature formed during tumour angiogenesis enables cells to intravasate more easily [14]. However, the main way that hypoxia promotes invasion and metastasis is through driving EMT [8].

EMT is the process by which epithelial cells progressively lose their epithelial characteristics (e.g. apical-basal polarity, strong cell-cell adhesions, attachment to the basement membrane) and acquire mesenchymal characteristics (e.g. migration ability, fibroblast-like morphology, front-back polarity, weakened cell-cell adhesion) [80]. This process is dynamic and reversible (termed mesenchymal-epithelial transition, or MET) meaning it can be transient or stable, and complete or partial [71]. This process is key in embryonic development and normal processes such as wound healing, but it can also be co-opted by cancer cells to break away from the tumour mass, invade surrounding tissue, and spread to form distant metastases [80, 81].

EMT involves vast changes in gene expression (approximately 400 genes) [82]. These changes are largely orchestrated by a few master transcription factors (EMT-TFs). These are: Snail (*SNAI1*), Slug (*SNAI2*), Twist1 (*TWIST1*), Twist2 (*TWIST2*), Zeb1 (*ZEB1*), and Zeb2 (*ZEB2*) [71, 81]. Several of these transcription factors directly repress E-cadherin, a key marker of epithelial cells, while activating expression of mesenchymal markers such as N-cadherin and vimentin [81]. There is evidence that hypoxia is able to drive activation of some of these transcription factors, thus driving EMT in the hypoxic tumour microenvironment [82]. For example, *SNAI1*, *TWIST1*, and *ZEB1* contain HREs in their promoter regions so are implicated as direct HIF targets [82], and experimentally, *SNAI1* was shown to be transcribed by HIF1 [83]. The expression of other EMT signalling molecules such as TGFβ and NFκB is also stimulated by hypoxia [82]. Finally, hypoxic regulation of miRNAs can also affect EMT. For example, known HIF target miR-210 directly suppresses E-cadherin expression through the canonical miRNA pathway [84].

The promotion of EMT by hypoxia also enables cells to resist cell death as mesenchymal cells are more resistant to apoptotic stimuli than their epithelial counterparts [71, 85, 86]. The mechanisms by which this occurs are not well understood [87], but studies have suggested that regulation of BCL-2 family proteins by the EMT-TFs plays a role. One study showed that the pro-apoptotic protein BIM (Bcl-2 interacting mediator of cell death) was directly repressed by Zeb1 in lung cancer cells [87], while decreased Twist expression led to a decrease in the anti-apoptotic protein BCL-2 [88]. In accordance, EMT has been associated with chemotherapy resistance in a several cancer types [89].

1.5.4 Genome instability

The emerging hallmark genome instability refers to the increased likelihood of genetic alterations occurring in a cell [71]. This enables tumour progression by increasing the rate at which possibly advantageous mutations can accumulate in the tumour. This hallmark has been associated with hypoxia both *in vitro* [37, 90] and *in vivo* [76, 91]. *In vitro* evidence showed downregulation of DNA repair genes occurred under chronic hypoxia (20 passages) in association with increased frame shift insertions and copy number variations [37]. This downregulation is in part mediated by HIF induced miRNA expression. The known HIF target miR-210 suppresses expression of the homology-directed repair protein RAD52 thus attenuating DNA damage repair [92]. This is supported by *in vivo* evidence whereby DNA mismatch repair genes such as *RAD51*, *MLH1*, and *MSH1* are expressed at lower levels in hypoxic tumours [76]. Additionally, Bhandari et al. [76] showed a significant correlation between copy number aberrations and the tumour hypoxia score *in vivo*, with this correlation being highest in breast cancer.

1.6 Aims

TNBC has the worst prognosis of the breast cancer subtypes and lacks specialised treatments [3, 4]. As indicated by analysis of hypoxia regulated gene expression, it is also the most hypoxic of the breast cancer subtypes (Figure 6). This means TNBC is a good model to use both in studying adaptation to chronic hypoxia and in identifying specific vulnerabilities which exist in hypoxic cells. These vulnerabilities represent “Achilles heels” in hypoxic cells which can be targeted therapeutically. As tumours are generally more hypoxic than their tissue of origin, they will be more sensitive to these treatments enabling a good therapeutic index to be established.

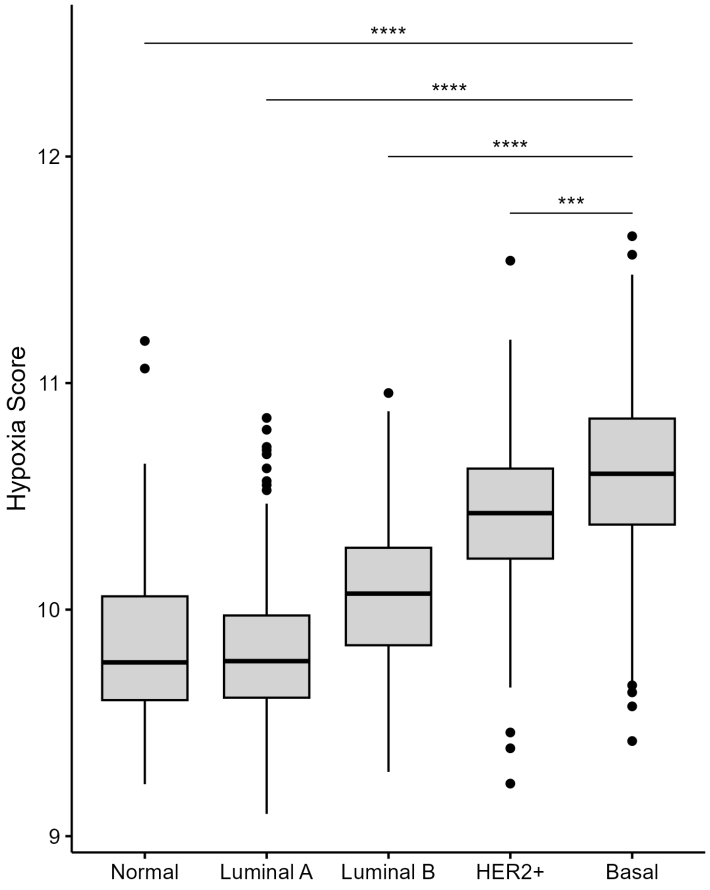

Figure 6: Hypoxia score by breast cancer subtype. Hypoxia scores of normal breast and breast cancer samples from The Cancer Genome Atlas calculated using SigQC and the Buffa hypoxia gene signature as described in Section 4.2.9. The basal subtype (predominantly TNBC) has a significantly higher hypoxia score than all other breast cancer subtypes. Significance was tested using a Wilcoxon rank sum test and multiple test correction was carried out using Holm’s method (**** = adjusted $p \leqslant 0.0001$).

Existing treatments which have aimed to target hypoxia have had limited success [93, 94]. This includes hypoxia-activated prodrugs (drugs which undergo a chemical reaction to become active specifically under hypoxic conditions [95]) and methods which aim to increase tumour oxygenation such as hyperbaric oxygen breathing or oxidative phosphorylation inhibitors. Some drugs which modulate HIF activity have been developed but have produced mixed results in clinical trials [93]. This highlights a need for further research into hypoxia specific cell dependencies in the context of cancer.

With that in mind, this research aimed to:

1. Characterise the transcriptional response to chronic hypoxia
2. Identify genes which are selectively essential under hypoxia
3. Identify miRNAs which are selectively essential under hypoxia

2 Methods

Most of the methods used in this project apply to only a single chapter and are therefore included within the relevant chapter's methods section. Only methods that were used across multiple chapters are included here. Suppliers and catalogue numbers for all consumables, antibodies, and equipment are found in Tables S1 to S3. Details for all computational programs used are found in Table S4.

2.1 Cell culture

MCF7, MDA-MB-231, HCC1806, and MDA-MB-468 breast cancer cell lines were used throughout this project and were authenticated by NorthGene in 2019. All cells were cultured in a humidified environment at 37℃ with 5% CO_2, in DMEM (5 mM glucose) supplemented with 10% foetal bovine serum (FBS) and 1% penicillin/streptomycin (P/S) (final concentration: 100 units/mL penicillin and 100 µg streptomycin). Cells were passaged at around 80% confluency. To passage, cells were rinsed using phosphate buffered saline (PBS) and detached from the surface of the flask using trypsin. Culture media was used to deactivate the trypsin and the cell solution was centrifuged at 275 *g* for 4 minutes to pellet the cells. The cell pellet was resuspended in culture media. If cells were being seeded for an experiment, the cell number was calculated using the LUNA II automated cell counter. The average of two readings was taken for each cell solution. Cells were regularly tested for mycoplasma contamination using the MycoAlert Mycoplasma Detection Assay.

2.2 Hypoxic cell culture

For hypoxic culture, cells were transferred to the Invivo$_2$ 400 workstation (hereon referred to as the hypoxia chamber) set to 1% O_2, 5% CO_2 with 50% humidity. All cell handling took place within the hypoxia chamber using culture media, PBS and trypsin that had been pre-conditioned to prevent re-oxygenation events. Unless otherwise specified, media was changed at least every 48 hours to prevent nutrient depletion.

2.3 Western blots

The conditions used for western blots varied depending on the protein of interest. All methods used for western blots are described below, and the combination of methods used for each protein are described in the relevant methods section.

2.3.1 Collection of protein lysates in urea buffer

Protein lysates were collected from cells cultured in 6 well plates using 200 µL of urea lysis buffer. The lysis buffer was made in house using 5 mL glycerol, 0.5 mL 1 M Tris-HCL (pH 7.4), 250 µL 1 M DTT (dithiothreitol), 2.5 mL 20% SDS (sodium dodecyl sulphate), one cOmplete EDTA-free protease inhibitor tablet, and made up to 50 mL using 8 M urea. The protein lysates were pulled through a 19G needle using a 1 mL syringe to shear the sample. Lysates were stored at -20℃.

2.3.2 Collection of protein lysates in RIPA buffer

Protein lysates were collected from cells cultured in 6 well plates using 200 µL of RIPA buffer. One cOmplete EDTA-free protease inhibitor tablet and five Pierce Phosphatase Inhibitor Mini Tablets were dissolved per 50 mL of RIPA buffer. Protein lysates were stored at -20℃.

2.3.3 Preparation of samples

The concentration of protein in the lysates was quantified using the colorimetric DC Protein Assay according to the manufacturer's instructions. A 1:4 dilution of the protein sample was used in the assay. A standard curve was produced each time the assay was run using BSA standards from the Quick Start Bovine Serum Albumin Standard Set. After quantification, a 20 µL solution was prepared for each sample which contained 5 µL of 4X Laemmli buffer (supplemented with 1:20 β-mercaptoethanol) and the desired mass of protein. The volume was made up to 20 µL using the appropriate lysis buffer. Samples were incubated at 95℃ for 10 minutes to denature the proteins then placed on ice prior to electrophoresis.

2.3.4 Running the gel

15 µL of protein ladder was loaded into the first well of the chosen gel type (NuPAGE 4-12% Bis-Tris Gel or NuPAGE 3-8% Tris-Acetate Gel), and 20 µL of protein sample was loaded into subsequent wells. Gels were loaded into an Invitrogen XCell SureLock gel electrophoresis tank and the tank filled with 1X running buffer (NuPAGE MES SDS Running Buffer or NuPAGE MOPS SDS Running Buffer). Gels were run for 10 minutes at 80 V, then at 120 V until complete. Gels were removed from the tank and prepared for transfer.

2.3.5 Transfer preparation

An Immobilon-P PVDF (polyvinylidene difluoride) membrane was activated by briefly soaking in 100% methanol, then washed in water and soaked in the chosen preparation of NuPAGE Transfer Buffer before use. One of the three protocols below was then followed for transfer.

2.3.6 Wet transfer

Proteins were transferred to an Immobilon-P PVDF membrane using the Mini Trans-Blot Cell tank system. The NuPAGE Transfer Buffer is comprised as follows: 1 part methanol, 0.5 parts NuPAGE Transfer Buffer (20X), 8.5 parts water. The gel and membrane were assembled between western blotting thick filter paper and foam pads. The proteins were transferred at 50 V for 2 hours on ice.

2.3.7 Novex semi-dry transfer

Proteins were transferred to an Immobilon-P PVDF membrane using the Novex Semi-Dry blotting system. The NuPAGE Transfer Buffer is comprised as follows: 1 part methanol, 1 part NuPAGE Transfer Buffer (20X), 8 parts water. The gel and membrane were assembled between Immobilon Blotting Filter Paper. The proteins were transferred for 90 minutes at 20 V.

2.3.8 Turbo transfer

Proteins were transferred to an Immobilon-P PVDF membrane using the Trans-Blot Turbo Transfer System. The NuPAGE Transfer Buffer is comprised as follows: 1 part methanol, 1 part NuPAGE Transfer Buffer (20X), 8 parts water. The gel and membrane were assembled

between Immobilon Blotting Filter Paper. The proteins were transferred for 30 minutes at 25 V.

2.3.9 Blocking, probing, washing and developing

Following transfer, the membranes were incubated on a rocking platform for at least 30 minutes in a blocking buffer at room temperature. Blocking buffer was made using a base of tris buffered saline containing 0.1% Tween 20 (TBS-T) with either 5% nonfat dried milk powder powder or 5% bovine serum albumin (BSA). The primary antibody of choice was diluted in blocking buffer and incubated with the membrane overnight on a rocking platform at 4℃. The membrane was washed three times using TBS-T for 10 minutes. The secondary antibody was diluted in blocking buffer and incubated for at least one hour on a rocking platform at room temperature. The membrane was again washed three times in TBS-T for 10 minutes. The protein was visualised using Amersham ECL prime western blotting detection reagent and the Invitrogen iBright1500 according to manufacturer's instructions.

2.3.10 β-actin

β-actin was used as a loading control for all western blots. The mouse monoclonal HRP-conjugated antibody to β-actin was diluted 1:50000 in 5% nonfat dried milk powder blocking buffer and incubated with the membrane on a rocking platform for at least one hour. The membrane was washed three times using TBS-T for 10 minutes. The protein was visualised using Amersham ECL prime western blotting detection reagent and the Invitrogen iBright1500 according to manufacturer's instructions.

3 Optimisation of a bioinformatics pipeline to quantify micro-RNAs in small RNA sequencing data

3.1 Introduction

The sequencing of small RNAs presents unique problems in terms of processing and alignment. The nature of short sequences means they are more likely to map to multiple locations in the genome (multi-map). Additional problems arise with miRNAs, as approximately 76% of genes encoding them are contained within other genes or regulatory elements (Figure 7) [96, 97]. This means that when a miRNA is mapped to the genome, even in a single location, it may map to a larger gene as well as the miRNA gene. These factors mean that the alignment parameters used for standard polyA RNA-seq data are unsuitable. This is because the parameters used often exclude multi-mapping reads and reads which align to overlapping genes. The amount of information lost when using these exclusion parameters on small RNA-seq data means alternative methods must be used.

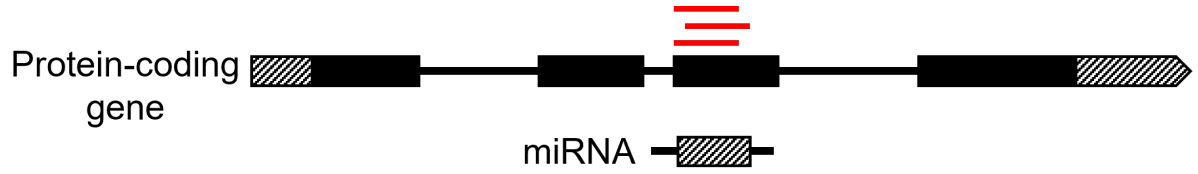

Figure 7: miRNA genes are often embedded within protein-coding or non-coding genes.

Some pipelines already exist to handle small RNA-seq data. However, many of these tools are web-based [98, 99], limiting both reproducibility and the ability to automate data processing. Some packages use alignment tools which have been shown to have sub-optimum performance in miRNA datasets [100, 101]. Other pipelines effectively align miRNAs but do not tackle the issue of multi-mapping reads [102, 103].

Here, I systematically identified the best alignment strategy for processing small RNA-seq data with the specific aim of quantifying miRNA expression. I relied on existing research regarding the best alignment tools and strategies for dealing with multi-mapping reads to inform my decisions [101, 104, 105].

3.2 Methods

Small RNA-seq data of samples SRR873382, SRR873383, SRR873388, and SRR873389 from dataset GSE47602 were downloaded from the Gene Expression Omnibus and processed into FASTQ files using the SRA-Toolkit (version 2.9). These samples contain small RNA-seq data from MCF7 cells exposed to hypoxia [106]. FastQC (version 0.11.9) was used to report on read quality. Cutadapt (version 2.10) was used to trim adapters and run quality control with the following options: `-q 20 -m 16 -M 28 --discard-untrimmed`. Bowtie (version 1.2.3) (options: `-a -m 20 -v 1 --best --strata`) and Bowtie2 (version 2.3.5.1) (options: `--very-sensitive-local`) were used for alignment as described below. The featureCounts function within the Subread package (version 1.6.4) was used to count alignments with the options described in Sections 3.3.3 and 3.3.4. Manatee was used to manage multi-mapping reads with the following options: `-collapse yes -mismatches 1 -m 20`. The human genome assembly Hsa.GRCh38 version 104 and miRBase release 22.1 were used for alignment/counting. The miRBase files were modified before use (Section 3.2.1). Testing of differential gene expression was conducted using DESeq2 (version 1.38.3).

3.2.1 Modification of miRBase files

The miRBase FASTA file contained the hairpin RNA sequences. This meant that uracil ("U") was present instead of thymine ("T"). The sequences were converted to the DNA format by replacing all "U" bases with "T" bases. Any miRNA sequences not designated as human were also removed. The GFF3 file was modified by removal of the primary transcripts (i.e. only mature miRNA sequences were retained). This allowed quantification of only mature, and therefore functional, miRNAs and prevented instances of multi-mapping caused by reads aligning to both their primary and mature transcript. When using Manatee, the GFF3 file was further modified to adhere to program requirements. For each line, "ID" was changed to "gene_id", "Name" was changed to "gene_name", and "gene_biotype=miRNA" was appended.

3.3 Results

Processing of RNA-seq data involves three main steps: quality control, alignment, and quantification (Figure 8).

Figure 8: miRNA alignment. Schematic showing the three main steps involved in processing RNA sequencing data. FASTQ files are fed into the pipeline and a matrix of gene counts is produced.

The first processing step involves removing the adapter sequences from the reads. Cutadapt was used to trim the adapter sequences and filter out reads shorter than 16 bases or longer than 28 bases after trimming. This encompasses the length of all miRNAs annotated within miRBase (Figure 9).

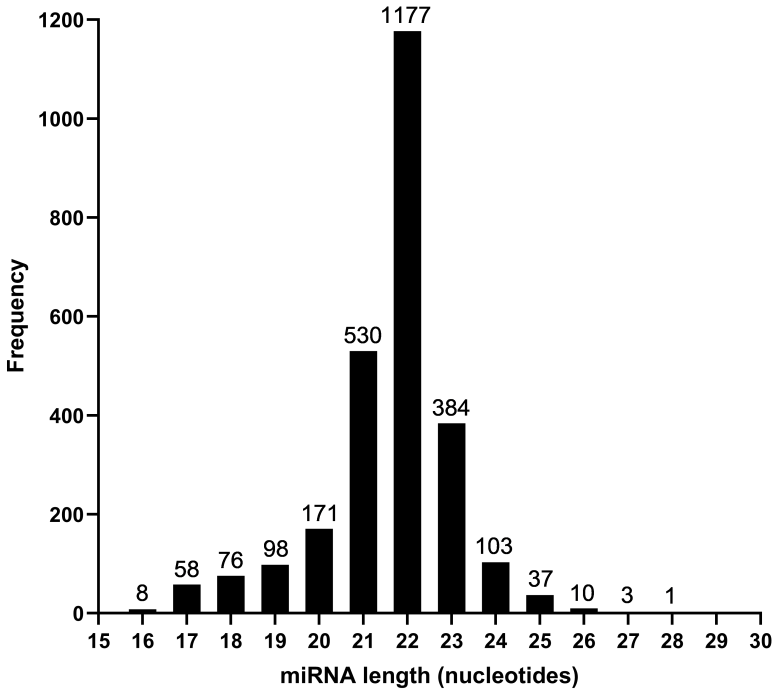

Figure 9: Length of miRNAs in miRBase. The length of miRNAs annotated in miRbase (release 22.1). The majority of miRNAs are 21-23 nucleotides in length, but there are instances of miRNAs as short as 16 and as long as 28 nucleotides.

3.3.1 Bowtie2

Bowtie2 was reported to be the best alignment tool for miRNAs so initially I used this tool to align reads to the genome [101]. Despite overall good alignment rates (average 69% uniquely mapped reads across the four samples), Bowtie2 is not able to align reads $<$ 22 nucleotides using its scoring algorithm in the default settings (Figure 10). This is apparent when looking at the scoring algorithm used, despite the fact the publication by Ziemann et al. [101] reports alignment of short reads by Bowtie2.

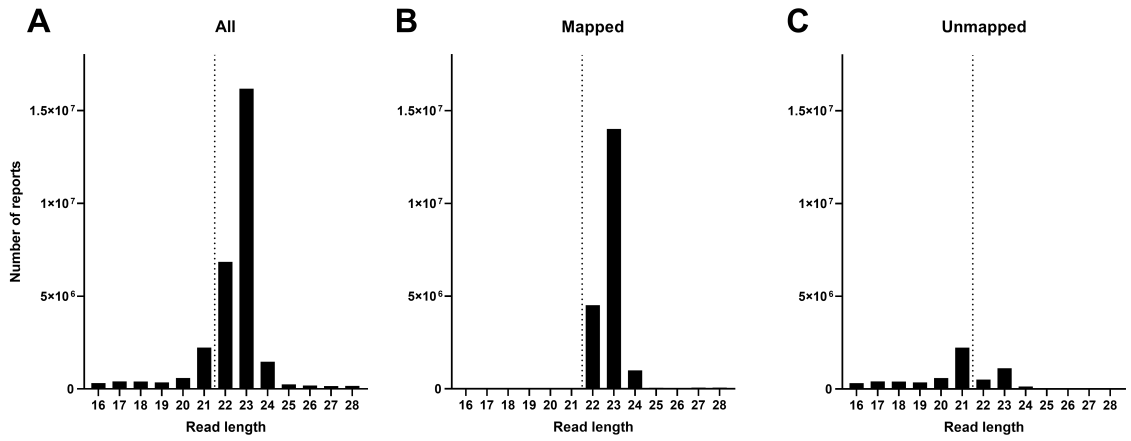

Figure 10: Bowtie2 mapping statistics. The data shown is for sample SRR873382 but is representative of all four samples. **(A)** The length distribution of all reads after trimming of the adapters and filtering out reads longer than 28 bases and shorter than 16 nucleotides. The majority of reads are 21-24 nucleotides in length. **(B)** The length distribution of mapped reads. Although the mapping rate is good in reads $\geqslant$ 22 nucleotides, reads $<$ 22 nucleotides do not map at all. Reads are mapped against the whole genome. **(C)** The length distribution of unmapped reads. All reads $<$ 22 nucleotides fail to map.

This is a problem in miRNA sequencing as 35% (941/2656) of miRNAs listed in miRBase fall below this threshold and would not be quantified using this method. According to the Bowtie2 manual, "In local alignment mode, the default minimum score threshold is $20 + 8.0 \times \ln(L)$, where L is the read length." In local alignment mode, points are allocated for each base in the read that aligns to the reference, and points are deducted for mismatches or indels. The default "match bonus" (i.e. the number of points added for each nucleotide that aligns) is 2. This means that a perfectly aligning read of 20 nucleotides can achieve a maximum score of

40. However, the minimum score threshold for a 20 nucleotide read is 44:

$$\begin{aligned} Threshold &= 20 + 8 \times \ln(20) \\ Threshold &= 20 + 8 \times 2.996 \\ Threshold &= 20 + 23.968 \\ Threshold &= 43.968 \end{aligned}$$

3.3.2 Alignment strategy

Due to the issues with Bowtie2, I chose instead to use Bowtie to align reads. This program also scored highly in the publication by Ziemann et al. [101] and was found to perform better than Bowtie2 by Potla et al. [102]. Bowtie is able to successfully align reads as short as 16 nucleotides making it more appropriate for the alignment of small RNAs (Figure 11).

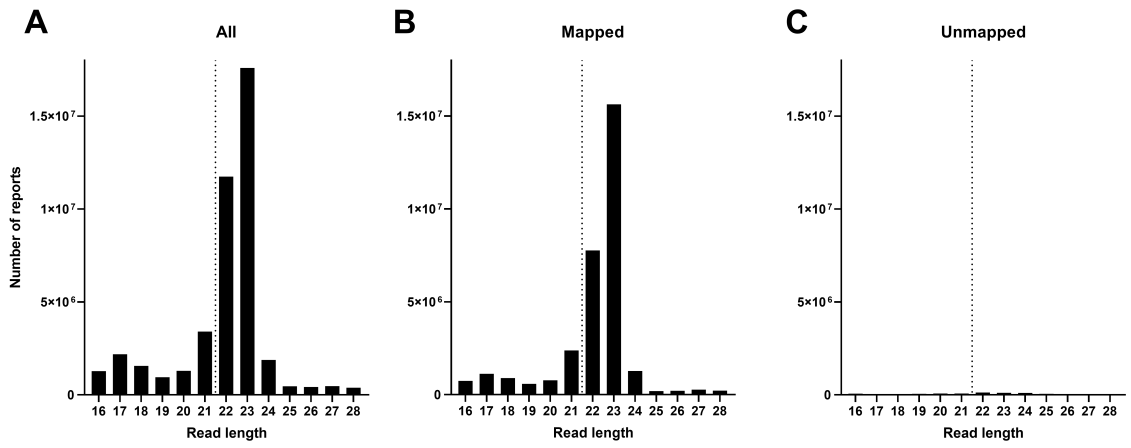

Figure 11: Bowtie mapping statistics. The data shown is for sample SRR873382 but is representative of all four samples. **(A)** The length distribution of all reads after trimming of the adapters and filtering out reads longer than 28 bases and shorter than 16 nucleotides. The majority of reads are 21-24 nucleotides in length. **(B)** The length distribution of mapped reads. The mapping rate is consistent across reads of all lengths. Reads are mapped against the whole genome. **(C)** The length distribution of unmapped reads. Very few reads fail to map and these are distributed across read lengths.

The second question is whether to align reads against the whole genome, or whether to align them specifically against a database of miRNAs (miRBase). Alignment of reads against miRBase reduces the amount of multi-mapping and ambiguity in assigning reads, but also limits the information which can be gained (i.e. not all small RNAs are miRNAs, and these

will fail to align). It could also create false positives as other locations in the genome where a read may align are not considered. Aligning samples from this dataset against either the whole genome or miRBase did not affect the number of uniquely mapped reads (Figure 12).

As expected, the number of reads which fail to align is greater when aligning against miRBase. Multi-mapping reads make up 21% of the dataset when aligning against the whole genome compared to 15% when aligning against miRBase. Since there are more potential mapping locations in the whole genome reference, this is expected. Because aligning against the whole genome allows the most information to be captured without compromising the number of uniquely mapped reads, this method was chosen going forward.

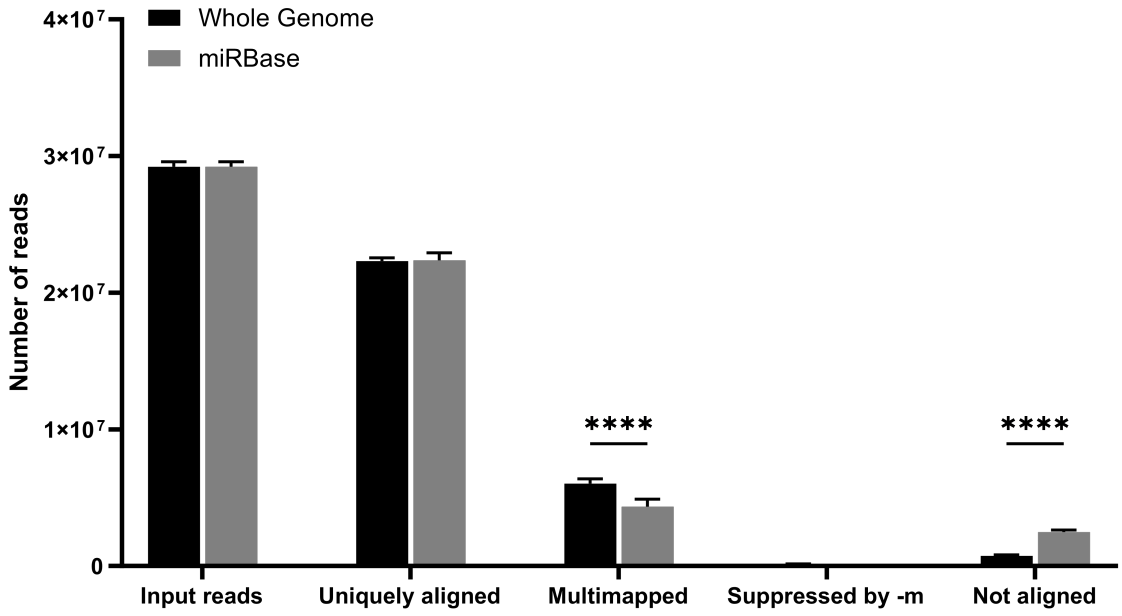

Figure 12: Choice of database to align to. Comparison of mapping when aligning short reads against miRBase or the whole genome. The number of multi-mapped reads decreases when aligning against miRBase, but the number of reads which fail to align increases. Reads that are "Suppressed by –m" have aligned $>$ 20 times to the database of choice. This figure is higher when aligning against the whole genome but remains a very small percentage of the total reads. Significance was calculated using two-way ANOVA and corrected for multiple testing using Sidak's method *(n = 4)* (**** $p \leqslant 0.0001$). Error bars show standard deviation.

3.3.3 Quantification strategy

Once aligned, the reads aligning to a particular feature (e.g. a gene) must be quantified. Different factors are taken into account when choosing whether a read is counted towards a specific feature. If a read maps to multiple locations in the genome, it may be omitted from the total. Similarly, if a gene aligns to a location which contains two genomic features (e.g. a miRNA which is embedded within a protein-coding gene) it may be discarded due to ambiguity. In this instance, the read could equally represent either the miRNA or be a fragment of mRNA from the protein coding gene. As such, these reads are typically assigned to neither feature.

As with alignment, this quantification can be conducted using either the whole genome or miRBase as a reference. The number of assigned reads is significantly higher when aligning to miRBase compared to the whole genome (Figure 13). The majority of reads are lost due to "ambiguity" when aligning against the whole genome. Since miRNA genes are often contained within other genes, featureCounts is unable to determine which gene the read belongs to and does not count reads which align in these areas. This is demonstrated by the recovery of these reads when using the `-O` option when running featureCounts. Using the `-O` option assigns the read to both genes with which it overlaps (i.e. the read is counted towards both the miRNA and the gene it overlaps with). This distorts the total read count so should be avoided where possible.

From these results it was determined that the optimum process involves using the whole genome as a reference to align against as this prevents loss of information and allows other types of small RNA to be quantified if desired. However, reads should subsequently be counted against miRBase as this removes the complicating issue of overlapping genes and improves miRNA quantification.

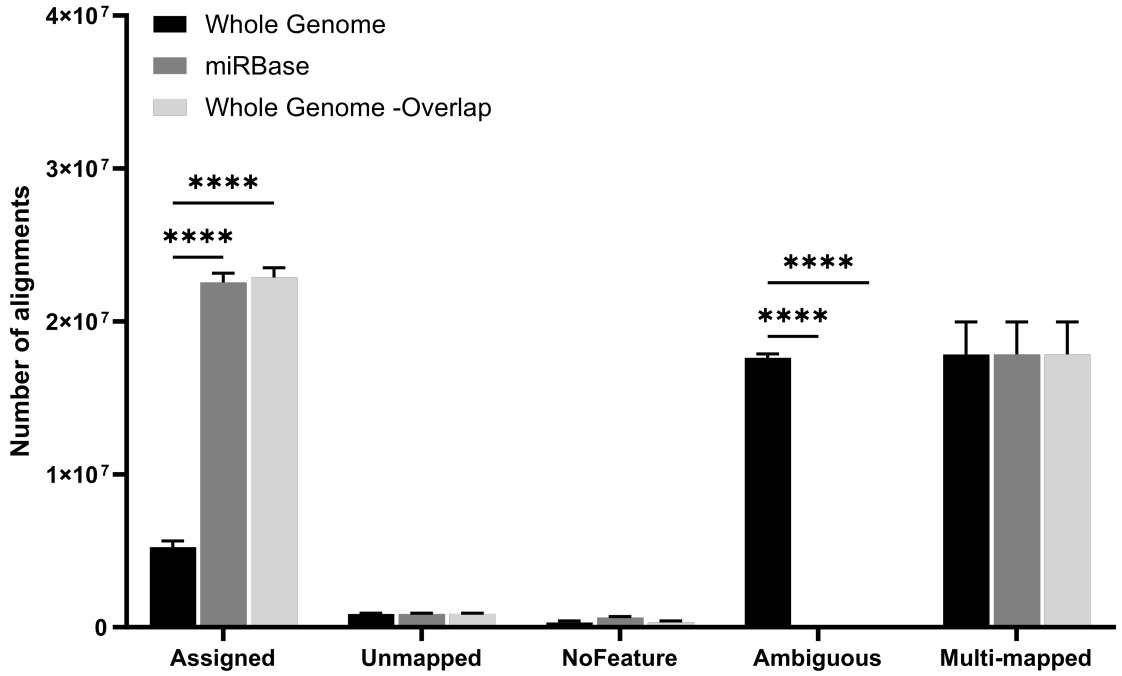

Figure 13: Quantification of miRNAs using the whole genome or miRBase as a reference. Reads are aligned to the whole genome and quantified against either the whole genome or miRBase. The number of assigned reads is significantly higher when aligning to miRBase compared to the whole genome. This number is recovered when using the -O option in featureCounts, as this allows reads to be counted when they align to a section of the genome that contains two or more genes. There are slightly more reads which do not map to a feature (i.e. gene) when counting against miRBase, but this is not significant. Significance was calculated using two-way ANOVA and corrected for multiple testing using Tukey's method *(n = 4)* (**** $p \leqslant 0.0001$). Error bars show standard deviation.

3.3.4 Multi-mapping reads

Bowtie does not label alignments that are multi-mapped in a way that can be used by featureCounts. A custom python script (Section 9.2) was written to add information about multi-mapping to each alignment in the Bowtie output. This then allowed two counting strategies which consider multi-mapping reads to be considered. Using featureCounts, multi-mapped reads can be excluded entirely, as it the norm in exome sequencing. Multi-mapped reads may be counted and their full count value assigned to each location where the read maps. Alternatively, a fraction approach can be used to divide the number of counts between the mapped locations. The final strategy is optimal in many ways, as it allows all possible locations to be considered without distorting total read counts. However, power is lost by the division of

reads and the unique method chosen here of aligning reads to the whole genome, but counting against miRBase, means this method is not appropriate. The reads would be divided between the mapped locations regardless of whether they map to a miRNA or not. Counting all multi-mapped reads ensures no information is lost and recovers many previously uncounted reads. However, this distorts the read number by inflating the number of counts. The difference in the number of assigned reads when counting or excluding those which multi-map is demonstrated in Figure 14.

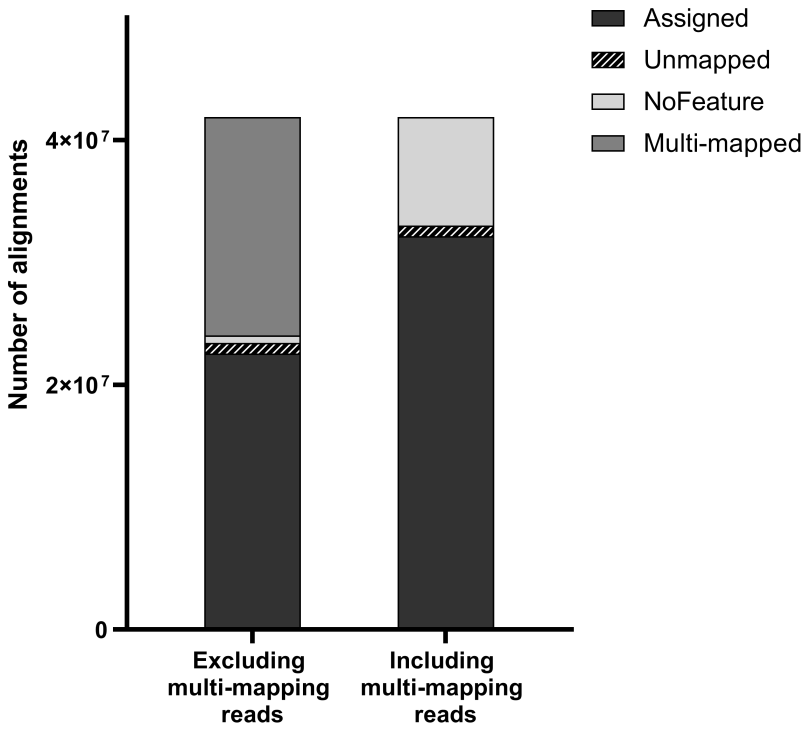

Figure 14: The effect of including or excluding multi-mapping reads. The number of alignments counted when multi-mapped reads are included or excluded. Reads were aligned against the whole genome then counted against miRBase. Multi-mapped reads are counted using the -M function and the reads assigned in their entirety. Approximately half of the multi-mapped reads map to no feature while the other half become successful alignments to a miRNA gene.

Other researchers have applied strategies to rescue these multi-mapped reads in a more systematic manner [104]. Read rescue is a strategy which counts multi-mapped reads based on whether there are also other uniquely aligned reads in the same location and distributes the multi-mapped reads according to this. Manatee is one such program [105]. It uses Bowtie to align the reads, then uses a read rescue counting strategy. The number of assignments

increases when using Manatee compared to excluding multi-mapped reads altogether, but is less than if all multi-mapped reads are counted (Figure 15).

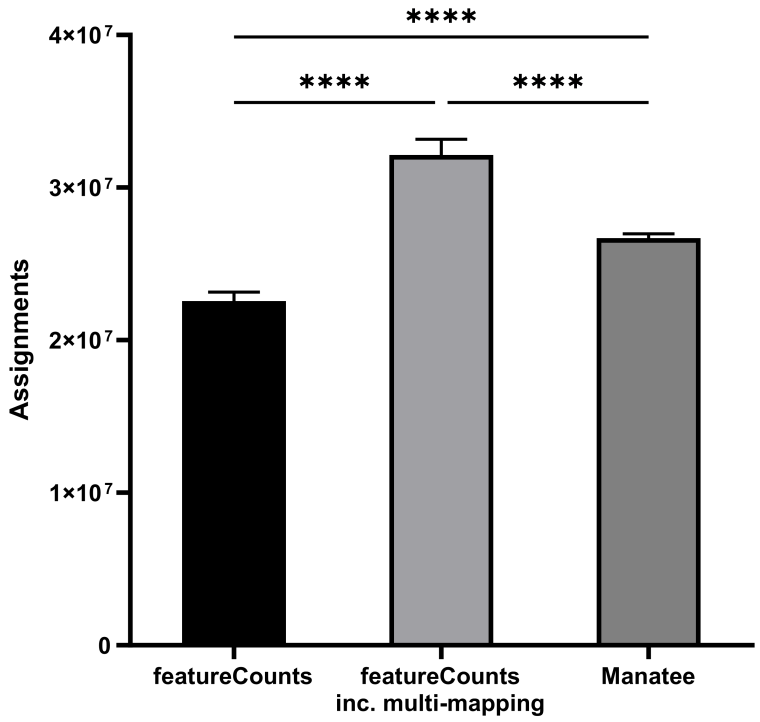

Figure 15: Assignments using featureCounts and Manatee. The number of reads assigned to miRNA genes made when using featureCounts with the default settings, featureCounts in multi-mapping mode, and Manatee for quantification. Significance was calculated using an ordinary one-way ANOVA and corrected for multiple testing using Tukey's method *(n = 4)* (**** $p \leqslant 0.0001$). Error bars show standard deviation.

This offers rescue of those reads where we can have some level of certainty regarding their mapping, but excludes those which remain ambiguous. When using these counts to run differential expression analysis, the differentially expressed miRNAs are largely overlapping between the three strategies (Figure 16).

The number of significantly differentially expressed miRNAs is lower when excluding multi-mapping reads (198), which is unsurprising given the decreased number of assigned reads. Manatee does not produce as many significantly differentially expressed miRNAs as featureCounts in multi-mapping mode (226 vs 241). Again, this is expected as the counts included in the analysis are lower.

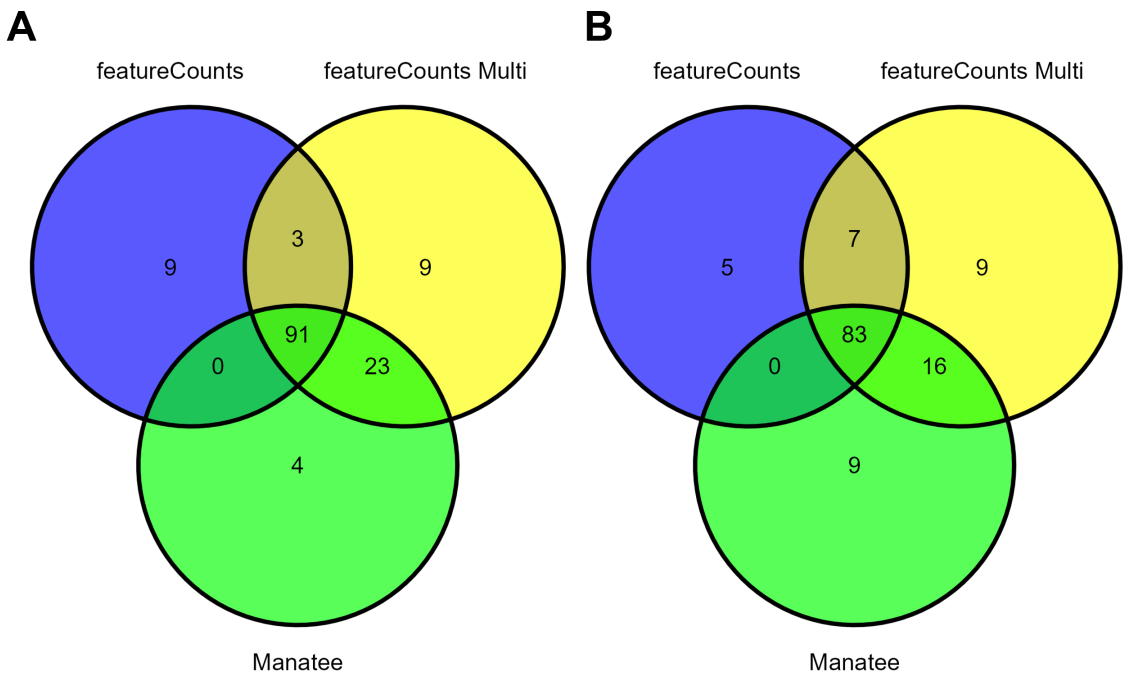

Figure 16: Differentially expressed genes using featureCounts and Manatee. Differentially expressed (adjusted p value $\leqslant 0.05$) miRNAs between normoxia and hypoxia when using the results from featureCounts, featureCounts in multi-mapping mode, and Manatee for quantification. The results are largely overlapping between the different methods. The methods that produce more assigned reads also identify more differentially expressed miRNAs. **(A)** Significantly upregulated miRNAs. **(B)** Significantly downregulated miRNAs.

3.4 Discussion

The results of these analyses suggest that an optimum processing strategy for miRNA sequencing data uses Bowtie to align reads to the whole genome, in agreement with other researchers [102, 103]. The optimum strategy also quantifies reads against miRBase, in agreement with Potla et al. [102], but imposes an additional read rescue strategy to salvage multi-mapped reads. As Manatee is capable of conducting both of these steps, it will be used in all analyses going forward.

4 Transcriptional adaptation to chronic hypoxia

4.1 Introduction

Although the hypoxic response is relatively well studied [107, 108], there are still areas where data are lacking. For example, the effect of chronic hypoxia has not been extensively investigated [12]. Some studies refer to their hypoxia timepoints as "chronic", but their exposure times are relatively short ($\leqslant$ 24 hours) [13, 21]. At these timepoints, substantial levels of the HIF proteins remain and true adaptation has not had time to occur. HIF1 protein levels are reported to drop after around 12 hours of hypoxia in endothelial cells [24, 31–33]. HIF2 protein levels peak at a later time point [24, 32, 33], but typically both HIF subtypes are present at low levels after 48 hours [24, 33]. This means the response to chronic hypoxia may be different from the response to acute hypoxia. There is also evidence that chronic hypoxia can promote a more aggressive cancer phenotype through selective pressure [8, 22, 109]. A systematic search of the Gene Expression Omnibus database (Di Giovannantonio, M. 2020, unpublished) revealed only four transcriptomic datasets where cells had been cultured under continuous hypoxia for $\geqslant$ 7 days, of which three had associated publications [110–112].

One of these studies used a clear cell renal cell carcinoma cell line with VHL mutations, meaning the response to hypoxia is not representative of a wild type VHL cell [110]. Nonetheless, they demonstrated little overlap in the mRNA and miRNA profiles taken after exposure to acute (24 hours) and chronic (3 months) hypoxia (1% O_2) [110]. The second study (Danielsson et al. [112]) did not compare acute and chronic hypoxia, but did show that chronic hypoxia (3% O_2 for 6 passages) caused different transcriptional responses in an isogenic fibroblast cell line at different stages of transformation, indicating a variable hypoxic response even in cells with an almost identical genetic background. The final study, which used the MCF7 breast cancer cell line, identified differentially expressed genes between the acute (24 hours) and chronic (10 weeks) hypoxia (0.5% O_2) timepoints [111]. They also found that overexpressing HER2 increased expression of hypoxia response genes, again demonstrating that the hypoxic response is context dependent.

Since then, a fourth study into the transcriptional response to chronic hypoxia has been published [37]. Khouzam et al. [37] exposed six cell lines, including MCF7 cells, to hypoxia

(1% O_2) for 20 passages. They did not do a comparison with acute hypoxia, but did show that some cell lines have a propensity to generate a stronger response to hypoxia (i.e. there were more differentially expressed genes). Although they considered using a TNBC cell line in their experiments (MDA-MB-231), this was excluded before microarray analysis.

Overall, research into chronic hypoxia, particularly in the context of TNBC, is lacking. This chapter aims to characterise the response to chronic hypoxia in four breast cancer cell lines. Of the cell lines chosen, one is oestrogen receptor positive (MCF7). The remaining three are triple-negative and cover a range of TNBC subtypes (i.e. MDA-MB-231 cells are mesenchymal stem like, MDA-MB-468 cells are basal like 1, and HCC1806 are basal like 2 [113, 114]).

4.2 Methods

4.2.1 Collection of samples (MDA-MB-231)

These samples were collected by Dr Anjali Arora (Buffa lab). Cells were seeded and left to adhere overnight before relevant samples were transferred to the hypoxia chamber. Cells were maintained in either normoxia or hypoxia as per Section 2.1 and Section 2.2. Cell pellets were collected after 6 hours, 24 hours, 7 days, or 14 days of culture in both conditions and stored at -80℃.

4.2.2 Collection of samples (MCF7; HCC1806; MDA-MB-468)

Cells were seeded and left to adhere overnight before relevant samples were transferred to the hypoxia chamber. Cells were maintained in either normoxia or hypoxia as per Section 2.1 and Section 2.2. Cell pellets were collected after 24 hours, 7 days, or 14 days of culture in hypoxia or after 24 hours in normoxia and stored at -20℃ following a protocol set by Dr Ana Miar Cuervo (Buffa lab).

4.2.3 RNA extraction (MDA-MB-231)

RNA was extracted from frozen cell pellets of approximately 3 million cells using the miRNeasy Mini Kit according to manufacturer's instructions. The cell pellets were lysed using the QIAzol lysis buffer and at least 2 minutes of vortexing. Vortex time was extended if the cell pellet had not fully disintegrated. The optional wash step using Buffer RWT and the optional dry step were both included. For samples collected at 6 hours and 24 hours, RNA was diluted in 100 µL of RNase free water (miRNeasy Mini Kit). For samples collected after 7 days or 14 days, RNA was diluted in 150 µL of RNase free water. The A260/A280 and A260/A230 ratios were recorded using a NanoDrop Spectrophotometer. 1 µL of RNase free water was used to blank the instrument and 1 µL of each RNA sample was used for testing. The RNA was quantified using a Qubit 4 Fluorometer using the Qubit RNA Broad Range Assay Kit according to the manufacturer's instructions. The fluorescence reading is temperature sensitive, so all samples were incubated at room temperature for at least 5 minutes to ensure temperature equilibration before reading. 1 µL of sample was added to 199 µL of working reagent. 3.75 µg of each RNA sample was loaded into a full skirted 96 well PCR plate for sequencing.

4.2.4 RNA extraction (MCF7; HCC1806; MDA-MB-468)

RNA was extracted from frozen cell pellets using TRI Reagent according to the manufacturer's instructions and resuspended in DEPC treated water. To improve the A260/A230 ratio, sodium acetate precipitation was used. 0.1 volumes of 3 M sodium acetate was added along with 3 volumes of 100% ethanol and the solution incubated overnight at -20℃. Samples were centrifuged at 21130 *g* for 15 minutes at 4℃. The supernatants were discarded and the pellets washed in 1 mL of ice-cold 75% ethanol. Samples were centrifuged at 21130 *g* for 10 minutes at 4℃. The supernatants were discarded and the pellets dried before resuspending in DEPC treated water. The samples were treated with the TURBO DNA-free Kit to remove any contaminating genomic DNA following the "routine DNase treatment" method in the manufacturer's instructions. 3.5 µg of each RNA sample was loaded into a full skirted 96 well PCR plate for sequencing.

4.2.5 mRNA sequencing

Library preparation and sequencing were conducted by the Oxford Genomics Centre, Oxford. Material was quantified using the Quant-it RiboGreen RNA Assay Kit on the FLUOstar OPTIMA microplate reader and the size profile and integrity analysed on the TapeStation using the RNA ScreenTape assay. Input material was normalised prior to library preparation. Strand specific library preparation was completed using the QuantSeq 3'mRNA-Seq Kit (forward) following manufacturer's instructions. Libraries were amplified on a Tetrad PCR Thermal Cycler using in-house unique dual indexing primers (based on [115]). Individual libraries were normalised using Qubit and the size profile was analysed on the TapeStation. Individual libraries were normalised and pooled accordingly. The pooled library was diluted to approximately 10 nM for storage. The 10 nM library was denatured and further diluted prior to loading on the sequencer. 75 bp paired end sequencing was carried out on the NextSeq.

4.2.6 Small RNA sequencing

Library preparation and sequencing was conducted by the Oxford Genomics Centre, Oxford. The libraries were prepared using the NEBNext Small RNA Library Kit following the manufacturer's instructions. Size selection was carried out using BluePippin. Library quality control metrics were analysed using TapeStation before being multiplexed ready for

sequencing on NextSeq.

4.2.7 Processing of mRNA sequencing data

Reads from 3' mRNA sequencing were processed using the pipeline described in Section 9.3.1. The quality of the reads was assessed using FastQC (version 0.11.8). Adapter sequences, polyA sequences, and low quality bases were removed using cutadapt (version 2.7). Reads shorter than 15 bases after trimming were also removed. The cutadapt options used were: `-m 15 -n 2 --nextseq-trim=20 -a AGATCGGAAGAGC -a AAAAAAAAAAAAAAAAAAAA`. Reads were aligned to the Hsa.GRCh38 assembly of human genome (version 104) using STAR (version 2.7.3a). Aligned reads were quantified using the featureCounts program from the Subread package (version 1.6.4). The options used for featureCounts were: `-t exon -g gene_id -primary -s 1`. Downstream processing was conducted in R. Differentially expressed genes were identified using DESeq2 (version 1.38.3). Low expressing genes (i.e. genes that did not have at least 5 counts in at least 3 samples after normalisation) were removed from the analysis. Genes were annotated (i.e. their Ensembl IDs converted to gene names, gene symbols, and Entrez IDs) using AnnotationDbi (version 1.60.2) and the org.Hs.eg.db database (version 3.16.0).

4.2.8 Processing of small RNA sequencing data

Reads from small RNA sequencing were processed using the pipeline optimised in Section 3. The quality of the reads produced from sequencing was assessed using FastQC (version 0.11.9). Adapter sequences and low quality bases were removed using cutadapt (version 2.10). Only reads 16-28 bases in length after trimming were used for analysis. This ensured the majority of reads belong to the miRNA family. Manatee (version 1.2) was used both to align reads against the Hsa.GRCh38 assembly of human genome (version 104) and quantify reads using a modified (described in Section 3.2) version of the miRBase GFF3 file (version 22.1) as a reference. Reads with more than one mismatch were not included, and reads which mapped to more than 20 genomic loci were not considered. Specifically, the following options were used when running Manatee: `-collapse yes -mismatches 1 -m 20 -s yes`. Downstream processing was conducted in R. Differentially expressed miRNAs were identified using DESeq2 (version 1.38.3).

4.2.9 Analysis of TCGA data

RNA-seq profiles from The Cancer Genome Atlas (TCGA) were accessed using TCGAbiolinks (version 2.26.0) and separated by breast cancer subtype. Hypoxia scores were assigned to each tumour by applying the Buffa hypoxia signature [116] using SigQC [117]. This compiles RNA expression profiles of genes in the Buffa hypoxia signature and assigns each sample a hypoxia score based on this. The median score was taken as the hypoxia score. Tumours with scores in the upper and lower tertiles were labelled as tumours with “high” and “low” levels of hypoxia.

4.2.10 HIF1α western blot

15 µL of Rainbow Molecular Weight Marker was added to the first well of a NuPAGE 4-12% Bis-Tris Gel. Remaining wells were loaded with 30 µg of protein (collected in urea buffer as described in Section 2.3.1). Gels were run in 1X MES buffer. Protein was transferred using the semi-dry method (Section 2.3.7). The membrane was blocked using 5% nonfat dried milk powder then probed with a mouse anti-human HIF-1α primary antibody diluted 1:1000 in blocking buffer. The secondary antibody (goat anti-mouse IgG HRP) was diluted 1:1000 in blocking buffer. Full details for probing and developing the membrane are found in Section 2.3.9.

4.2.11 HIF2α western blot

15 µL of Rainbow Molecular Weight Marker was added to the first well of a NuPAGE 4-12% Bis-Tris Gel. Remaining wells were loaded with 40 µg of protein (collected in urea buffer as described in Section 2.3.1). Gels were run in 1X MES buffer. Protein was transferred using the wet transfer method (Section 2.3.6). The membrane was blocked using 5% nonfat dried milk powder then probed with a rabbit HIF2α primary antibody diluted 1:1000 in blocking buffer. The secondary antibody (goat anti-rabbit IgG HRP) was diluted 1:1000 in blocking buffer. Full details for probing and developing the membrane are found in Section 2.3.9.

4.2.12 E-cadherin western blot

15 µL of Rainbow Molecular Weight Marker was added to the first well of a NuPAGE 4-12% Bis-Tris Gel. Remaining wells were loaded with 30 µg of protein (collected in urea buffer as

described in Section 2.3.1). Gels were run in 1X MOPS buffer. Protein was transferred using the Turbo method (Section 2.3.8). The membrane was blocked using 5% nonfat dried milk powder then probed with the E-cadherin rabbit monoclonal antibody diluted 1:1000 in blocking buffer. The secondary antibody (goat anti-rabbit IgG HRP) was diluted to 1:1000 in blocking buffer. Full details for probing and developing the membrane are found in Section 2.3.9.

4.2.13 Vimentin western blot

15 µL of Rainbow Molecular Weight Marker was added to the first well of a NuPAGE 4-12% Bis-Tris Gel. Remaining wells were loaded with 30 µg of protein (collected in RIPA buffer as described in Section 2.3.2). Gels were run in 1X MES buffer. Protein was transferred using the Turbo method (Section 2.3.8). The membrane was blocked using 5% nonfat dried milk powder then probed with the vimentin rabbit monoclonal antibody diluted 1:1000 in blocking buffer. The secondary antibody (goat anti-rabbit IgG HRP) was diluted to 1:1000 in blocking buffer. Full details for probing and developing the membrane are found in Section 2.3.9.

4.2.14 Wound healing assay

$6.5x10^4$ cells were seeded per well of an Incucyte ImageLock 96-well plate in 100 µL of culture media (24 replicate wells per plate). Cells had either been cultured in hypoxia for 13 days prior to seeding or were the equivalent passage in normoxia. Cells were left to adhere overnight before a wound was scratched into the cell monolayer. The wound was scratched using an Incucyte 96-Well Woundmaker Tool according to manufacturer's instructions. Half of the media (50 µL) was removed from the well and replaced with 75 µL of fresh culture media. This removed some of the scratched cells which were now floating. This differs from a typical protocol, where all the media would be removed and replaced, because cells in hypoxia had a tendency to detach from the plate if all media was removed. The plate was immediately transferred to an Incucyte (Incucyte S3 or Incucyte Zoom) in an incubator (set to 36℃, humidified, 5% CO_2), where images were taken at 2 hour intervals for at least 36 hours. This meant that hypoxic cells were transferred to a normoxic environment for the assay. Analysis was conducted using a different set of training images for each cell type. "Masks" of the wound and cells were created at each timepoint (Figure 17) and the relative wound density calculated. This method accounts for the density of the wound at the initial time point (i.e. incomplete scratching or cells re-settling) and for changes in the density of the cell

monolayer inside and outside of the wound region throughout the experiment.

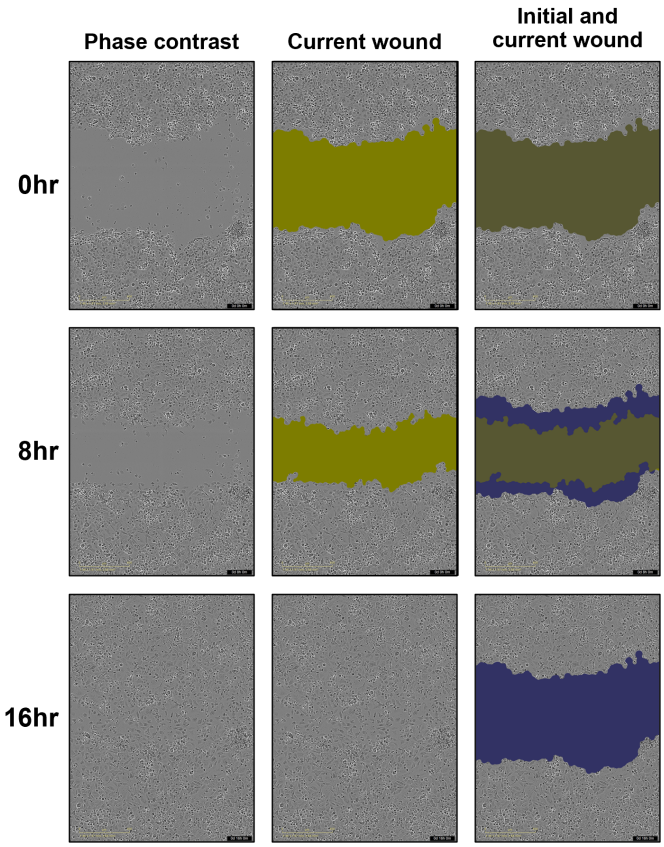

Figure 17: Masking in wound healing analysis. The wound is scratched at 0hr, and the initial and current wound masks fully overlay. Over time, cells enter the wound region and the overlay is reduced. The difference in cell density between the current and initial wound regions can be used to calculate the relative wound density.

4.2.15 Adhesion assay

$1x10^4$ cells were seeded per well of a 96 well plate in 100 µL of culture media. Quadruplicate wells were prepared in each plate. Four separate plates were seeded for each timepoint. Cells had either been cultured in hypoxia for 14 days prior to seeding or were the equivalent passage in normoxia. After 30 minutes, 1 hour, and 2 hours, plates were inverted to remove the media along with any non-adhered cells. Each well was rinsed twice using 200 µL of PBS and again emptied by inversion. At 2 hours, one of the plates was centrifuged at 300 *g* for 4 minutes before the rinse step. This plate represents the total number of cells. Plates were stored (with empty wells) at -80℃ overnight. Plates were thawed and the cells quantified using the CyQUANT Cell Proliferation Assay according to manufacturer's instructions. Briefly,

cell lysis buffer and CyQUANT GR dye (a proprietary dye that fluoresces when bound to nucleic acids) were added to the wells and incubated at room temperature for 5 minutes. The fluorescence was measured using a POLARstar Omega microplate reader with the following settings: excitation 480 nm, emission 520 nm. To calculate the percentage of cells that had adhered by each timepoint, the fluorescence at each timepoint was compared to the “total” cells control.

4.3 Results

4.3.1 Chronic hypoxia reduces cell growth

Since there have been few studies on truly chronic hypoxia, and those that use longer timepoints often used a hypoxic incubator where cells are exposed to a normoxic environment during passaging [118], it was not known whether our chosen panel of breast cancer cell lines could survive 14 days of exposure to hypoxia and whether they would proliferate under these conditions. The cell number was monitored in normoxia and hypoxia (1% O_2) over 14 days. The results show that cells were able to survive under chronic hypoxia and continued to proliferate, though at a reduced rate (Figure 18). The cells in this experiment are unlikely to have experienced nutrient deprivation or a drop in pH due to the high volume of media used per flask and the frequent media changes employed.

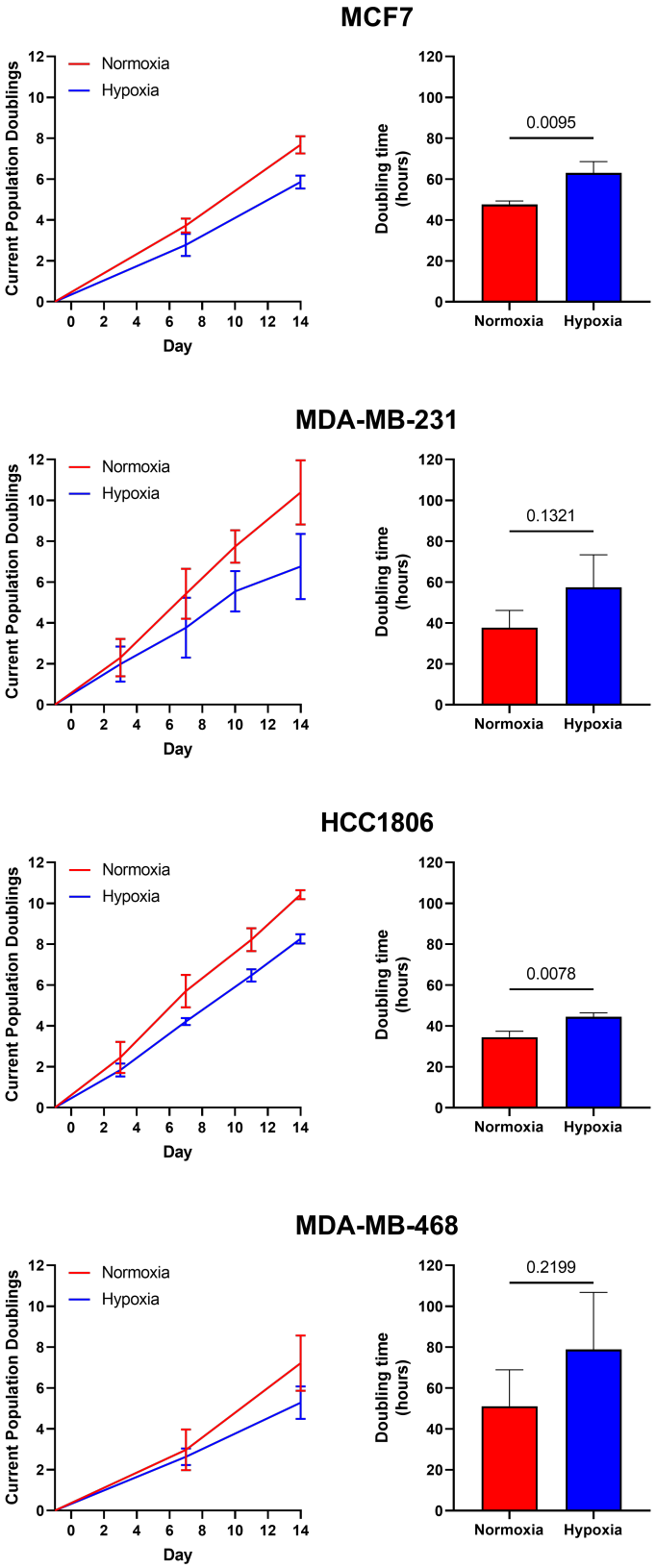

Figure 18: Cell growth under chronic hypoxia. Current population doublings, and the average doubling time in hours across the experiment for each cell line in normoxia and hypoxia. An unpaired t-test was used to test for differences in doubling times *(n = 3)*.

4.3.2 The classical hypoxia response is maintained for at least 14 days of hypoxia exposure

There is evidence that levels of the HIF proteins change depending on the length of hypoxia exposure [24, 25, 111]. Typically, levels of HIF1 peak before HIF2, but HIF2 is maintained for longer time periods. However, these studies still monitor HIF expression over relatively short time periods. To test whether the HIF1 and HIF2 proteins are maintained in chronic hypoxia timepoints, western blot analysis was conducted on samples collected under acute (6 hours and 24 hours) and chronic (7 days and 14 days) hypoxia (Figure 19).

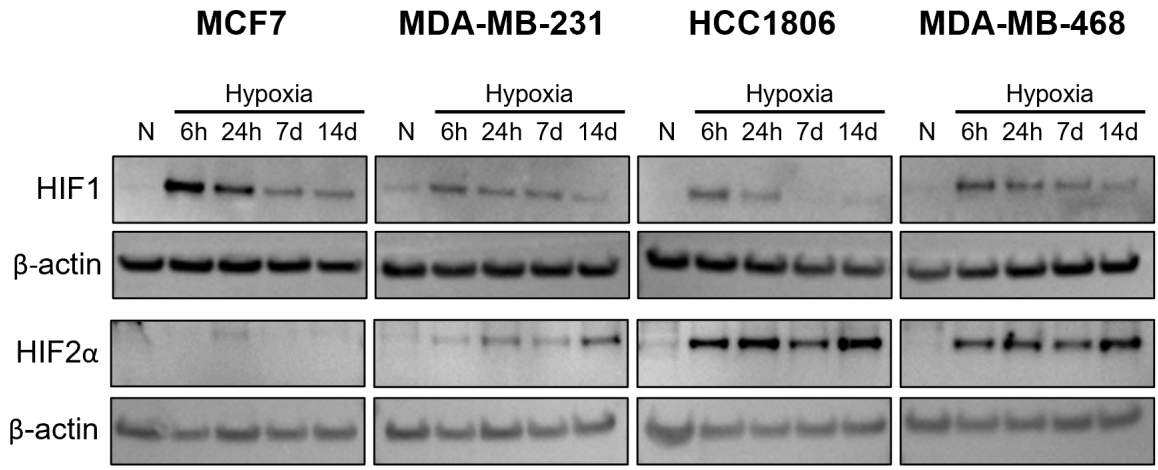

Figure 19: Kinetics of HIF1 and HIF2 under chronic hypoxia. Western blots showing expression levels of HIF1α and HIF2α along with the β-actin control.

In general, HIF1 levels were highest at the 6 hour timepoint and depleted over time (Figure 20). In the case of HCC1806, HIF1 dropped back to baseline levels by day 7. In the other three cell lines, HIF1 levels were lower in the chronic timepoints, but some expression was maintained. In contrast, HIF2 expression was more stable between acute and chronic hypoxia, but showed larger cell line variability. HIF2 expression was almost entirely absent in MCF7 cells. However, MCF7 cells had the highest overall expression of HIF1 which could compensate for the lack of HIF2 activity. The other cell lines expressed HIF2 by the 6 hour timepoint and the expression remained relatively stable up to 14 days (Figure 21). HCC1806 cells had the highest overall level of HIF2, which could compensate for the lack of HIF1 at the chronic timepoints. Together, these data show that across cell lines, at least one of the two HIF subtypes is maintained up to 14 days of hypoxia exposure.

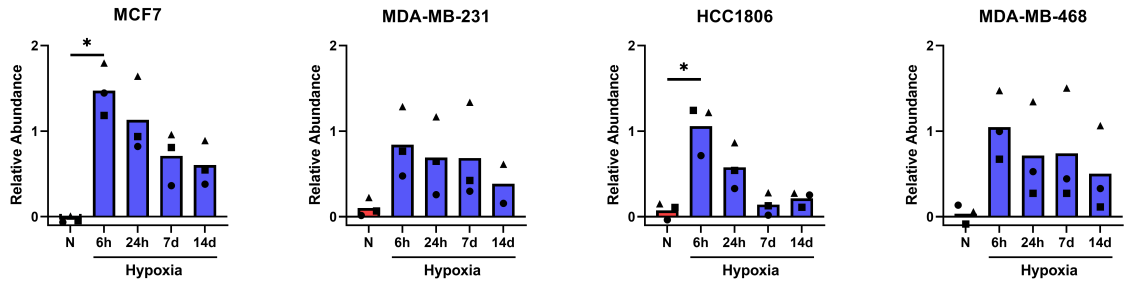

Figure 20: HIF1 western blot quantification. Levels of HIF1α compared to β-actin measured by western blot. Each of the three replicates is displayed as a different shape. Significance was tested using one-way ANOVA with repeated measures and corrected for multiple testing using Dunnett's method (* $p \leqslant 0.05$).

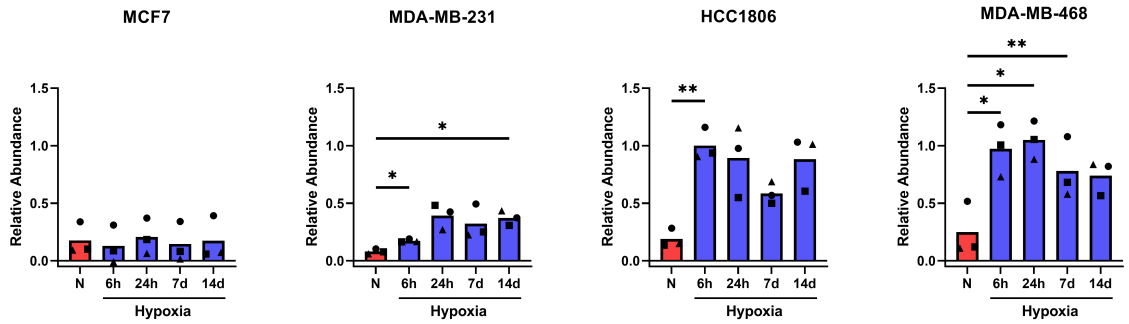

Figure 21: HIF2 western blot quantification. Levels of HIF2α compared to β-actin measured by western blot. Each of the three replicates is displayed as a different shape. Significance was tested using one-way ANOVA with repeated measures and corrected for multiple testing using Dunnett's method (* $p \leqslant 0.05$, ** $p \leqslant 0.01$).

To investigate the response to chronic hypoxia in greater depth, RNA sequencing was conducted on cells cultured under hypoxia for 24 hours, 7 days, and 14 days. Analysis of the results using decoupleR (version 2.4.0), a program which infers transcription factor activity using expression levels of the target genes, predicts that either HIF1 or HIF2 (*EPAS1*) are the most active transcription factors in all hypoxic samples across all timepoints (Figures 22 and 23). Interestingly, this program was also able to pick up on some of the differences in HIF1 and HIF2 activity between the cell lines. decoupleR predicted that HIF2 was more active than HIF1 in MDA-MB-468 cells, which have high levels of HIF2, while MCF7 cells were predicted to have high HIF1 activity compared to HIF2 in accordance with western blot analyses (Figures 20

and 21). There does not appear to be a switch in the most active HIF gene over time in this analysis. Over-representation analysis using clusterProfiler (version 4.6.2) showed that the gene ontology term “cellular response to hypoxia” was significantly (adjusted p value $\leqslant 0.05$) enriched in both acute and chronic hypoxia when compared to normoxia, and this was true across all four cell lines.

Finally, I assessed the correlation of $\log_2$ fold changes of known hypoxia response genes between the different hypoxia timepoints. Three different lists of hypoxia response genes were used:

1. Direct HIF targets (ChIP Atlas) [119]
2. A curated HIF regulon (DoRothEA) [120]
3. An established hypoxia gene signature (Buffa) [116]

Overall, there was a very strong correlation between the $\log_2$ fold change of these genes after 24 hours of hypoxia exposure and the chronic timepoints (Figures 24 to 26). This means that genes which are upregulated after 24 hours of hypoxia exposure are upregulated to a similar degree in the chronic hypoxia timepoints. This suggests that the same pathways are needed to maintain survival under hypoxia regardless of exposure time. This also indicates that existing hypoxia gene signatures which have been generated using acute hypoxia datasets are adequate for classification of samples experiencing chronic hypoxia.

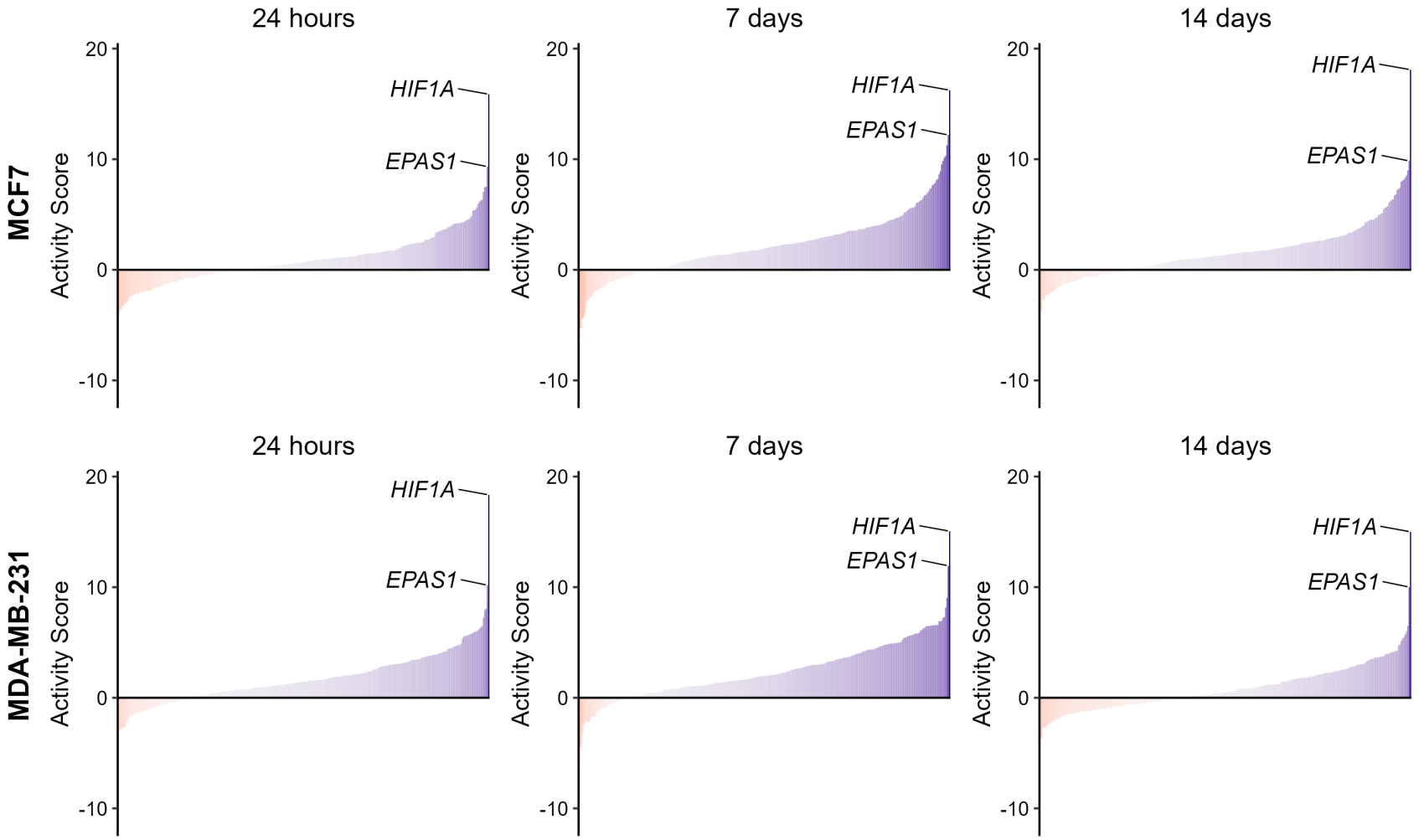

Figure 22: Active transcription factors in MCF7 and MDA-MB-231 cells under hypoxia. The activity of 429 transcription factors in hypoxic samples compared to the normoxic control as predicted by decoupleR. *HIF1A* (HIF1)and *EPAS1* (HIF2) have the highest predicted activity of all transcription factors. Blue represents transcription factors which are more active in hypoxia, while red represents those more active in normoxia.

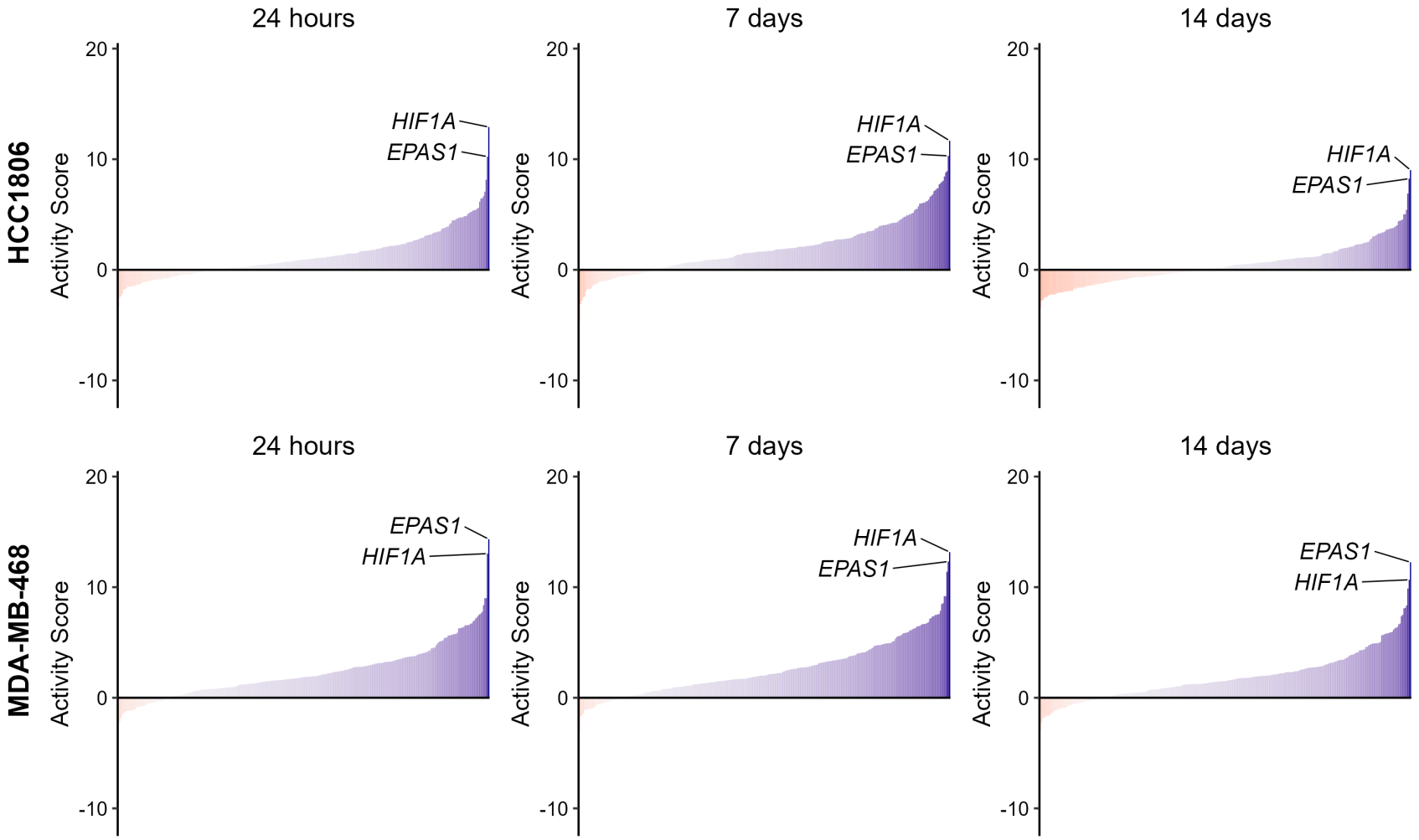

Figure 23: Active transcription factors in HCC1806 and MDA-MB-468 cells under hypoxia. The activity of 429 transcription factors in hypoxic samples compared to the normoxic control as predicted by decoupleR. *HIF1A* (HIF1)and *EPAS1* (HIF2) have the highest predicted activity of all transcription factors. Blue represents transcription factors which are more active in hypoxia, while red represents those more active in normoxia.

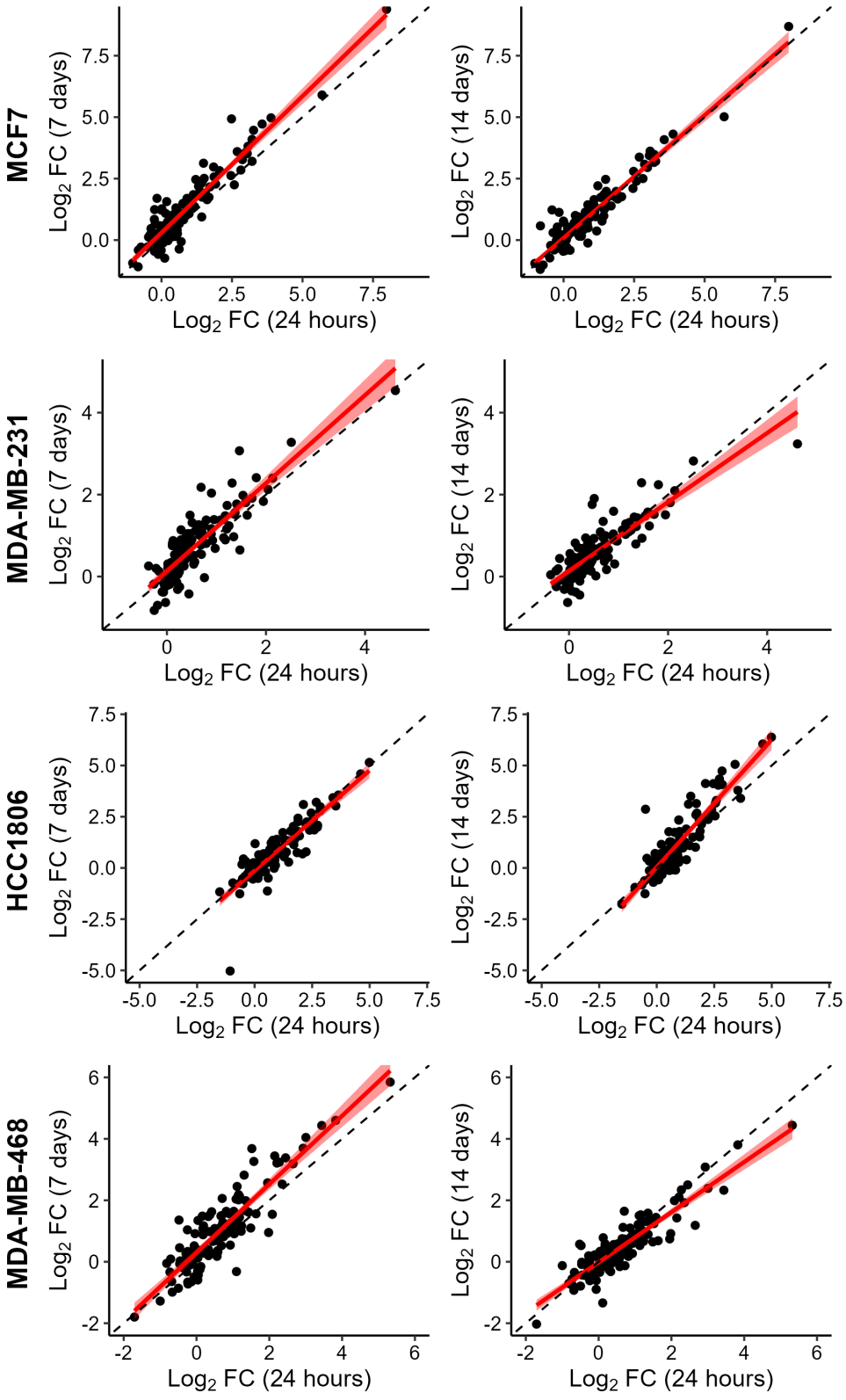

Figure 24: Correlation of the $\log_2$ fold changes of HIF target genes (ChIP Atlas) between different hypoxia timepoints. The $\log_2$ fold change of HIF target genes (ChIP Atlas) between acute and chronic hypoxia is strongly correlated. The dashed line represents $x = y$. The red line shows the linear model with surrounding confidence intervals.

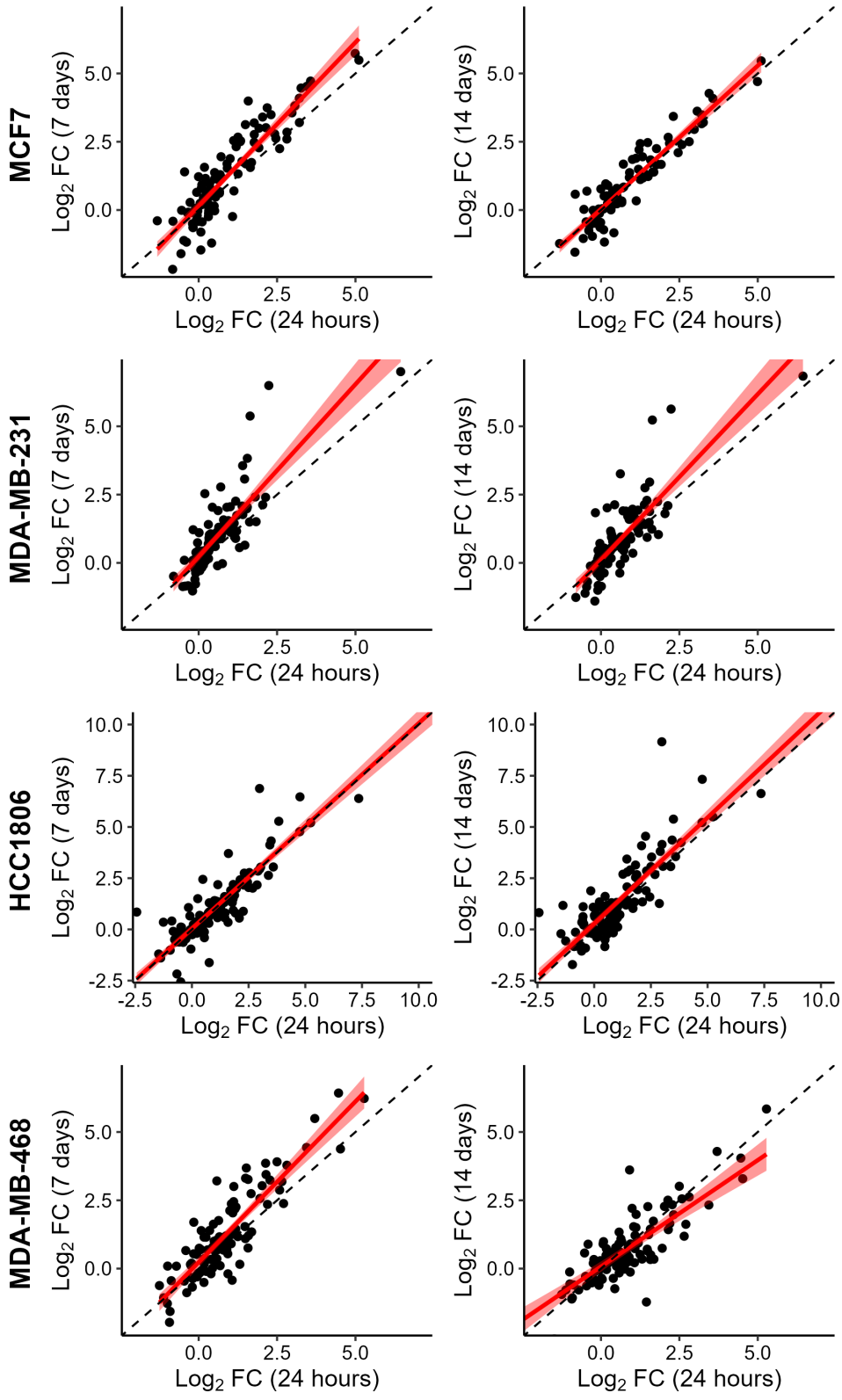

Figure 25: Correlation of the $\log_2$ fold changes of HIF target genes (DoRothEA) between different hypoxia timepoints. The $\log_2$ fold change of HIF target genes (DoRothEA) between acute and chronic hypoxia is strongly correlated. The dashed line represents $x = y$. The red line shows the linear model with surrounding confidence intervals.

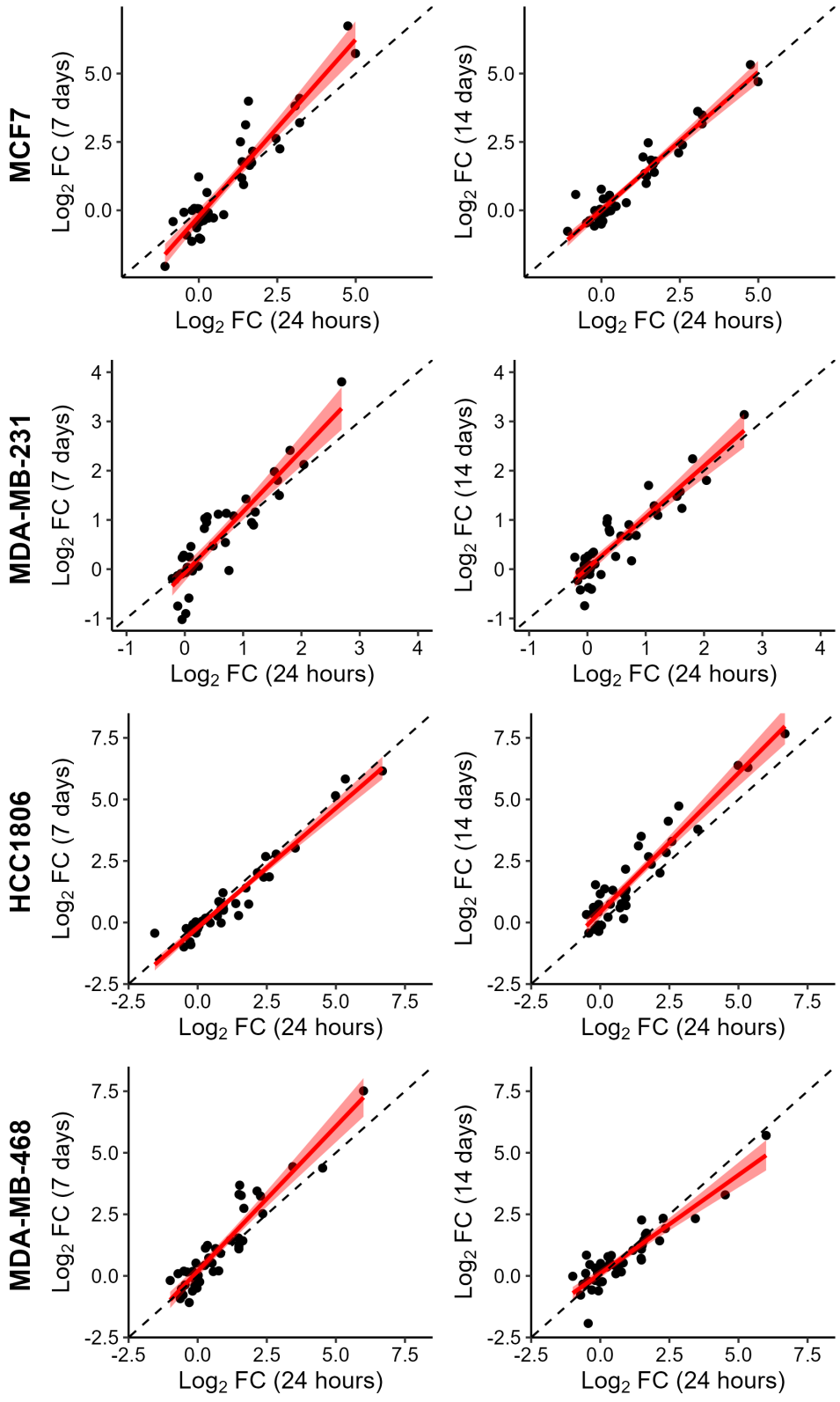

Figure 26: Correlation of the $\log_2$ fold changes of hypoxia signature genes (Buffa) between different hypoxia timepoints. The $\log_2$ fold change of hypoxia signature genes (Buffa) between acute and chronic hypoxia is strongly correlated. The dashed line represents $x = y$. The red line shows the linear model with surrounding confidence intervals.

Together, these data show that the levels of HIF1/HIF2 found at the chronic timepoints are sufficient to maintain the classical hypoxia response, and that the hypoxic response remains the most dominant activated pathway for at least 14 days after the beginning of hypoxia exposure.

4.3.3 Chronic hypoxia drives changes in the expression of structural genes

After establishing that the classical hypoxia response remained largely the same in acute and chronic hypoxia, I asked if there were any changes in cells exposed to chronic hypoxia that did not occur under acute hypoxia. In general, cells had the greatest number of differentially expressed genes after 7 days of hypoxia exposure, and these were more likely to be found in common with differentially expressed genes at day 14 than at day 1 (Figures 27 to 30). This suggests that there is a transcriptional profile that is induced by chronic hypoxia, but not acute.

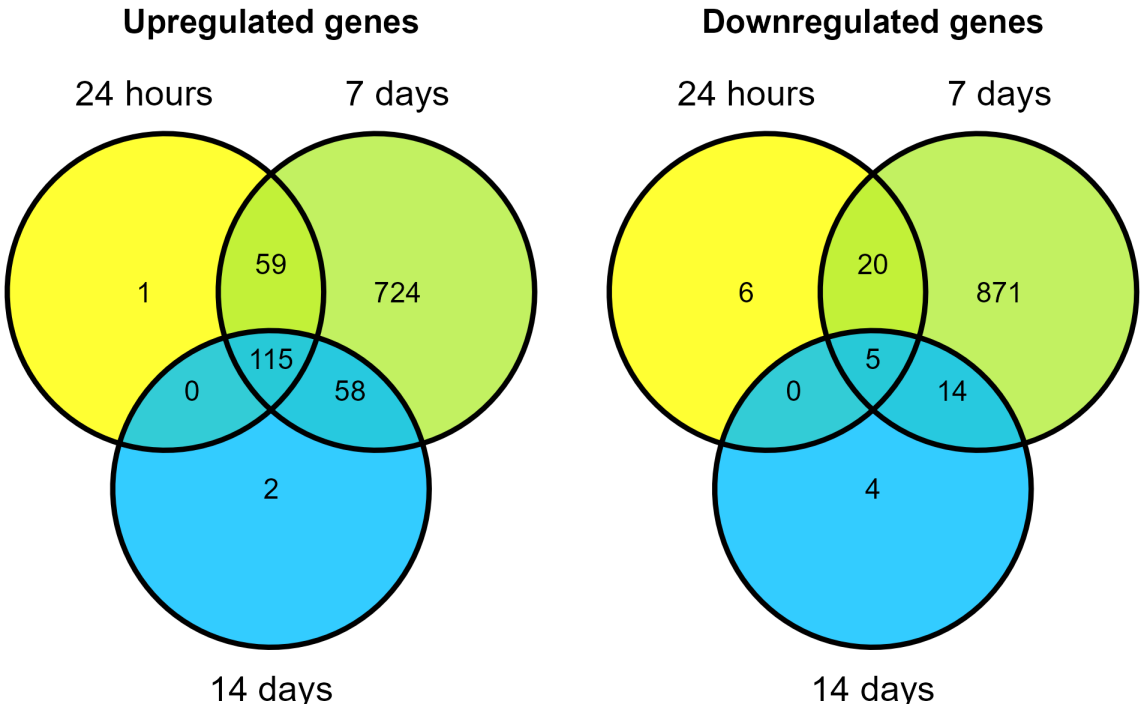

Figure 27: Overlap of differentially expressed genes at each hypoxia timepoint (MCF7). The overlap of significantly (adjusted p value $\leqslant 0.05$) differentially expressed genes at each hypoxia timepoint in MCF7 cells.

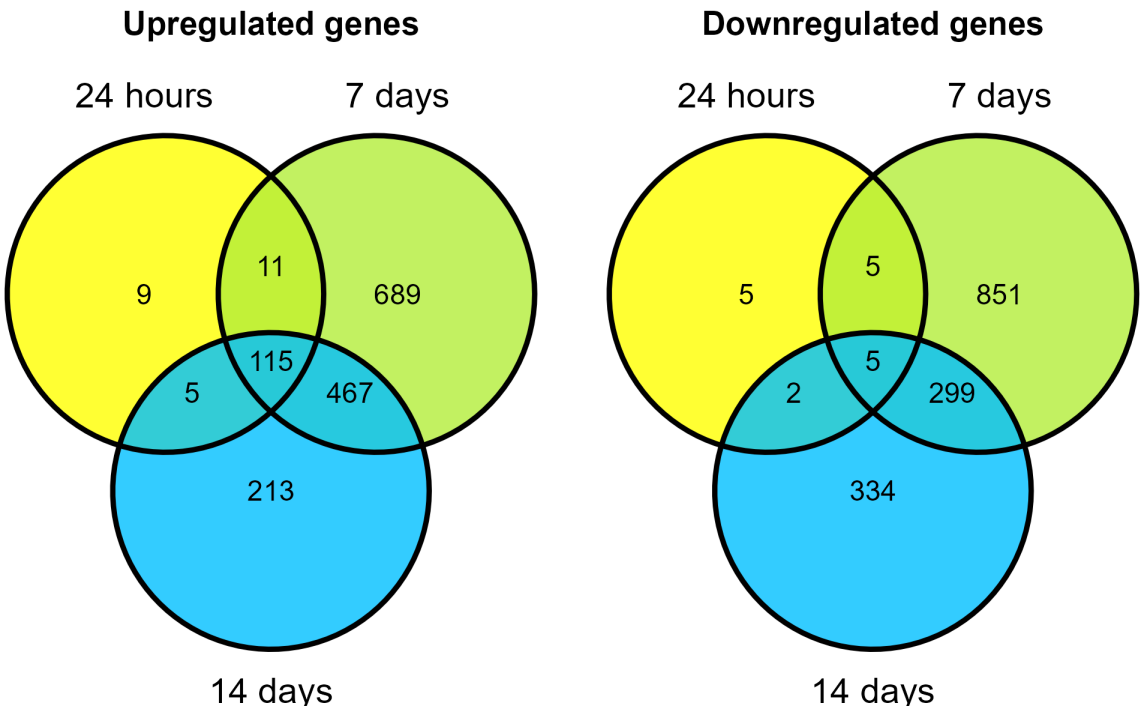

Figure 28: Overlap of differentially expressed genes at each hypoxia timepoint (MDA-MB-231). The overlap of significantly (adjusted p value $\leqslant 0.05$) differentially expressed genes at each hypoxia timepoint in MDA-MB-231 cells.

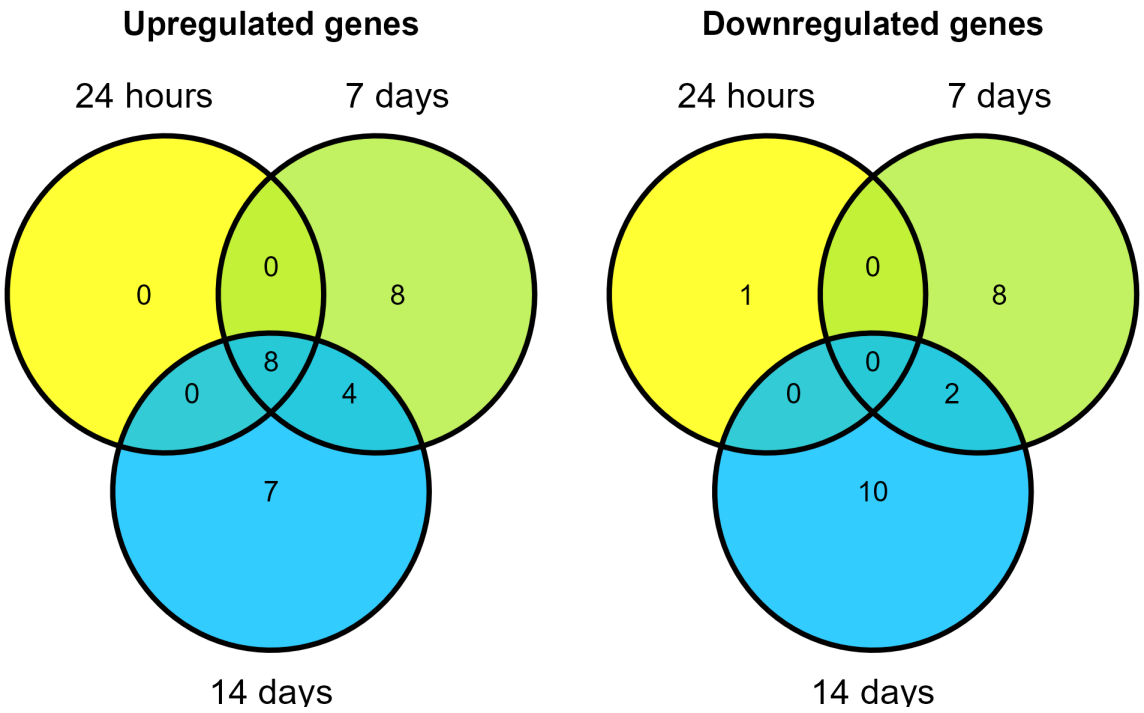

Figure 29: Overlap of differentially expressed genes at each hypoxia timepoint (HCC1806). The overlap of significantly (adjusted p value $\leqslant 0.05$) differentially expressed genes at each hypoxia timepoint in HCC1806 cells.

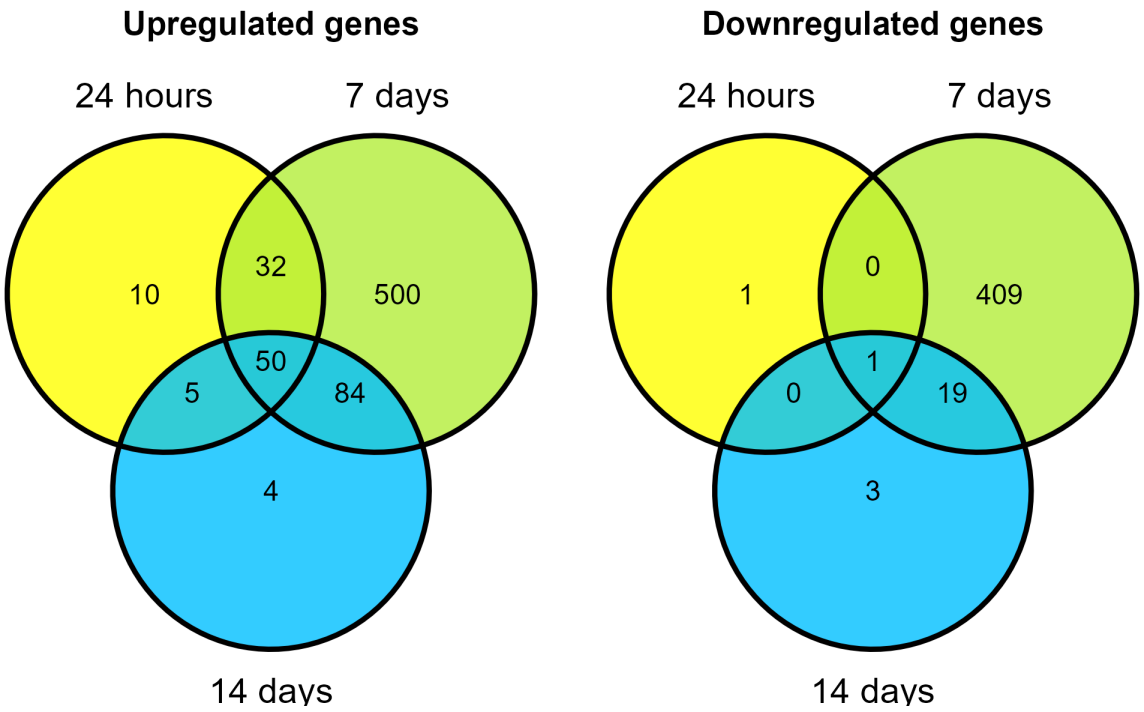

Figure 30: Overlap of differentially expressed genes at each hypoxia timepoint (MDA-MB-468). The overlap of significantly (adjusted p value $\leqslant 0.05$) differentially expressed genes at each hypoxia timepoint in MDA-MB-468 cells.

I next assessed the overlap of genes upregulated specifically under chronic hypoxia between cell lines. These results show surprisingly little overlap. Only one gene is identified as chronic hypoxia specific in more than two of the cell lines (Table 2).

QSOX1 is the only gene which was upregulated specifically under chronic hypoxia across all three TNBC cell lines (Figure 31). QSOX1 (quiescin sulfhydryl oxidase 1) catalyses disulfide bond formation during protein folding, reducing molecular oxygen to hydrogen peroxide in the process [121]. QSOX1 has already been associated with tumorigenesis and is reported to play a role in cancer cell invasion [121], although there is some conflicting evidence on its role in breast cancer [122, 123]. In fibroblasts, QSOX1 was shown to be essential for integration of laminin into the basement membrane but not for laminin production [124]. This led to reduced migration of cancer cells across ECM produced by QSOX1 knockdown fibroblasts. This is interesting as two other genes in Table 2, *LAMB3* and *LAMC2*, encode subunits of laminin [125]. This strongly suggests that dysregulation of basement membrane production is occurring under chronic hypoxia, which *in vivo* could lead to increased invasion and metastasis.

Table 2: Chronic hypoxia specific upregulated genes. Genes that are significantly (adjusted p value $\leqslant 0.05$) upregulated after 7 and 14 days of hypoxia exposure, but not after 24 hours. Only genes which fulfil these criteria in at least two of the cell lines are shown.

MCF7	MDA-MB-231	HCC1806	MDA-MB-468
ANKRD37	ANKRD37		
ARHGAP29	ARHGAP29		
ATP1B1	ATP1B1		
BCL6	BCL6		
	CDCP1		CDCP1
	CDKN1A	CDKN1A	
	CPA4		CPA4
DYNC2I1	DYNC2I1		
EFNA1	EFNA1		
	EGLN3	EGLN3	
	ETS1		ETS1
FER1L4	FER1L4		
	FKBP9P1		FKBP9P1
	GJB3		GJB3
	GPRC5A		GPRC5A
	ICAM1		ICAM1
ID4	ID4		
IGFBP5			IGFBP5
	IL7R		IL7R
	INHBA		INHBA
	KLF7		KLF7
	KRT34		KRT34
	LAMB3		LAMB3
	LAMC2		LAMC2
	MGLL		MGLL
	MSN		MSN
	MT1X	MT1X	
PLOD2			PLOD2
	PNRC1	PNRC1	
	PPP1R14B-AS1		PPP1R14B-AS1
PPP1R15A	PPP1R15A		
PPP1R3E	PPP1R3E		
	PSG5		PSG5
	QSOX1	QSOX1	QSOX1
	RAB3B		RAB3B
RORA	RORA		
	SAMD4A	SAMD4A	
SNORD13			SNORD13
	TNFAIP3		TNFAIP3
	UPP1		UPP1
		VIM	VIM

In addition, *LAMB3* and *LAMC2* both encode subunits of laminin 332 [125]. Laminin 332 is reported as the ECM component most able to stimulate cell adhesion, migration, and invasion, and is a biomarker for EMT in cancer [126, 127]. In accordance, laminin 332 is associated with invasive breast ductal carcinomas and has been reported to increase migration in breast cancer cells [127, 128]. Knockdown of *LAMC2* was able to reduce the growth and invasive properties of thyroid cancer cells [129].

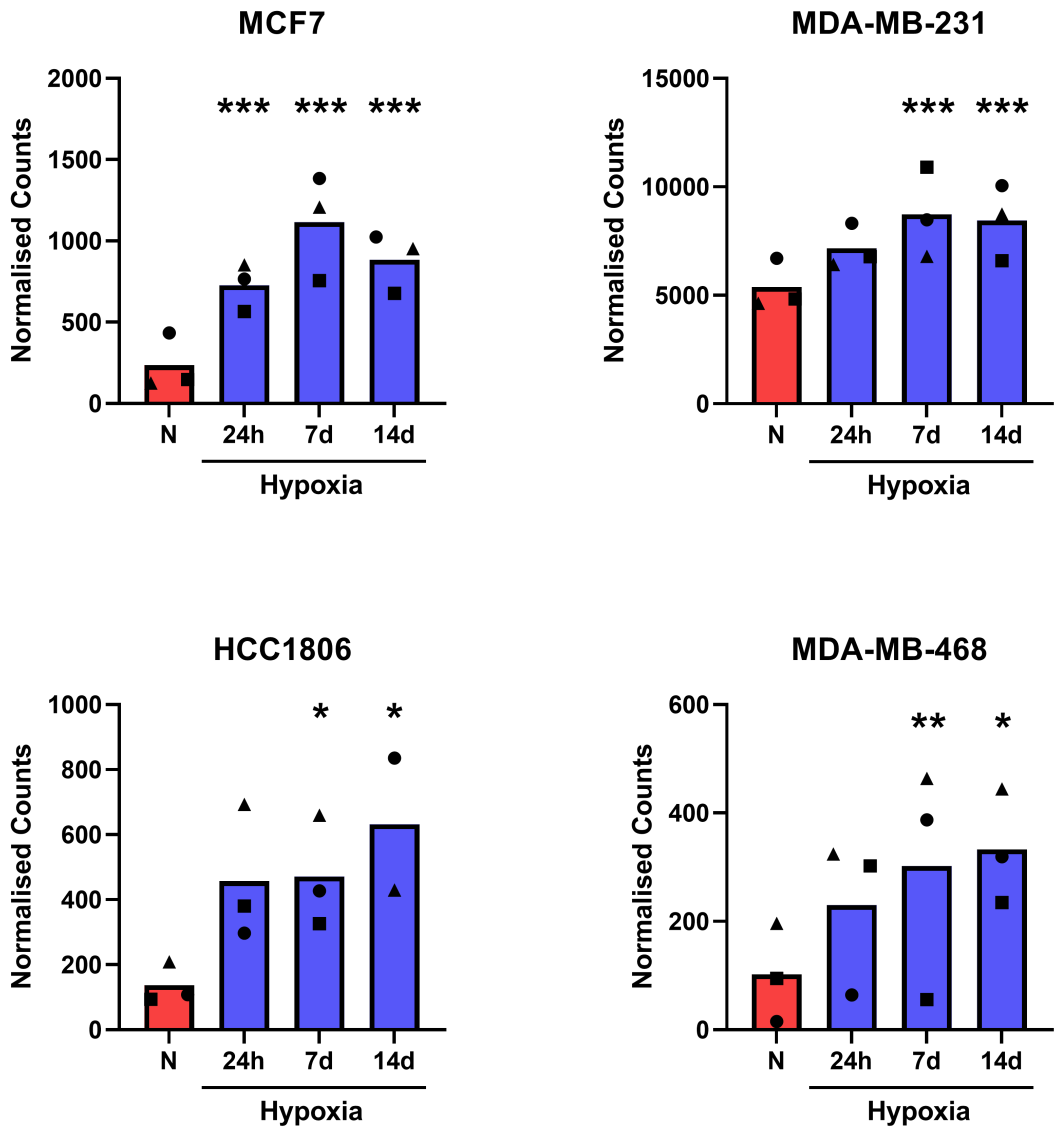

Figure 31: QSOX1 expression. The expression of *QSOX1* across each cell line. *QSOX1* was significantly (adjusted p value $\leqslant 0.05$) upregulated after 7 days and 14 days, but not 24 hours, of hypoxia in the TNBC cell lines. In MCF7 cells, QSOX1 was significantly (adjusted p value $\leqslant 0.05$) upregulated at all hypoxia timepoints (* $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$). Each shape represents a different biological replicate.

Several other genes found in Table 2 are associated with adhesion and EMT. For example, *PLOD2* is involved in the production of collagen, catalysing the critical lysyl hydroxylation step [130]. It is a known HIF target and is often over-expressed in cancers [130, 131]. A study by Xu et al. [132] showed that knockdown of PLOD2 pushed cervical cancer cells towards a more epithelial phenotype, suggesting that PLOD2 is a promoter of EMT. In the same study, knockdown of PLOD2 decreased cell adhesion to collagen. Through these mechanisms, PLOD2 is associated with cancer progression [130, 131].

Another example of an EMT associated gene in Table 2 is *VIM*. *VIM* encodes the intermediate filament protein vimentin, an important marker of EMT which is strongly associated with cell migration, invasion, and adhesion [133]. Vimentin is essential for the formation of invadopodia and the development of front-rear polarity in migrating cells [134]. Consistent with this function, vimentin is associated with poor prognosis [134], and knockdown of vimentin was shown to reduce migration and cell adhesion in MDA-MB-231 cells [135].

ICAM1 encodes an intracellular adhesion molecule which plays a role in both cell-cell and cell-ECM adhesion [136]. It is highly expressed in TNBC compared to other breast cancer subtypes and is associated with poor prognosis [137, 138]. Inhibition of ICAM1 led to reduced invasion in breast cancer cell lines, while knockdown of ICAM1 in a TNBC cell line reduced metastasis to the bone by impairing EMT [138, 139].

MSN encodes moesin, which aids the formation of invadopodia through its interaction with extracellular matrix protein 1, thereby facilitating cell migration [140]. In accordance with this, high expression of moesin was associated with metastasis and poor prognosis in several cancers, including breast cancer [140]. Knockdown of moesin caused mesenchymal-epithelial transition to occur in p53 mutant MCF7 cells [141], suggesting it plays a role in EMT.

INHBA codes for inhibin β-A, a member of the TGF-β superfamily. INHBA is over-expressed in many types of cancer and is often associated with poor prognosis [142–144]. Knockdown of INHBA in breast cancer cells reduced growth and invasive properties by inducing a more epithelial phenotype [144], while over-expression of INHBA produced the opposite results [144]. This suggests that INHBA plays a positive role in regulating EMT.

Unbiased analyses of pathway enrichment in acute and chronic hypoxia conditions highlight similar trends in the data (Figures 32 to 35). For example, HCC1806 cells show a difference in genes involved in cell-substrate adhesion after 7 days of hypoxia exposure. MDA-MB-468 cells show a clear transition between acute hypoxia, where most over-represented pathways are related to the classical hypoxia response, to chronic hypoxia where pathways relating to cell adhesion, migration, and cell structure become enriched. By day 14, MDA-MB-231 cells also show differences in cell adhesion by pathway enrichment. However, most of the chronic hypoxia specific pathways relate to the cell cycle, ER stress, and apoptosis. The lack of EMT associated pathways could be due to MDA-MB-231 cells already having a strong mesenchymal phenotype (i.e. it belongs to the mesenchymal stem like subtype of TNBC [113]), which means hypoxia does not induce further changes. MCF7 cells appear to have a more stable phenotype across the different hypoxia timepoints. This could reflect the less aggressive nature of MCF7 cells which may prevent changes associated with invasion and metastasis from developing as readily.

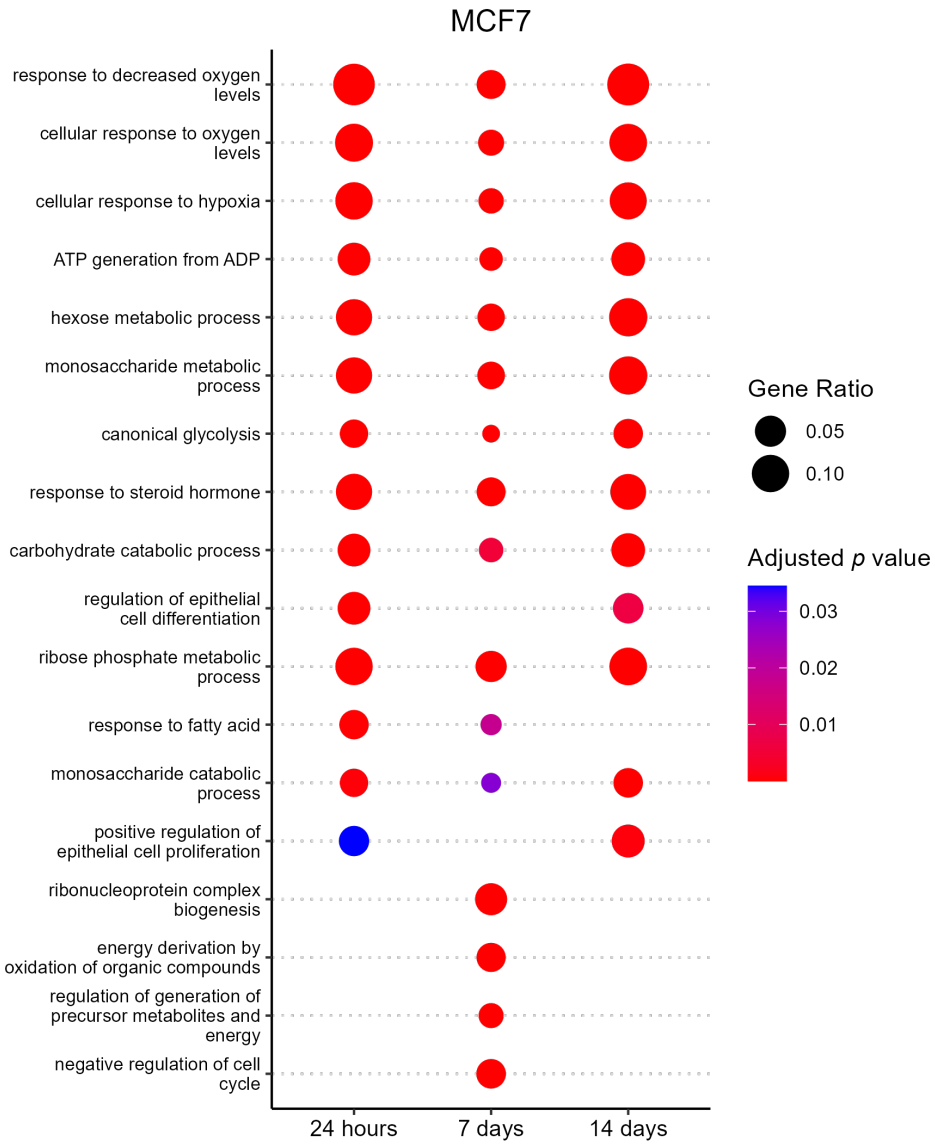

Figure 32: Over-represented pathways in acute and chronic hypoxia in MCF7 cells. Genes that are significantly (adjusted p value $\leqslant 0.05$) differentially expressed at each hypoxia timepoint are used in over-representation analysis. Pathways which are significantly (adjusted p value $\leqslant 0.05$) enriched are shown. The Gene Ratio refers to the number of differentially expressed genes found in the pathway, divided by the total number of differentially expressed genes in that condition.

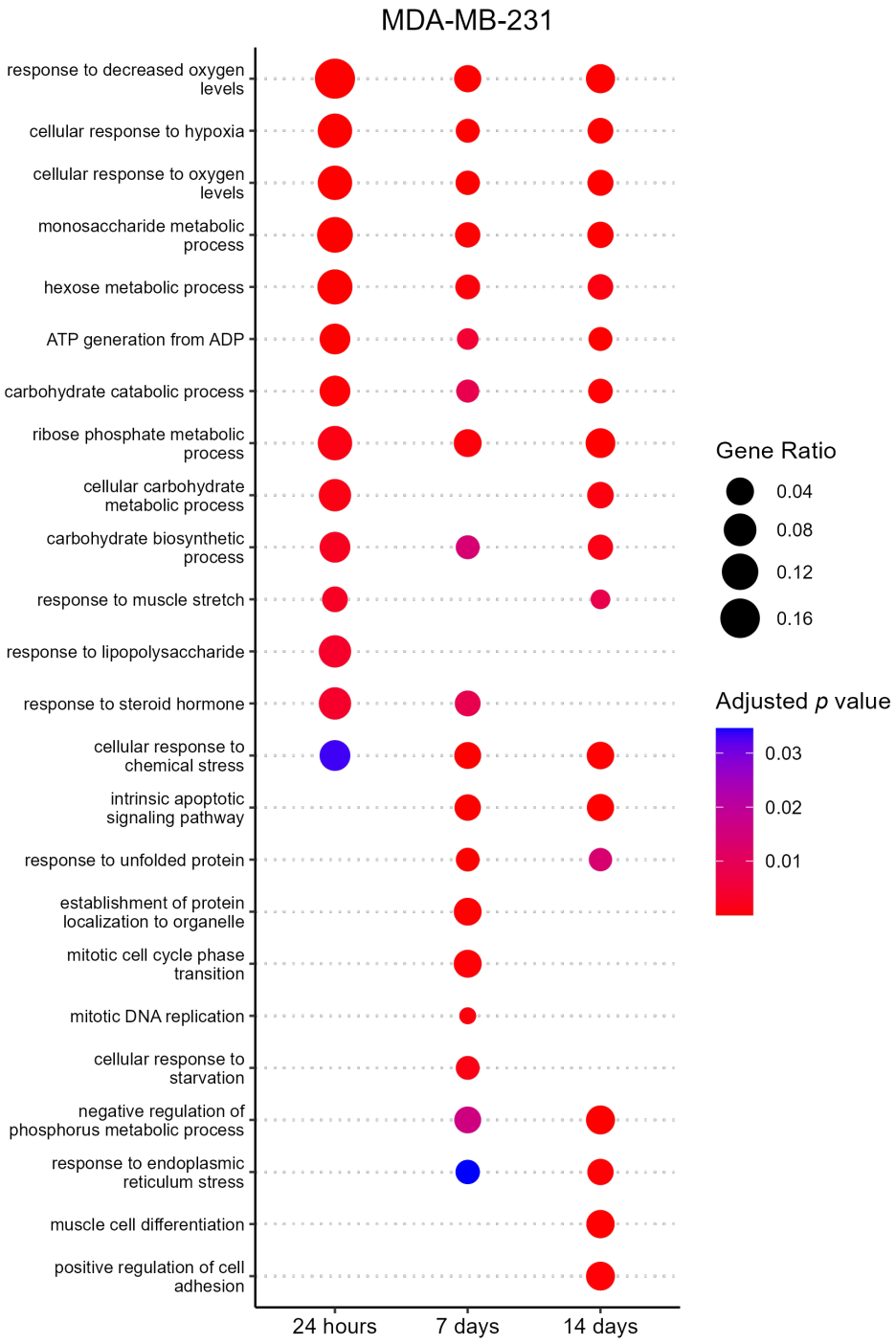

Figure 33: Over-represented pathways in acute and chronic hypoxia in MDA-MB-231 cells. Genes that are significantly (adjusted p value $\leqslant 0.05$) differentially expressed at each hypoxia timepoint are used in over-representation analysis. Pathways which are significantly (adjusted p value $\leqslant 0.05$) enriched are shown. The Gene Ratio refers to the number of differentially expressed genes found in the pathway, divided by the total number of differentially expressed genes in that condition.

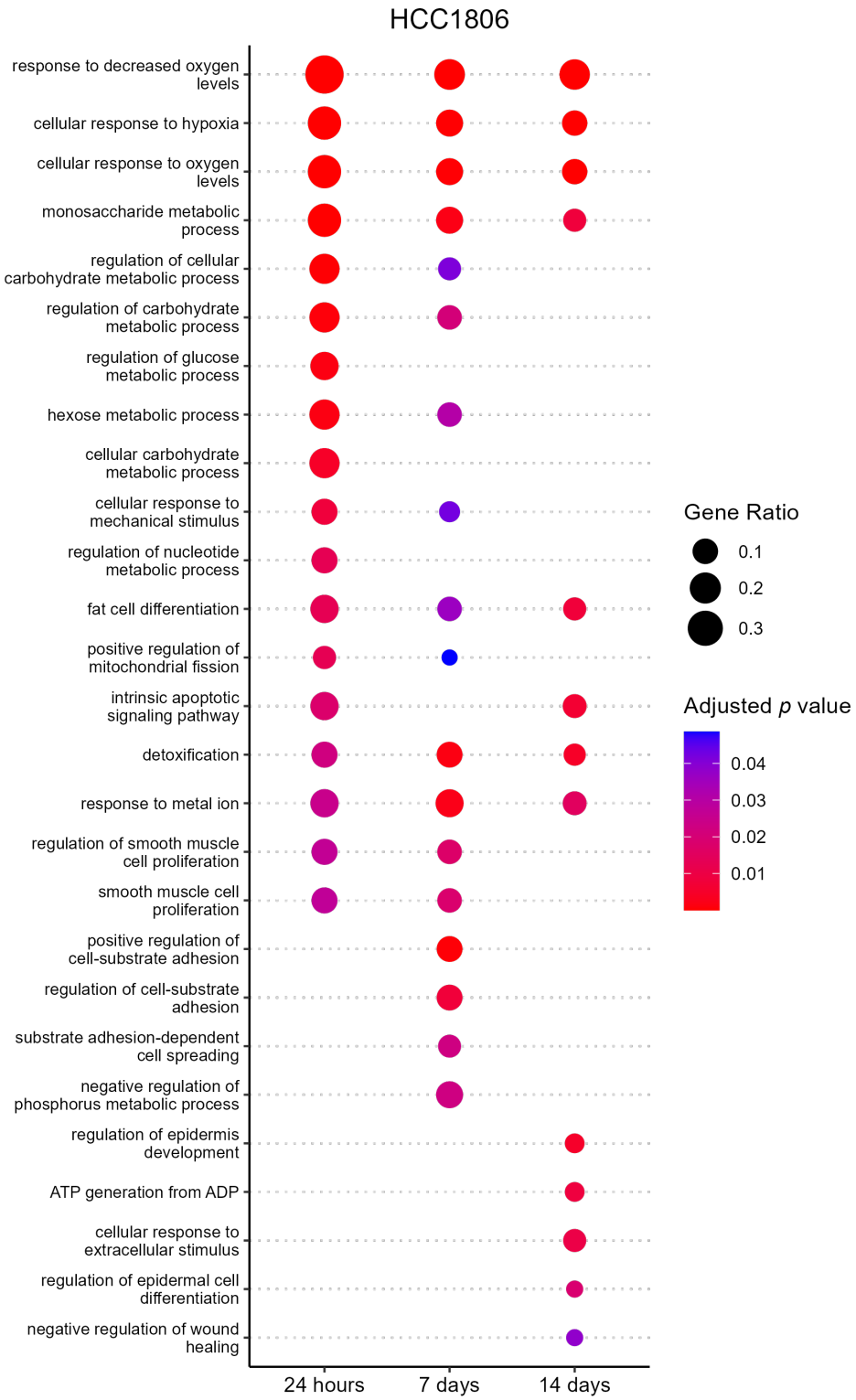

Figure 34: Over-represented pathways in acute and chronic hypoxia in HCC1806 cells. Genes that are significantly (adjusted p value $\leqslant 0.05$) differentially expressed at each hypoxia timepoint are used in over-representation analysis. Pathways which are significantly (adjusted p value $\leqslant 0.05$) enriched are shown. The Gene Ratio refers to the number of differentially expressed genes found in the pathway, divided by the total number of differentially expressed genes in that condition.

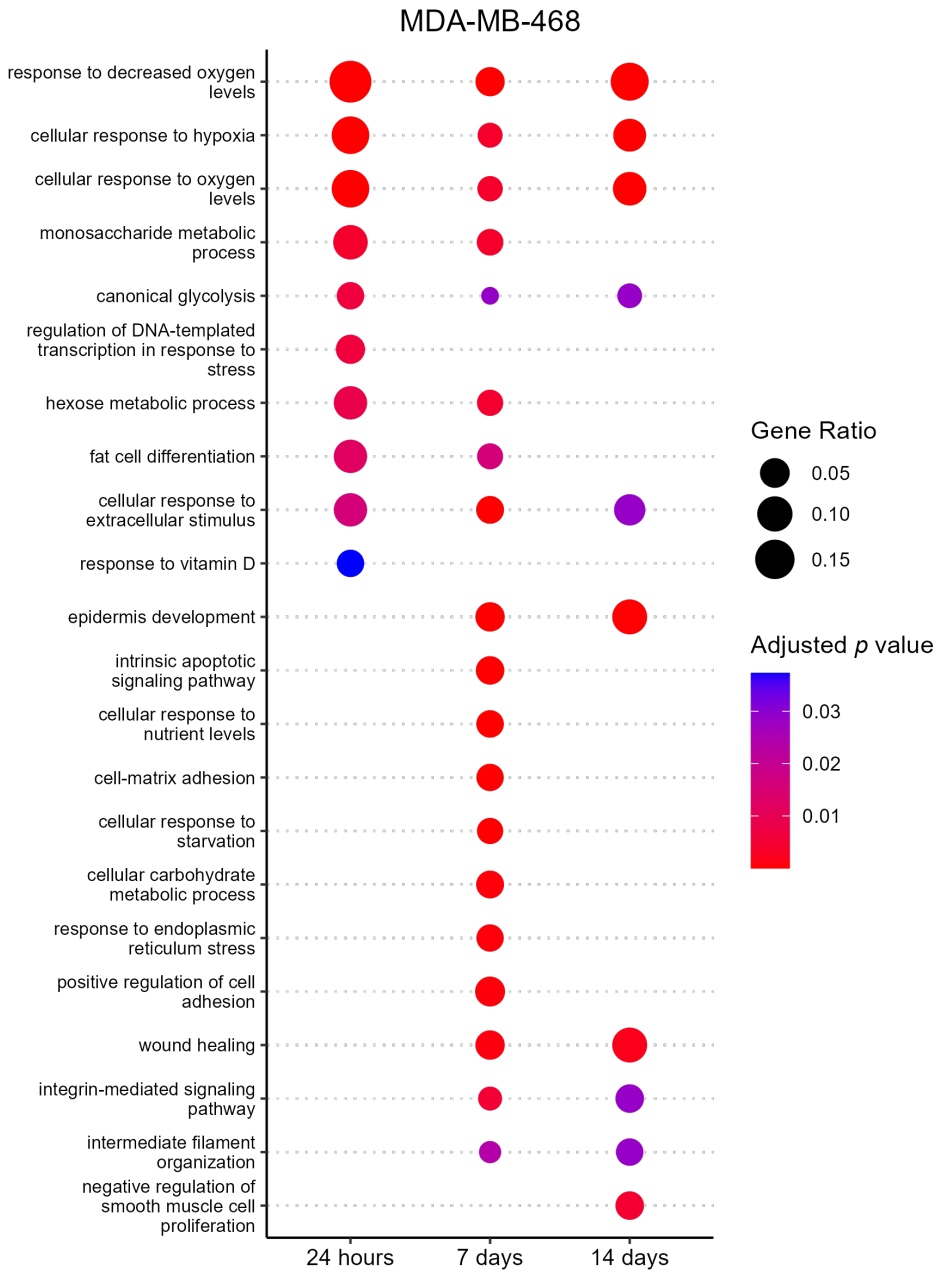

Figure 35: Over-represented pathways in acute and chronic hypoxia in MDA-MB-468 cells. Genes that are significantly (adjusted p value $\leqslant 0.05$) differentially expressed at each hypoxia timepoint are used in over-representation analysis. Pathways which are significantly (adjusted p value $\leqslant 0.05$) enriched are shown. The Gene Ratio refers to the number of differentially expressed genes found in the pathway, divided by the total number of differentially expressed genes in that condition.

To assess whether these genes were relevant in tumours and not just cell lines, I investigated the expression of genes in Table 2 in The Cancer Genome Atlas (TCGA) (Figure S3). After separating samples from the basal subtype according to their hypoxia score (Section 4.2.9), I assessed the expression of each the genes in Table 2. Twelve of the chronic hypoxia specific genes had significantly (adjusted p value $\leqslant 0.05$) higher expression in tumours with a hypoxia score in the upper tertile (Figure S3). These genes were: *CDCP1*, *CDKN1A*, *CPA4*, *EFNA1*, *EGLN3*, *GJB3*, *GPRC51A*, *KRT34*, *LAMC2*, *MTX1*, *PLOD2*, and *UPP1*. *ANKRD37* was also significantly upregulated in tumours with a high hypoxia score, however it is found within the Buffa hypoxia gene signature which was used to calculate the hypoxia scores so is discounted. Unexpectedly, six of the genes in Table 2 were significantly downregulated in tumours with a higher hypoxia score (*ETS1*, *ID4*, *IGFBP5*, *IL7R*, *PPP1R3E* and *RORA*), in contrast to the *in vitro* results. This could be caused by differences in the type of hypoxia experienced in the tumour. The hypoxia signature, which is already a proxy for true hypoxia measurement, is comprised of classical hypoxia genes which means cells experiencing acute, or intermittent, hypoxia will also be assigned a high hypoxia score. This could drown out the signal of tumours which are experiencing chronic hypoxia. Similarly, samples from the TCGA will also contains elements of stroma which could also drown out the signal from hypoxic tumour cells.

4.3.4 Vimentin levels are increased after exposure to chronic hypoxia

Given the RNA-seq results suggested processes such as EMT, migration, and adhesion were altered under chronic hypoxia, I attempted to functionally validate these changes *in vitro*. First, I investigated the expression of two key markers of EMT, E-cadherin and vimentin. E-cadherin is highly expressed in epithelial cells, while vimentin is expressed by mesenchymal cells. I exposed cells to 14 days of hypoxia and assessed the levels of E-cadherin and vimentin by western blot. Although there was some depletion of E-cadherin in MDA-MB-468 cells exposed to chronic hypoxia, in general the expression of E-cadherin did not change (Figures 36 and 37). As expected based on their subtypes, MCF7 cells express E-cadherin to a high level, while MDA-MB-231 cells do not express E-cadherin at all.

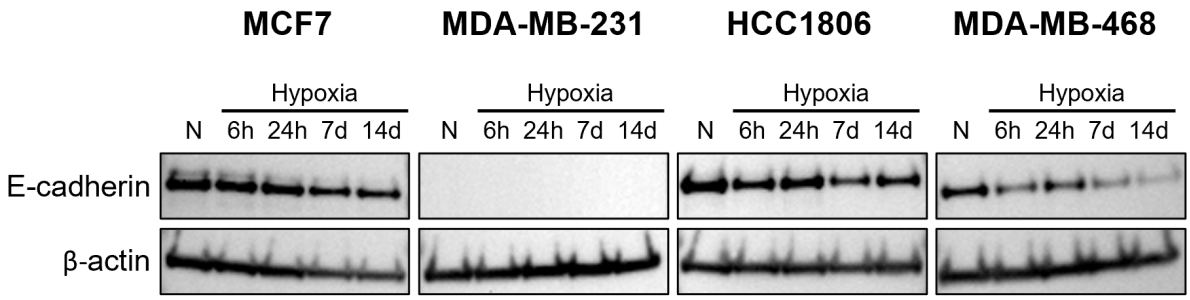

Figure 36: Expression of E-cadherin under chronic hypoxia. Expression of E-cadherin across the four cell lines exposed to hypoxia for 6 hours, 24 hours, 7 days, and 14 days. E-cadherin was not detected in MDA-MB-231 cells.

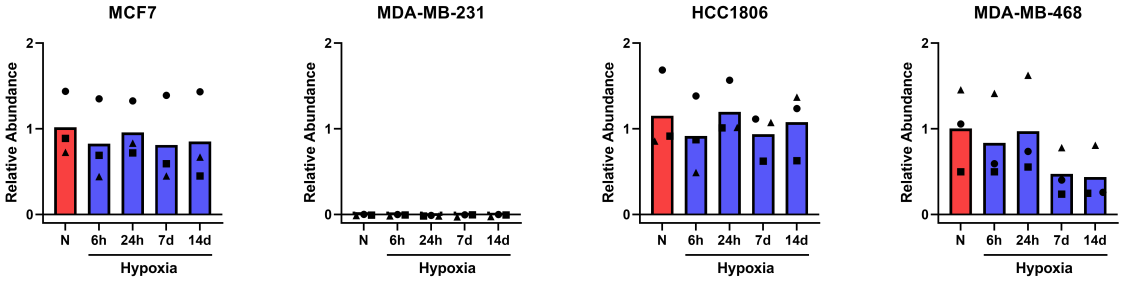

Figure 37: E-cadherin western blot quantification. Relative levels of E-cadherin compared to β-actin measured by western blot. There were no significant differences in E-cadherin levels between normoxia and any hypoxia timepoint. The symbol shapes represent different biological replicates. Significance was tested using one-way ANOVA with repeated measures and corrected for multiple testing using Dunnett's method *(n = 3)*.

In contrast, levels of vimentin did change after exposure to hypoxia (Figure 38) in line with the RNA-seq results which identified *VIM* as a chronic hypoxia specific gene. Vimentin levels were significantly upregulated after 7 days of hypoxia exposure in HCC1806 cells, and after 7 and 14 days in MDA-MB-468 cells (Figure 39). MDA-MB-231 cells also appear to increase their vimentin expression in line with hypoxia exposure time, but this result is not significant. Vimentin expression was not detected in MCF7 cells at any timepoint. The changes in vimentin expression indicate that chronic hypoxia exposure pushes cell towards to more mesenchymal phenotype, though the response is cell line dependent.

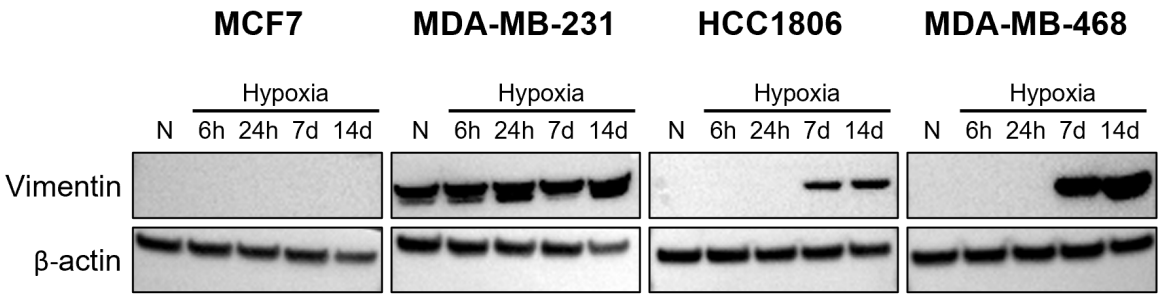

Figure 38: Expression of vimentin under chronic hypoxia. Expression of vimentin across the four cell lines exposed to hypoxia for 6 hours, 24 hours, 7 days, and 14 days. MDA-MB-231 cells have consistently high levels of vimentin, and a lower exposure time had to be used for imaging these membranes. This means that the MDA-MB-231 vimentin levels cannot be compared to the other cell lines. Vimentin was not detected in MCF7 cells at any timepoint.

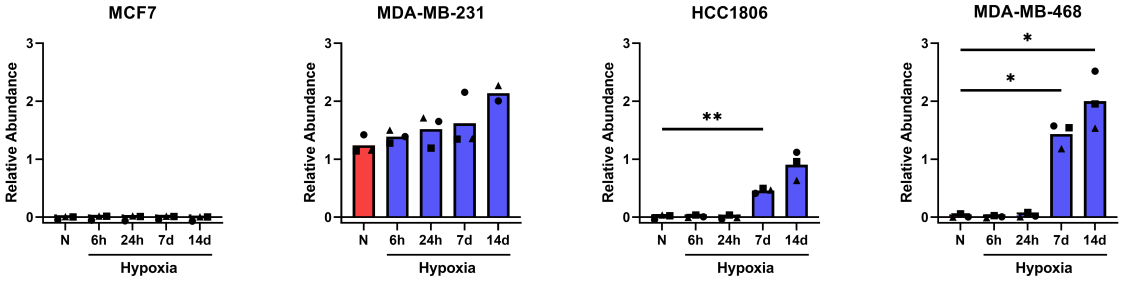

Figure 39: Vimentin western blot quantification. Relative levels of vimentin compared to β-actin measured by western blot. The symbol shape represents the biological replicate. Significance was tested using one-way ANOVA with repeated measures and corrected for multiple testing using Dunnett's method *(n = 3)* (* $p \leqslant 0.05$, ** $p \leqslant 0.01$).

4.3.5 Wound healing ability did not increase after exposure to chronic hypoxia

As EMT usually coincides with increased cell motility, and genes/pathways associated with wound healing were affected in the RNA-seq data, I conducted a wound healing experiment to investigate changes in the migratory potential of cells exposed to chronic hypoxia. The wound healing capacity of cells that had been exposed to hypoxia for 14 days was compared to passage matched normoxic controls. There was no increase in wound healing ability in any of the cell lines after hypoxia exposure (Figure 40). MDA-MB-468 cells did not migrate into the wound under any circumstances, exhibiting no cell motility. Though the extent of the effect varied, cell lines generally had a reduced wound healing capacity after exposure to hypoxia.

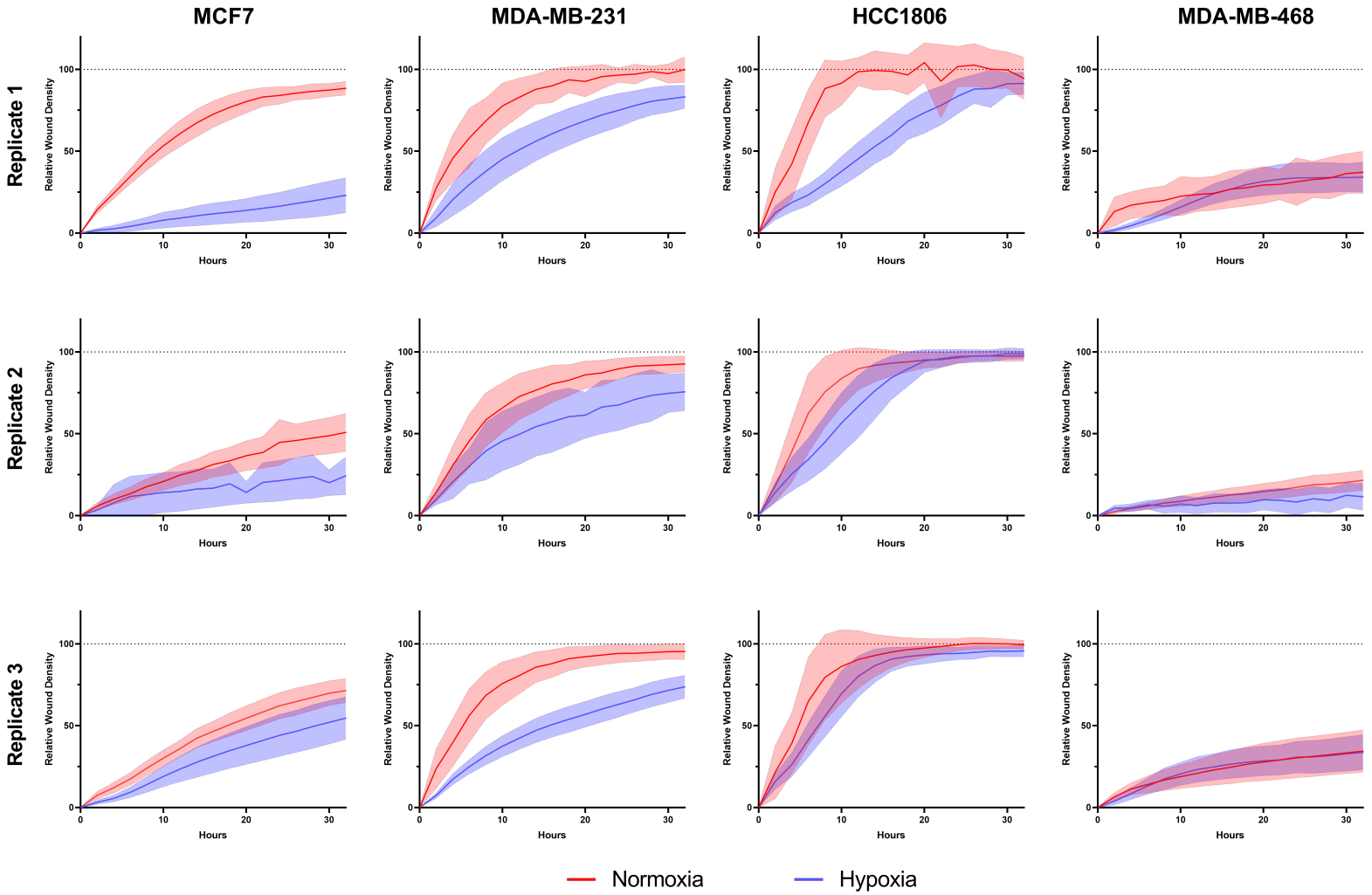

Figure 40: Wound healing. The mean relative wound density is shown over time with associated standard deviations. Hypoxic cells had been cultured in hypoxia for 14 days prior to the experiment, but were transferred to a normoxic environment for monitoring of wound density. Each of the three biological replicates are shown separately, and each contains a minimum of six technical replicates.

4.3.6 Cell adhesion was not affected by exposure to chronic hypoxia

The RNA-seq results implied changes in adhesion were taking place after exposure to chronic hypoxia. This was also reflected in observations during cell handling, whereby cells had a greater tendency to lift from the cell surface under hypoxia. To investigate cell adhesion, an assay was used which measures the rate at which cells adhere to the cell surface after seeding. Despite observational changes, there were no significant differences in the adhesion rate of cells exposed to hypoxia for 14 days and the normoxic controls in any of the cell lines tested (Figure 41).

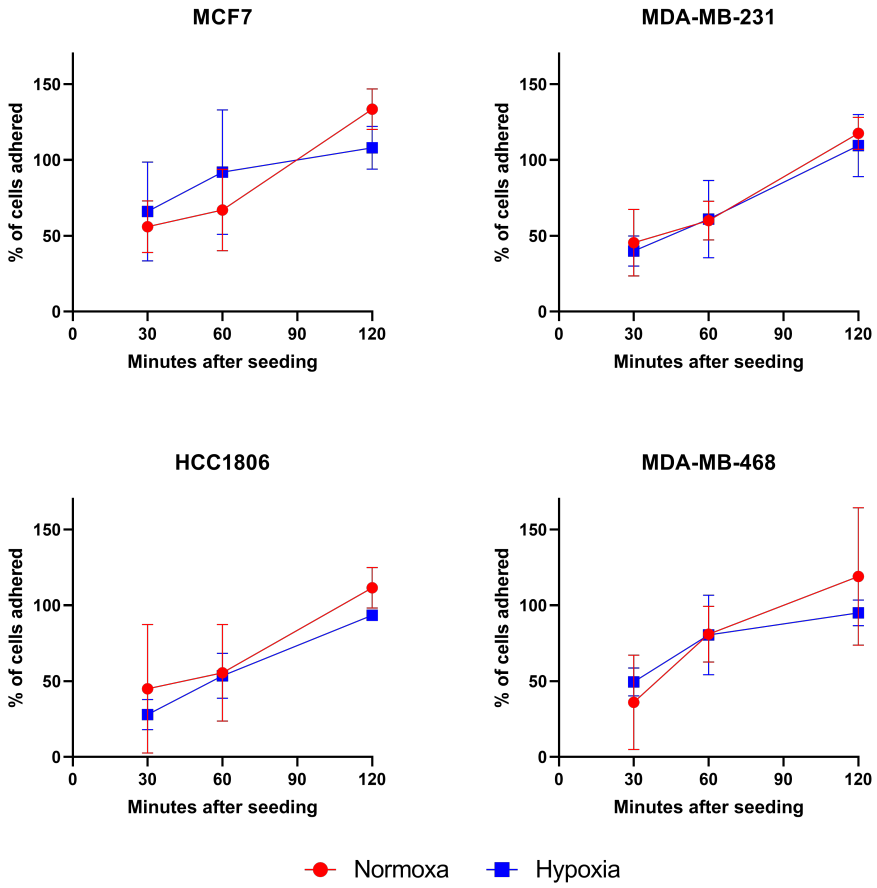

Figure 41: Cell adhesion. There were no differences in the rate of cell adhesion between cells cultured under hypoxia for 14 days and the normoxic controls *(n = 2)*. Error bars show standard deviation.

4.3.7 AP-1 is the most active EMT associated transcription factor under chronic hypoxia

Given the unusual results of the functional experiments, I returned to examine the RNA-seq data for expression of EMT associated transcription factors (EMT-TFs). Despite evidence that HIF1 directly upregulates TWIST via a HRE [145], significant TWIST upregulation was not observed in any cell line (Figure 42). In general, there were few consistent differences observed in the expression of the main EMT-TFs. *SNAI1* and *SNAI2* appear to be the most responsive, with one of the two being upregulated under hypoxia in each cell line, however these results are not significant. The only EMT-TF which was consistently upregulated was activator protein 1 (AP-1), which is a heterodimer comprised largely of genes from the Fos and Jun families. Although not considered a canonical EMT-TF, AP-1 has been shown to promote EMT in several studies, including in mouse mammary cells [80, 146]. Although there remains some cell line variation (i.e. *FOSL1* is most upregulated in MDA-MB-468 cells, whereas *FOS* and *FOSB* are in the other cell lines), the general pattern of AP-1 components being upregulated under hypoxia is consistent across the four cell lines.

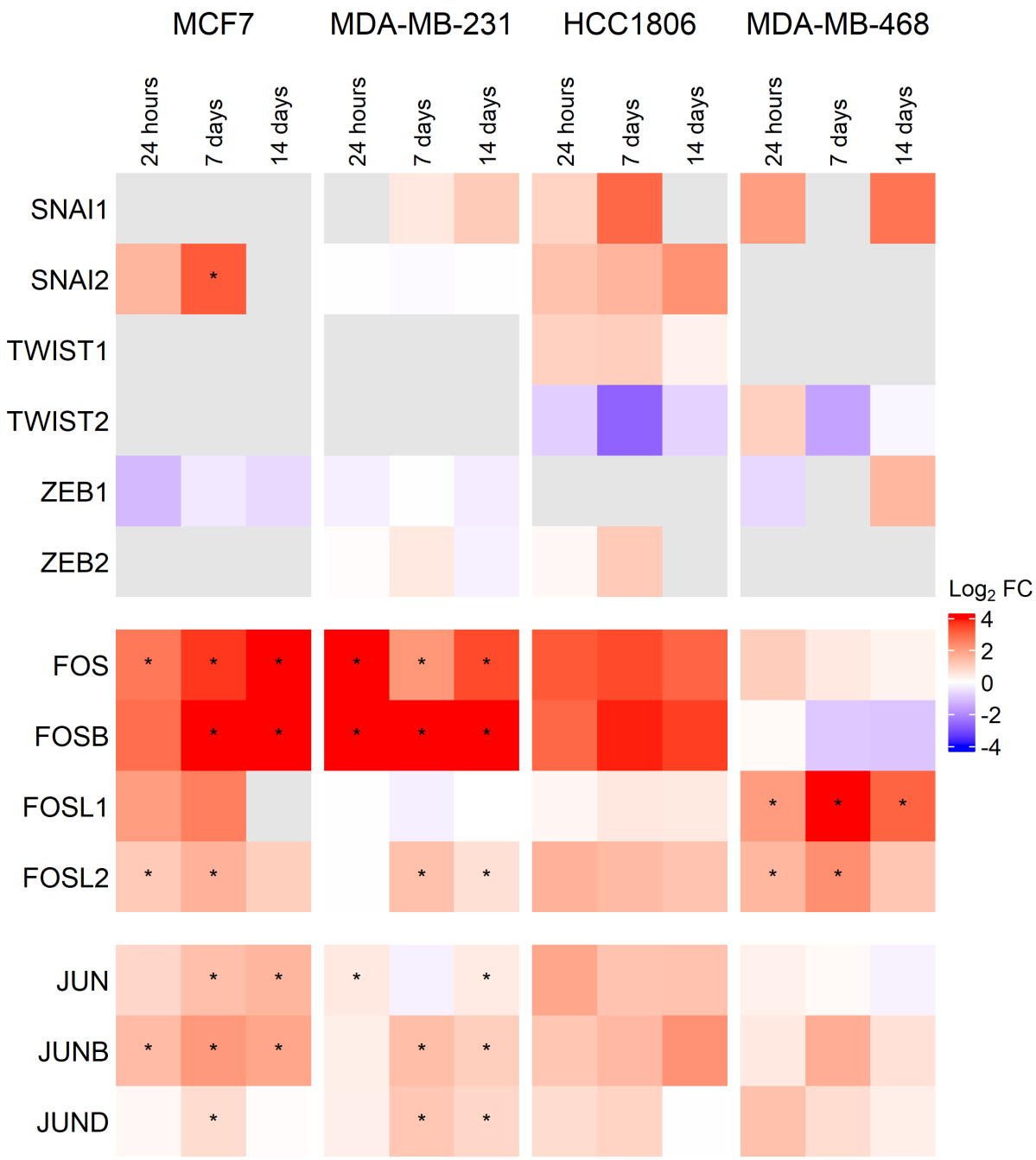

Figure 42: Expression of EMT associated transcription factors. Heatmaps showing the expression of EMT associated transcription factors (EMT-TFs) in each hypoxia timepoint compared to normoxia. Asterisks denote genes that were significantly (adjusted p value $\leqslant 0.05$) differentially regulated in that condition. $Log_2FC = log_2$ fold change.

4.3.8 miRNAs which repress EMT are active in MCF7 and MDA-MB-231 cells

Given that most EMT-TFs did not display clear patterns, I analysed the miRNA expression profiles of the cells to investigate whether this could explain the unusual patterns in the data. In general, there were very few significantly differentially expressed miRNAs in HCC1806 and MDA-MB-468 cells. One very notable exception is miR-100-5p, which was significantly (adjusted p value $\leqslant 0.05$) upregulated in MDA-MB-468 cells under chronic hypoxia (Figure 43). This miRNA has been shown to induce EMT in breast cancer cells (i.e. decrease E-cadherin while increasing vimentin expression) yet simultaneously reduce invasion and motility [147].

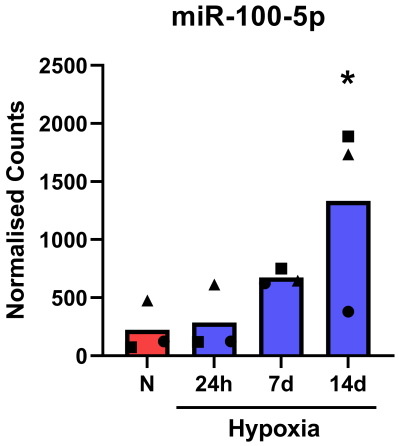

Figure 43: Expression of miR-100-5p in MDA-MB-468 cells. Expression of miR-100-5p is significantly (adjusted p value $\leqslant 0.05$) upregulated after 14 days of hypoxia exposure in MDA-MB-468 cells.

In mammary cells, AP-1 was shown to drive expression of miR-221/222 [146], miRNAs which have been shown to promote EMT in breast cancer [148]. This does not agree with my data, whereby components of the AP-1 complex are upregulated but miR-221/222 are significantly downregulated in several hypoxia comparisons and never upregulated (Figures S4 and S5).

In the case of miRNA regulation under hypoxia, there is a surprising amount of overlap between MDA-MB-231, the most mesenchymal of the cell lines, and MCF7, the most epithelial. Additionally, the miRNAs which are differentially regulated generally behave in a way that would push cells towards a more epithelial phenotype. For example, miR-16-5p is a tumour suppressor miRNA and exogenous expression was shown to drive cells towards an epithelial phenotype [149, 150]. This miRNA is significantly (adjusted p value $\leqslant 0.05$) upregulated after

14 days of hypoxia in both MCF7 and MDA-MB-231 cells, and may explain why they fail to show any markers of EMT (Figure 44).

miR-199b-3p is also upregulated in chronic hypoxia in both MCF7 and MDA-MB-231 cells (Figure 44). Again, this miRNA is generally tumour suppressive, and is able to target Snail to prevent EMT [150, 151]. This could counter the increase in *SNAI1* mRNA seen in MDA-MB-231 cells under chronic hypoxia (Figure 42). In MCF7 cells, the increase in Slug (*SNAI2*) (Figure 42), could be countered by increased expression of its targeting miRNA, miR-203a-3p [152], again suggesting an inhibition of EMT. miR-200c-3p exists in a negative feedback loop with Zeb [153, 154] and is associated with the epithelial phenotype. Again, both MCF7 and MDA-MB-231 upregulate this miRNA under chronic hypoxia (Figure 44). As Zeb is a direct target of miR-200c-3p, the slight decrease seen in *ZEB1* mRNA (Figure 42) could be explained by increased miR-200c-3p activity. None of these miRNAs are differentially regulated in HCC1806 or MDA-MB-468 cells (data not shown).

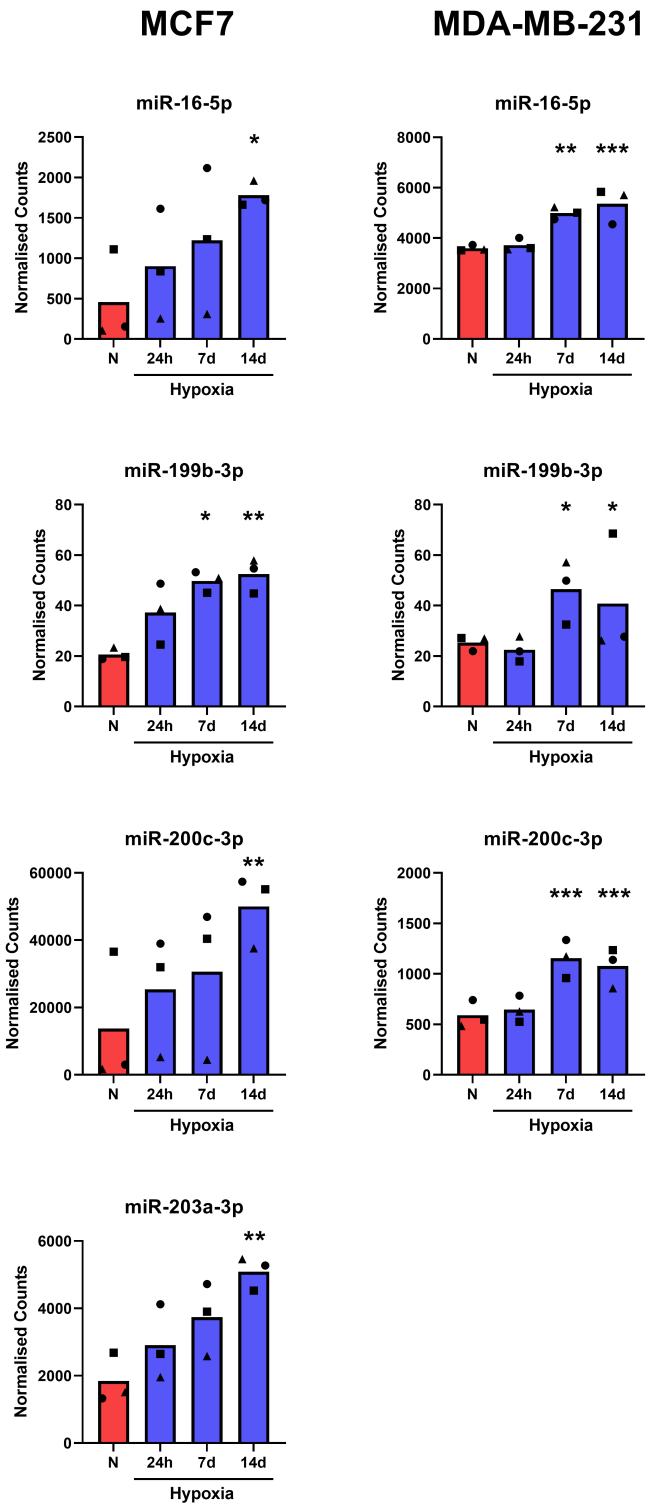

Figure 44: Expression of EMT associated miRNAs in MCF7 and MDA-MB-231 cells. Expression of miRNAs associated with inhibition of EMT are upregulated under chronic hypoxia in MCF7 and MDA-MB-231 cells (* adjusted p value $\leqslant 0.05$, ** adjusted p value $\leqslant 0.01$, *** adjusted p value $\leqslant 0.001$.)

Together, these data support the hypothesis that MCF7 cells are not undergoing EMT in response to chronic hypoxia, as these cells do not show EMT associated changes at the mRNA, miRNA, protein (vimentin and E-cadherin) or functional level. Interestingly, the data also support a hypothesis that MDA-MB-231 cells are undergoing MET, thus pushing the cells into a hybrid E/M state. This state has been suggested to provide stress resistance [80, 155, 156] so may aid adaptation to chronic hypoxia. This would explain why there are minimal changes in pathway analysis (i.e. only changes in adhesion) and also why characteristics associated with the mesenchymal phenotype are not increased. This puts MCF7 and MDA-MB-231 cells in contrast to HCC1806 and MDA-MB-468 cells, who display EMT associated changes at the mRNA level, do not upregulate EMT inhibiting miRNAs, and increase their vimentin expression at the protein level. However, it does not explain why the expected functional changes did not occur in the wound healing and adhesion assays.

4.4 Discussion

Extensive RNA-seq experiments have been conducted on cells exposed to acute hypoxia, which means a core set of commonly regulated hypoxia genes is well established. However, few studies have conducted RNA-sequencing on cells exposed to hypoxia for more than 72 hours. This means the chronic response to hypoxia is largely uncharacterised. My results show that the core, classical hypoxia response is maintained for at least 14 days under consistent exposure to 1% O_2. The expression level of the core hypoxia response genes does not appear to change between acute (24 hours) and chronic (7 days and 14 days) hypoxia exposure, suggesting the intensity of the hypoxic response remains consistent over at least two weeks (Figures 24 to 26). This hypoxia response is orchestrated by presence of HIF, which is maintained in at least one of its subtypes across all four cell lines for the duration of the experiment. As observed by other researchers, a HIF switch occurs whereby the expression of HIF1 in the acute timepoints is replaced by expression of HIF2 in the later timepoints [25]. However, this does not occur in every cell line as MCF7 cells fail to upregulate HIF2 at any timepoint studied (Figures 19 and 21). This highlights cell line variability in the HIF response and shows that the switching of HIF subtypes is not a universal mechanism.

Of the existing studies into the transcriptomic response to chronic hypoxia, one uses a model that is unsuitable for comparison. Chen et al. [110] use a model of clear cell renal cell carcinoma which is VHL mutant, does not express HIF1, and constitutively expresses HIF2. This means that their normoxic cells have a pseudohypoxic phenotype due to the active HIF2. Despite this, there is some overlap with my results. The cells in their chronic hypoxia experiment (1% O_2 for 3 months) did not show differences in migration when measured by wound healing assay, but they did have greater migratory capacity when measured by transwell assay. The researchers attribute this to the chemoattractant used which provides a stimulus for migration. This indicates that the wound healing assay used in my experiments may not be sufficient to conclude there were no changes in migratory capacity.

Jarman et al. [111] exposed MCF7 cells to hypoxia (0.5% O_2) for 10 weeks. In line with my data, they show high levels of overlap between HIF targets that are upregulated in acute and chronic hypoxia. They showed little to no HIF2 expression in wild type MCF7 cells as in my data, but did show an increase in HIF2 when they overexpressed HER2 in these cells.

Danielsson et al. [112] also observed differences in the response to chronic hypoxia (3% O_2 for six passages) in their isogenic cell lines, showing how important the existing transcriptome of the cell is for influencing the hypoxic response. This could explain why there is generally little overlap in chronic hypoxia response genes between my four cell lines.

When analysing chronic hypoxia response genes, I focused on genes which were relevant in at least two of the four cell lines. This increases the applicability of the results, as they are more likely to be consistent between different patients if they are relevant across different cell lines. Of the genes that were chronic hypoxia responsive, there were a large number of genes whose functions were implicated in adhesion, motility, and EMT. EMT has already been associated with hypoxia, with many EMT-TF genes (e.g. *SNAI1*, *TWIST1* and *ZEB1*) containing HREs in their promoter regions [82]. Additionally, TGFβ, a driver of EMT, is stimulated by hypoxia. Furuta et al. [157] showed that exposure of lung cancer cells to 10 days of hypoxia was able to drive expression of TGFβ. There is some evidence that TGFβ exists in a positive feedback loop with HIF through TGFβ driven inhibition of PHD2 expression [158, 159]. This would augment pro-EMT signalling specifically under chronic hypoxic conditions.

Given the RNA-seq data implied changes in EMT, adhesion, and motility, validation experiments were conducted focusing on these cell characteristics. The results of these experiments are interesting and in some cases contradictory. Expression of two main markers of the epithelial and mesenchymal phenotypes, E-cadherin and vimentin, was assessed at the protein level. The results of this experiment implied that EMT was occurring in at least HCC1806 and MDA-MB-468 cells as the expression of vimentin significantly increased upon exposure to chronic hypoxia (Figures 38 and 39). In the case of MDA-MB-468 cells, there was also a concomitant decrease in E-cadherin at the chronic hypoxia timepoints (Figures 36 and 37). The same effect was not seen in MDA-MB-231 cells, however these cells are already mesenchymal and express very high levels of vimentin at baseline. Despite existing in an epithelial state, MCF7 cells did not display a change towards the mesenchymal phenotype in this assay, with no decrease in E-cadherin or detectable expression of vimentin at any timepoint. The genetic background of the cell has been reported to affect a cell's journey through EMT [160]. For example, wild type p53 was shown to prevent EMT through its regulation of miR-200c, while mutant p53 promotes EMT [161, 162]. MCF7 cells are the only cells in my panel which express wild type p53 (MDA-MB-231 and MDA-MB-468 are p53 mutant,

HCC1806 are p53 null), which could at least partially explain the lack of evident EMT in these cells. It is possible that in order for hypoxia to drive EMT, cells must already be primed to enact such changes. In this way, the aggressiveness of the TNBC subtype, as well as non-functioning p53, allows this process to occur in HCC1806 and MDA-MB-468 cells, but not in MCF7 cells.

As the EMT marker vimentin was upregulated in HCC1806 and MDA-MB-468 cells, I expected to also see changes in migration and adhesion. However, the wound healing experiment showed no increase in migration under chronic hypoxia in any of the cell lines tested (Figure 40). MDA-MB-468 failed to exhibit any motility in the assay, whereas hypoxic HCC1806 cells migrated to cover the wound but at a slower rate than their normoxic counterparts. The lack of increased motility could be caused by energy conserving mechanisms. It possible that, though the cells have a greater *potential* to migrate, they do not act on this in order to preserve ATP. Although cells had been cultured in hypoxia for 14 days prior to the wound healing assay, the cells were removed after creation of the wound to be assessed in the Incucyte, which is contained in an incubator at atmospheric oxygen levels. The stress caused by this re-oxygenation may have reduced migratory capacity which otherwise would have been increased in the hypoxic setting. Wound healing experiments which maintain cells in hypoxia for the duration of the experiment could determine whether re-oxygenation affected my results. Alternatively, a transwell assay could be used. This would better recapitulate the tumour state, whereby cells in chronic hypoxia, which would likely be deprived of nutrients as well as oxygen, would be migrating towards areas of greater nutrient density.

The adhesion assay did not reveal differences in adhesion in any of the cell lines (Figure 41), despite RNA-seq data suggesting changes in adhesion were taking place. Again, different assays may be required to uncover any differences, as observation during cell culture strongly suggested that changes in adhesion were occurring. For example, it was difficult to conduct the wound healing assay on MDA-MB-468 cells as the scratching of the wound often caused a large sheet of cells to be removed, rather than only cells which contacted the wound maker. This strongly implies enhanced cell-cell adhesion, with a possibility of reduced cell adhesion to the ECM. Similarly, MDA-MB-231 cells were observed to release from the cell surface much more easily during media changes and during wound formation, again indicating weaker connections to the ECM. These changes were not reflected in the adhesion assay conducted. However, this assay specifically captures information on how quickly cells adhere to the cell

surface after seeding rather than the strength or abundance of adhesions. Early after seeding it is likely that all cell connections to the ECM are weaker as cells have had less time to establish multiple connections and connections of different types. Immunofluorescent staining of cell-cell and cell-ECM adhesion molecules could shed light on the differences observed.

There is mounting evidence that cells do not exist in a polar state of either epithelial or mesenchymal, but rather there are steady intermediate, or hybrid, states between these two phenotypes [80, 155, 160, 163]. Of the breast cancer subtypes, TNBC is most associated with the hybrid state [163]. These E/M hybrid states have been purported to be more plastic, thus allowing cells to adapt to different micro-environmental challenges [80, 155, 156]. Tagliazucchi et al. [160] found that tumours which fell into the E/M hybrid state were more likely to have higher levels of hypoxia. This also correlated with the expression of stemness marker CD44, a marker which has been identified in our lab by single-cell RNA-seq as being increased after 7 days of hypoxia in MDA-MB-231 cells (data not published). Interestingly, Tagliazucchi et al. [160] observed variation in hypoxia driven EMT, concluding that although hypoxia predisposes cells to begin EMT, it is not absolutely determined and is highly context specific.

Different research groups have put forward ways of identifying cells in a E/M hybrid state. For example, Tagliazucchi et al. [160] showed that the E/M hybrid state was distinguishable from fully transformed mesenchymal cells by genes involved in cytoskeleton remodelling and cell adhesion. This is in line with my data, where genes differentially regulated under chronic hypoxia are heavily enriched with structural genes and those involved in adhesion (Figures 34 and 35). The OVOL transcription factors are also thought to maintain the hybrid E/M state [163], however, my data did not display clear patterns with regards to these genes.

Another suggested marker of the E/M hybrid state is P-cadherin [164], a protein which is expressed by collectively migrating cells [163]. In some settings (e.g. development, wound healing) cells are able to migrate without losing all of their cell-cell attachments, allowing them to move as a collective [81, 165]. This has also been observed in certain cancer types such as squamous cell carcinomas [71]. MDA-MB-468 express P-cadherin at higher levels under chronic hypoxia, with expression being significantly (adjusted p value $\leqslant 0.05$) upregulated after 7 days of hypoxia exposure. This increase in P-cadherin expression supports the hypothesis that MDA-MB-468 cells enter an E/M hybrid state during exposure to chronic hypoxia. It also

means they may have the capacity to migrate as a collective, but may fail to do so in the wound healing assay due to lack of chemoattractant stimuli.

Jolly et al. [163] discuss another endpoint of EMT, namely the amoeboid endpoint. Spindle shaped mesenchymal cells typically have strong cell-ECM adhesion, whereas amoeboid cells retain a round shape and have lower levels of ECM adhesion. Amoeboid cells are typically unable to degrade ECM and are less invasive, but they still maintain motile capabilities [166]. This method of migration has been suggested as the preferred method under hypoxia for reasons of energy conservation [166]. As MDA-MB-468 cells did not visibly change their phenotype towards spindle shaped, but appeared to lose some cell-ECM adhesion, it could be that they are on an EMT trajectory towards the ameboid phenotype.

Analysis of EMT-TF expression did not yield clear results in my data. The only common trend seen was an increase in expression of Fos family genes. Proteins from the Fos and Jun families dimerise to form the AP-1 transcription factor, a known driver of EMT [80]. High expression of AP-1 components has been associated with poor survival in cancer, and exogenous expression of AP-1 alone was shown to induce EMT, even in non-cancerous mammary cells, through AP-1 driven transcription of *ZEB1* and *ZEB2* [146, 167].

Brown et al. [156] conducted in depth characterisation of SUM149PT breast cancer cells at different stages of EMT. Their results show that components of AP-1 were activated at the start of EMT, and their expression was maintained through all of the steady E/M hybrid states and into the mesenchymal phenotype. The consistency of this gene expression pattern may explain why it is the only universally active EMT-TF across my four cell lines. In accordance with my data and others, they were unable to identify clear patterns in the expression of other EMT-TFs at the mRNA level [81, 156]. Their experiments showed that cells in the E/M hybrid states had greater migratory and invasive capacity, which does not agree with my data. However, they did not use hypoxia in their model.

The mesenchymal phenotype plays a role in stress resistance. Using lineage tracing, an *in vivo* study showed that breast cancer cells which metastasised to the lung had not previously undergone EMT, disputing the hypothesis that cells undergo EMT, metastasise, then undergo MET during colonisation [168]. This is in line with another *in vivo* study which showed that

knockout of either Twist or Snail did not affect invasion or metastasis in a mouse model of pancreatic cancer [169]. However, both of these studies suggested that mesenchymal cells were more resistant to chemotherapy. In Fisher et al.'s study, [168], a small number of mesenchymal cells would arise in the primary tumour, but not metastasise. When the mice were treated using chemotherapy, the mesenchymal cells would survive and subsequently go on to seed metastasis. This was enabled by a combination of lower growth rates and upregulation of genes involved in chemoresistance in the mesenchymal state. Zheng et al. [169] showed that Twist or Snail knockout increased cell proliferation and made cells more susceptible to chemotherapy. These findings also support research by Chen et al. [147] who showed that changes in molecular EMT markers were not always associated with increases in migration, invasion, and tumourigenicity. These studies pose a possible new role for chronic hypoxia induced EMT which does not drive migration, but does increase resistance to microenvironmental stressors.

In conclusion, the response to chronic hypoxia is highly variable between different cell types. Induction of an atypical EMT appears in two of the four cell lines. The existence of these cells in an E/M hybrid state may explain why there are discrepancies between the undoubtable expression of some EMT markers, along with supporting gene expression changes, but without the expected changes in cell behaviour. The type of EMT induced in these cases may fulfil the purpose of increasing stress resistance in the cells rather than increasing invasion/migration. Just as mesenchymal cells were resistant to chemotherapy due to lower proliferation rates [168, 169], they may also be more resistant to hypoxia induced stress. Reduced growth rates would preserve energy in the hypoxic tumour microenvironment enabling cell survival.

5 Gene essentiality under chronic hypoxia

5.1 Introduction

Having identified that chronic hypoxia induces transcriptional changes associated with EMT, adhesion, and motility, I questioned whether these changes were essential for cell survival under hypoxia. If these processes constitute an adaptation to the chronic hypoxia setting, then inhibiting these changes may reduce cell growth or survival in this context. Tagliazucchi et al. [160] demonstrated that some genes associated with EMT, or the E/M hybrid state, were important for cell viability. EMT is already associated with chemoresistance [81], a property which enables cell survival under stress conditions, so it is possible that this process will also contribute to survival under hypoxic stress. Furthermore, all four cell lines maintained the classical transcriptional response to hypoxia for the full 14 day time course. The longevity of this response means cells could be reliant on either the master HIF transcription factors, or some of their downstream targets, for survival under chronic hypoxia.

CRISPR (clustered regularly interspaced short palindromic repeats) originated as an adaptive immune system in bacteria, but has since been modified to produce a powerful gene editing tool. Using CRISPR to knockout genes has become increasingly common due to its specificity and ease of use [170]. Only two components are required for gene knockout, both of which can easily be introduced to most eukaryotic cells (Figure 45). The first component is the Cas9 enzyme, an RNA-guided DNA endonuclease which causes double-strand breaks in DNA. The second component is a single guide RNA (sgRNA). The sgRNA contains a 20 nucleotide spacer sequence which directs Cas9 binding, and therefore the location of the double-strand break, in a sequence dependent manner [171]. Since this spacer region is modifiable, Cas9 can be targeted to almost anywhere in the genome. The cell repairs the double-strand break, typically using non-homologous end joining. This process often results in small insertion/deletion events surrounding the break site [171]. In a protein coding gene, this can cause a frame-shift event to occur. This results in either a premature stop codon, producing a truncated protein, or severe mutation of the final protein. In both cases, the target protein is likely to be non-functional. In this manner, genes may be selectively knocked out using CRISPR.

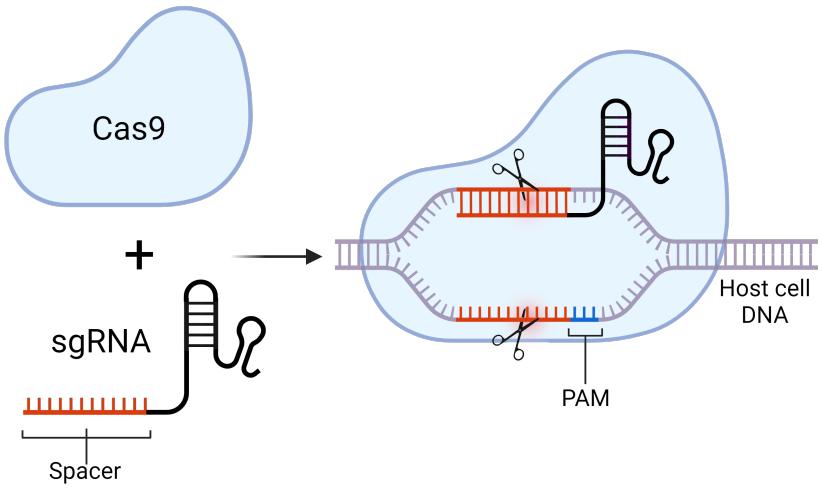

Figure 45: Components required for CRISPR. Only the Cas9 enzyme and a single-guide RNA are required to cause targeted double-strand breaks in host cell DNA. The spacer region targets Cas9 to a desired location in the genome based on sequence complementarity. A downstream three nucleotide protospacer adjacent motif (PAM) sequence must be present in the host genome. Figure created using BioRender.

CRISPR knockout screens utilise this technology to produce large scale loss-of-function experiments. These allow genes involved in a given process to be systematically identified. Screens are often genome-wide, targeting all protein-coding genes within an organism, although smaller, curated panels of genes may also be used. A simple CRISPR knockout screen involves synthesising a pool of sgRNAs which target the desired genes (Figure 46). These are packaged into lentiviruses which are used to infect a population of cells. The cells are either pre-engineered to express Cas9, or this is included in the lentiviral vector. In each cell, a different sgRNA will be expressed and therefore a different gene knocked out. The integration of the lentiviral genome into the host cell DNA means that sequencing of host cell DNA at a later stage can be used to identify which sgRNA was expressed in that cell. In pooled selection screens, cells are infected, then cultured under a chosen selective pressure for a period of time. During this time, some cells will have a survival advantage/disadvantage due to the gene knockout. This means the proportion of cells expressing each guide will change over time and in response to any selection pressures. At the end of the screen, the proportion of cells containing each guide is compared and this allows identification of genes which play a role in survival under the chosen selection pressure [171, 172].

The power of this technology was harnessed in this chapter to conduct genome-wide CRISPR knockout screens on three breast cancer cell lines (MDA-MB-231, MDA-MB-468, MCF7). Cells were cultured for 14 days in normoxia or hypoxia to allow genes which are involved in survival/adaptation to hypoxia to be identified.

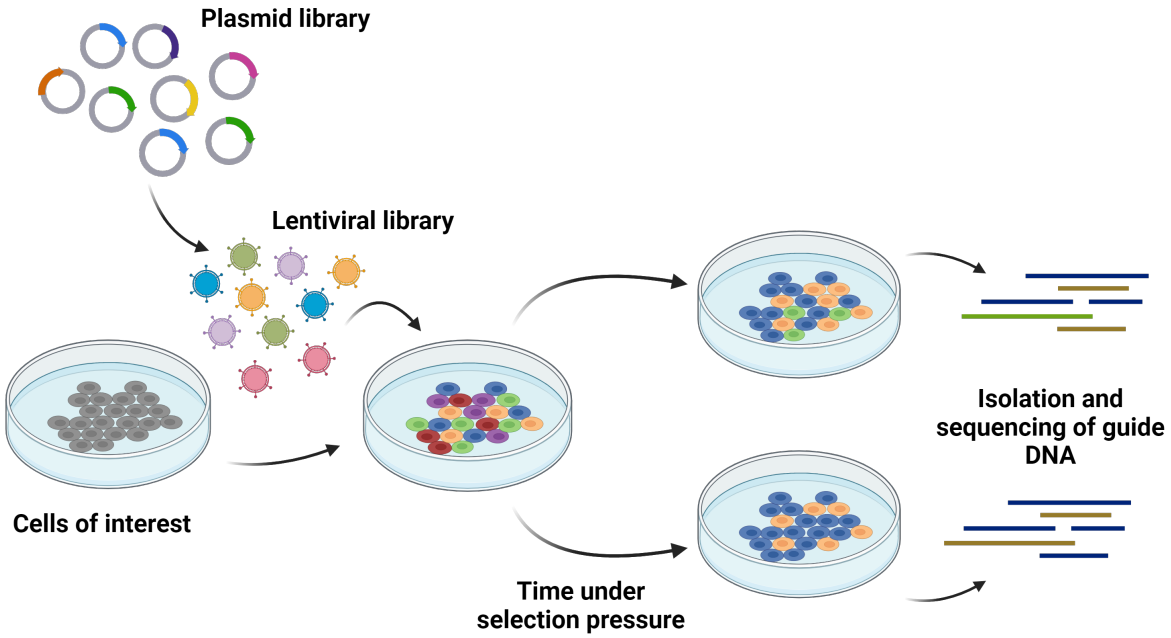

Figure 46: Steps involved in a CRISPR screen. The plasmid CRISPR library is packaged into lentiviruses and used to infect the cells of interest. After culturing cells under the desired selection pressure, genomic DNA is extracted. The guide sequence, which has integrated into the genomic DNA, is amplified and sequenced. Figure created using BioRender.

5.2 Methods

5.2.1 CRISPR screen

The laboratory work for the following CRISPR screens was a combined effort of postdocs in the Buffa lab (Dr Ana Miar Cuervo, Dr Helen Sheldon, Dr Giovanna Granata) who worked on the protocol and execution. A significant part of the laboratory work was contributed as a service by the High Throughput Screening group at the Target Discovery Institute (TDI), University of Oxford. At least fifty-two T175 flasks were seeded at the start of the experiment. The following day, cells were infected using the Toronto KnockOut Library v3 (TKOv3) at a multiplicity of infection (MOI) of 0.3. One flask was kept free of virus. Cells were incubated overnight. The following day the virus was removed and cells incubated overnight in fresh media. The next day, puromycin was added to all flasks and selection continued until all cells in the uninfected flask had died. When selection was complete, cells were trypsinised and pooled. At least $3.6x10^7$ cells were required to maintain 500x coverage. $4x10^7$ cells were removed and the cell pellet was stored at -80 ℃. This became the reference (T0) sample. The remaining cells were divided between the normoxia and hypoxia arms of the experiment. Media was changed at least every 48 hours in hypoxia. Cells were maintained in their assigned condition, and samples taken at day 3 (T3), day 7 (T7), day 10 (T10), and day 14 (T14) during passaging. Enough cells were re-seeded so as to maintain $\geqslant$ 500x coverage. The screen was conducted in biological duplicate on each of the following cell lines: MCF7, MDA-MB-231, and MDA-MB-468.

5.2.2 DNA extraction

$4x10^7$ cells were collected for genomic DNA extraction. DNA extraction was performed using the Wizard Genomic DNA Purification Kit and PureLink RNase A as per manufacturer's instructions. Cell pellets were thawed at room temperature for 5-10 minutes, then resuspended in 1.4 mL PBS by vortexing. 5 mL of Nuclei Lysis Solution was added, along with 32 µL of 20 mg/mL RNase A, to give a final concentration of 100 µg/mL. The solution was mixed by inversion then incubated at 37 ℃ for 15 minutes. 1.67 mL of Protein Precipitation Solution was added to the nuclear lysate and the solution was vortexed for 20 seconds. The lysate was centrifuged at 4500 *g* for 10 minutes at room temperature. The supernatant was transferred to a centrifuge tube containing 5 mL isopropanol and mixed by inversion. The solution was centrifuged at 4500 *g* for 5 minutes at room temperature. The supernatant was discarded

and 5 mL of room temperature 70% ethanol added to the DNA pellet. The solution was again centrifuged at 4500 *g* for 5 minutes at room temperature. The supernatant was discarded and the DNA pellet air dried for 10 minutes. 400 µL of Tris-EDTA (TE) buffer was added and the DNA allowed to dissolve at 65℃ for 1 hour. The DNA was quantified using both a NanoDrop Spectrophotometer and the Qubit dsDNA High Sensitivity Assay.

5.2.3 Library preparation

Two PCR reactions were performed first to enrich sgRNA regions in the genome (PCR1) and second to amplify sgRNAs and add adapter and index sequences. Primer sequences for PCR1 are shown in Table 3.

Table 3: PCR1 primer sequences - TKOv3

Name	Sequence
v2.1-F1	GAGGGCCTATTTCCCATGATTC
v2.1-R1	GTTGCGAAAAAGAACGTTCACGG

For PCR2, the forward primer utilised the Illumina TruSeq Universal Adapter (AATGATACG-GCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT) followed by one of nine different stagger sequences used to introduce complexity to the library (Table 4) [173]. The reverse primer contained a unique Illumina i7 index for each sample to allow pooling during sequencing.

Table 4: PCR2 stagger sequences - TKOv3

Name	Sequence
F01	TTCTTGTGGAAAGGACGAAACACCG
F02	ATTCTTGTGGAAAGGACGAAACACCG
F03	GATTCTTGTGGAAAGGACGAAACACCG
F04	CGATTCTTGTGGAAAGGACGAAACACCG
F05	TCGATTCTTGTGGAAAGGACGAAACACCG
F06	ATCGATTCTTGTGGAAAGGACGAAACACCG
F07	GATCGATTCTTGTGGAAAGGACGAAACACCG
F08	CGATCGATTCTTGTGGAAAGGACGAAACACCG
F09	ACGATCGATTCTTGTGGAAAGGACGAAACACCG

PCR1 reactions were set up as outlined below using a total of 100 µg genomic DNA. 3.5 µg of genomic DNA was used per 50 µL reaction.

25 µL	NEBNext Ultra II Q5 Master Mix (2X)
2.5 µL	PCR1 forward primer (10 µM)
2.5 µL	PCR1 reverse primer (10 µM)
3.5 µg	Genomic DNA
__ µL	DEPC treated water to make the total reaction volume 50 µL

PCR1 reactions were run on a thermocycler using the following settings:

Table 5: PCR1 settings - TKOv3

Step	Temperature	Time
Initial denaturing	98℃	30 seconds
Denaturation*	98℃	10 seconds
Annealing*	66℃	30 seconds
Extension*	72℃	15 seconds
Final extension	72℃	2 minutes
Hold	10℃	Infinite

*30 cycles

2 µL of PCR1 product was run on a 1% agarose gel to ensure the reaction yielded bands of the correct size (600 bp). All individual 50 µL reactions were pooled for PCR2. PCR2 reactions were set up as outlined below.

25 µL	NEBNext Ultra II Q5 Master Mix (2X)
2.5 µL	(F01-F09 pool) forward primer (10 µM)
2.5 µL	i7 reverse primer (10 µM)
5 µL	PCR1 product
15 µL	DEPC treated water to make the total reaction volume 50 µL

PCR2 reactions were run on a thermocycler using the following settings:

Table 6: PCR2 settings - TKOv3

Step	Temperature	Time
Initial denaturing	98℃	30 seconds
Denaturation*	98℃	10 seconds
Annealing*	55℃	30 seconds
Extension*	65℃	15 seconds
Final extension	65℃	5 minutes
Hold	10℃	Infinite

*10 cycles

The PCR2 product was run on a 2% agarose gel and the 200 bp band was excised and purified using a QIAquick Gel Extraction Kit. Products were quantified and the purity assessed using both a NanoDrop Spectrophotometer and the Qubit dsDNA High Sensitivity Assay. Samples were pooled and sequenced using a NovaSeq 6000 at the Oxford Genomics Centre, Oxford.

5.2.4 Processing of sequencing reads

Data were downloaded by file transfer protocol and md5sums checked to ensure files were not corrupted. FastQC (version 0.11.9) was used to assess the quality of the reads. Read 2 showed consistently low quality scores so was removed from the analysis. Read 1 was processed using cutadapt (version 2.10) to trim the adapter sequences and remove reads that were not 16-21 bases after trimming. The cutadapt options used were: `--discard-untrimmed -q 20 -m 16 -M 21 -g TCTTGTGGAAAGGACGAAACACC...GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGAGATCGGAAGAGCACACGTCTGAACTCCAGTCA`. Reads were aligned to the guide sequences using MAGeCK (version 0.5.9.2) using the `--trim-5 0,1` option. MAGeCK does not allow for mismatches so mapping was highly stringent. If contaminating guides were present (Section 5.3.1), they were manually removed at this point. Differential analysis was conducted using MAGeCK and *p* values were corrected for multiple testing internally using the Benjamini-Hochberg method.

5.2.5 GCN2 and p-GCN2 western blots

Duplicate NuPAGE 3-8% Tris-Acetate gels were prepared, and one was probed for GCN2 and the other for p-GCN2. 15 µL of Rainbow Molecular Weight Marker was added to the first well of the GCN2 gel, while 15 µL of Precision Plus Protein Dual Color Standards was added to the first well of the p-GCN2 gel. Remaining wells were loaded with 50 µg of protein (prepared using urea buffer as described in Section 2.3.1). Gels were run in 1X MES buffer. Protein was transferred using the Turbo transfer method (Section 2.3.8). The GCN2 membrane was blocked using 5% nonfat dried milk powder, while the p-GCN2 membrane was blocked using 5% BSA. The membranes were probed with primary antibodies targeting either GCN2 or p-GCN2, both at 1:1500 dilutions in their respective blocking buffer. The secondary antibody (goat anti-rabbit HRP) was diluted to 1:1500 in the respective blocking buffer for each membrane. Full details for probing and developing the membrane are found in Section 2.3.9.

5.2.6 Treatment of cells with GCN2i

A chemical inhibitor of GCN2 (GCN2i) [174] was purchased from Selleck Chemicals (GCN2iB, catalogue number S8929). The powdered drug was reconstituted in DMSO to produce a 10 mM solution and aliquots were stored at -80℃. A final GCN2i concentration of 0.5 µM was used. In accordance, the relevant controls were cultured in media with 0.005% DMSO.

5.2.7 Generation of GCN2 knockdown cells

$1.5x10^4$ cells were seeded per well of a 96 well plate and left to adhere overnight. Media did not contain P/S. MISSION lentiviral transduction particles were purchased from Sigma Aldrich. These lentiviral particles contain genes encoding short hairpin RNAs (shRNAs) which target GCN2 (Table 7).

Table 7: GCN2 shRNA sequences

Abbreviation	Gene Target	Clone ID	Target Sequence
sh850	GCN2	TRCN0000300850	GCCTAACTGGTGAAGAAGTAT
sh851	GCN2	TRCN0000300851	CCAAAGGTCTATCAAATGAAA

The lentiviral solutions were diluted in culture media (no P/S) containing 8 µg/mL polybrene to produce MOIs of 0.5, 1, 2, and 5. 100 µL of the relevant viral solution was added to the wells. The plate was centrifuged at 250 *g* for 2 minutes to bring the viruses in closer proximity to the cells. The viral solution was left on the cells overnight then removed and replaced with 200 µL of fresh culture media (no P/S). After 24 hours, all infected cells, along with the no infection control, were treated with 0.5 µg/mL of puromycin. After 24 hours, the puromycin concentration was increased to 1 µg/mL and selection continued for another 72 hours. Cells were expanded over subsequent weeks, and vials stored at -80℃.

To identify the MOI which produced the best knockdown, GCN2 protein levels were assessed by western blot. Samples were collected from confluent 6 well plates using RIPA buffer. 50 µg of protein was loaded per lane of a 4-12% Bis-Tris Gel and processed as per Section 5.2.5.

5.2.8 Incucyte growth assay

Cells (MCF7, 7.5x10^3 cells; MDA-MB-231, 5x10^3 cells; HCC1806, 5x10^3 cells; MDA-MB-468, 1.25x10^4 cells) were seeded into a 24 well plate in 1 mL of culture media and left to adhere overnight. The following day, any relevant drug treatments were applied and the confluency of the wells read using an Incucyte S3 using the “whole well” method. Plates were transferred to hypoxia if necessary. Media was changed after 3 days. After 6 days in hypoxia (i.e. 7 days total growth) the confluency of the wells was recorded again using the Incucyte S3.

5.2.9 Propidium iodide assay

Cells in 24 well plates which had been cultured for one week in Section 5.2.8 underwent the propidium iodide assay. Propidium iodide is added to wells to give a final concentration of 50 µM and incubated for approximately 30 minutes. A reading was then taken using a POLARstar OMEGA microplate reader using the following settings: 544 nm excitation, 620 nm emission. Propidium iodide binds to DNA and fluoresces. However, it is non-permeant to live cells. This means in the initial incubation it will only stain dead cells or those with compromised membrane integrity. The first reading represents the number of dead cells. The plate is then freeze-thawed three times to ensure all cells are killed. A second reading is taken which provides a measure of the total number of cells in each well.

5.3 Results

5.3.1 Mitigation of an issue with contaminating guides

During processing of the MDA-MB-231 CRISPR data, the first screen to be conducted, several guides were found to be over-represented beyond the level which could be explained by biological selection. It was discovered that guides targeting *NDUFA1* and *ACACA* had been used in an unrelated experiment at the TDI and this had contaminated the samples. This caused massive over-representation of reads mapping to these guides (average 62% of reads; range 42-83%) (Figure 47).

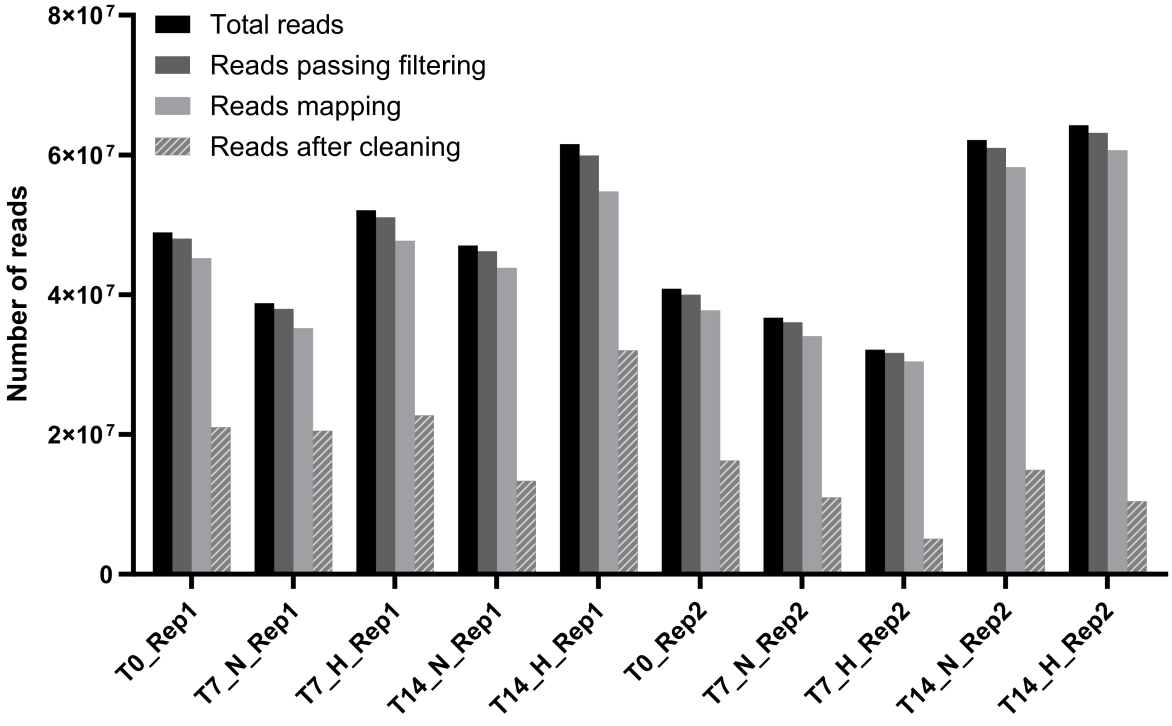

Figure 47: Quality control metrics - MDA-MB-231 CRISPR screen. The number of reads passing each quality control step from the MDA-MB-231 CRISPR data is shown. Most reads pass the initial filtering and map well to the CRISPR library. Removal of the contaminating guides (reads after cleaning) drastically reduces the mapping rate for all samples.

Various quality control steps were carried out to ensure the data were still usable. Firstly, the distribution of guides in the plasmid library was checked to ensure contamination had not occurred at the start of the experiment. The guide distribution in the plasmid library was as expected with no guides present at extremely high levels (Figure 48). Guides targeting *NDUFA1* and *ACACA* are highlighted, showing they were not over-represented in the original library. This is in contrast to the experimental samples, where over-representation of these guides is apparent (Figure 49). When all guides targeting *NDUFA1* and *ACACA* were removed, the guide distribution was comparable to that of the plasmid library (Figure 50). Further sequencing of samples at different stages of preparation showed that the contamination had occurred during the PCR1 step (data not shown).

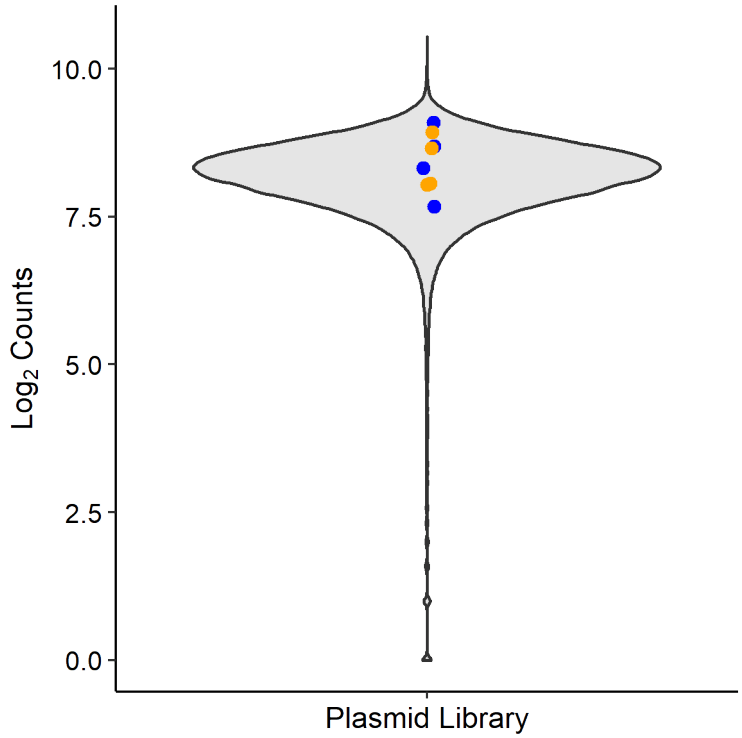

Figure 48: *NDUFA1* and *ACACA* targeting guides in the TKOv3 plasmid library. The guides targeting *NDUFA1* (blue) and *ACACA* (orange) are not over-represented in the TKOv3 plasmid library.

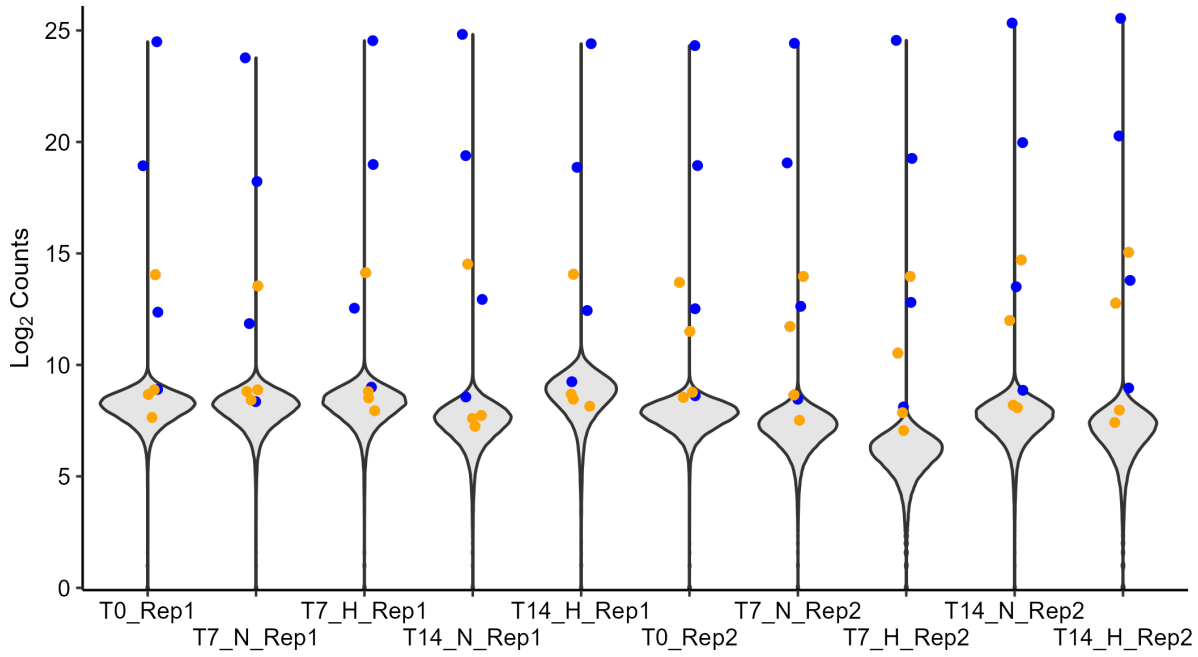

Figure 49: *NDUFA1* and *ACACA* targeting guides in experimental samples. Some of the guides targeting *NDUFA1* (blue) and *ACACA* (orange) are massively over-represented in the experimental samples. Data are presented on a $\log_2$ scale so cover a large range. One of the guides targeting *NDUFA1* has a minimum of 14 million counts per sample, compared with a median guide count across samples of 206.

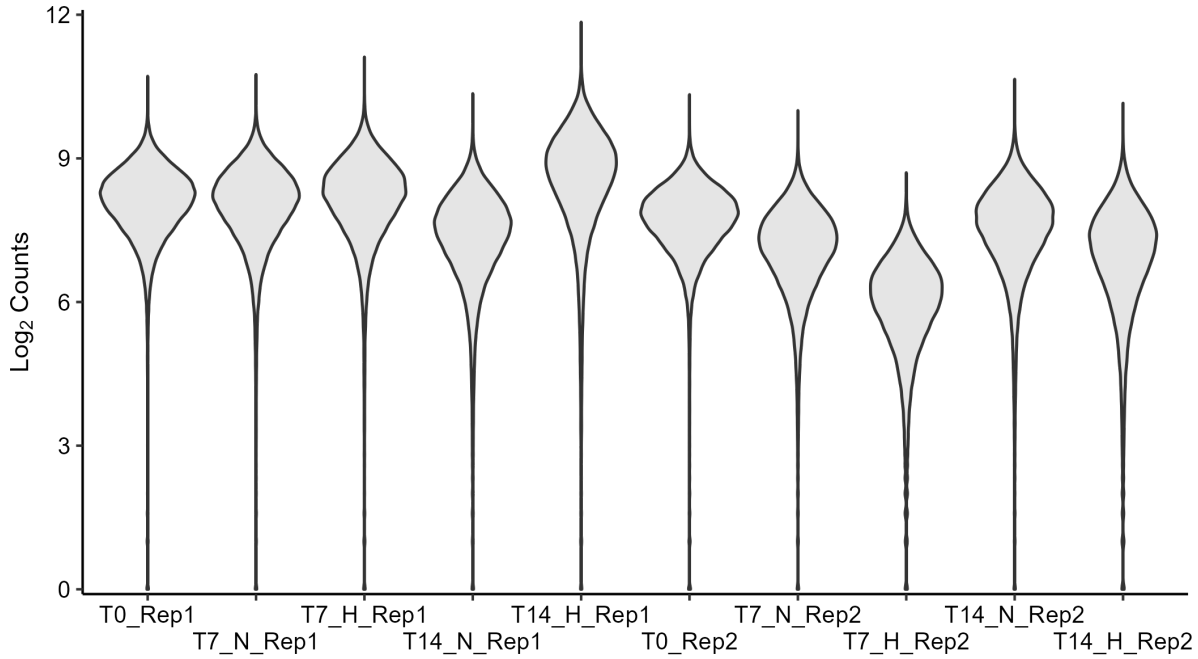

Figure 50: Guide distribution in experimental samples after removal of the contaminating guides. After removal of the contaminating guides, the guide distribution is comparable across samples and contains no extreme outliers.

The program used for analysis, MAGeCK, uses the median for its normalisation strategy so it is not affected by extreme outliers. These analyses show that the guide contamination has not greatly affected the output of the experiment, with the caveat that the coverage of the experiment has been reduced (mean coverage = 245). The contaminating guides dilute the true guide population during the PCR steps and occupy space in the flow cell during sequencing. This reduces the power of the experiment, making it more difficult to detect subtle effects. This issue did not affect the screens conducted on MCF7 or MDA-MB-468 cells.

5.3.2 Quality control

After sequencing, various measures were assessed for quality control. There is some variation in the number of reads produced for each sample, but the percentage of reads which pass filtering and map to the guide sequences is consistent (Figure 51). Less than 1% of bases fail quality checks across all samples and all cell lines, indicating good quality sequencing. The average mapping rate is 86.9% and 86.5% for MDA-MB-468 and MCF7 cells respectively. These screens showed no over-representation of any guide sequences, and their guide distribution was as expected (Figures S6 and S7). The Gini index (reported by MAGeCK) is used to indicate consistency in the read count distribution. The Gini index is $\leqslant 0.2$ in all samples, consistent with the value expected in a negative selection screen [175, 176]. The maximum number of guides with zero counts (i.e. missing guides) in any sample is 875 in MDA-MB-231 samples, 423 in MDA-MB-468 samples, and 341 in MCF7 samples. Even the highest number here (875) only represents 1.2% of the library. This indicates good representation of the library but suggests that some coverage may have been lost in the MDA-MB-231 screen, likely due to the contaminating guides.

Principal component analysis (PCA) was used to check whether samples varied in the expected manner. In all screens, samples separated along the first principal component in a time dependent manner (Figure 52). This suggests that the major driver of variation in the samples is time, rather than experimental condition (i.e. normoxia vs hypoxia). T7 samples cluster more closely than T14 samples. This is expected as more time allows more selection to take place and therefore more variation to arise between the samples. In the MDA-MB-231 and MDA-MB-468 screens, variation in the second principal component can be largely explained by the replicate number, but this is not the case for the MCF7 screen.

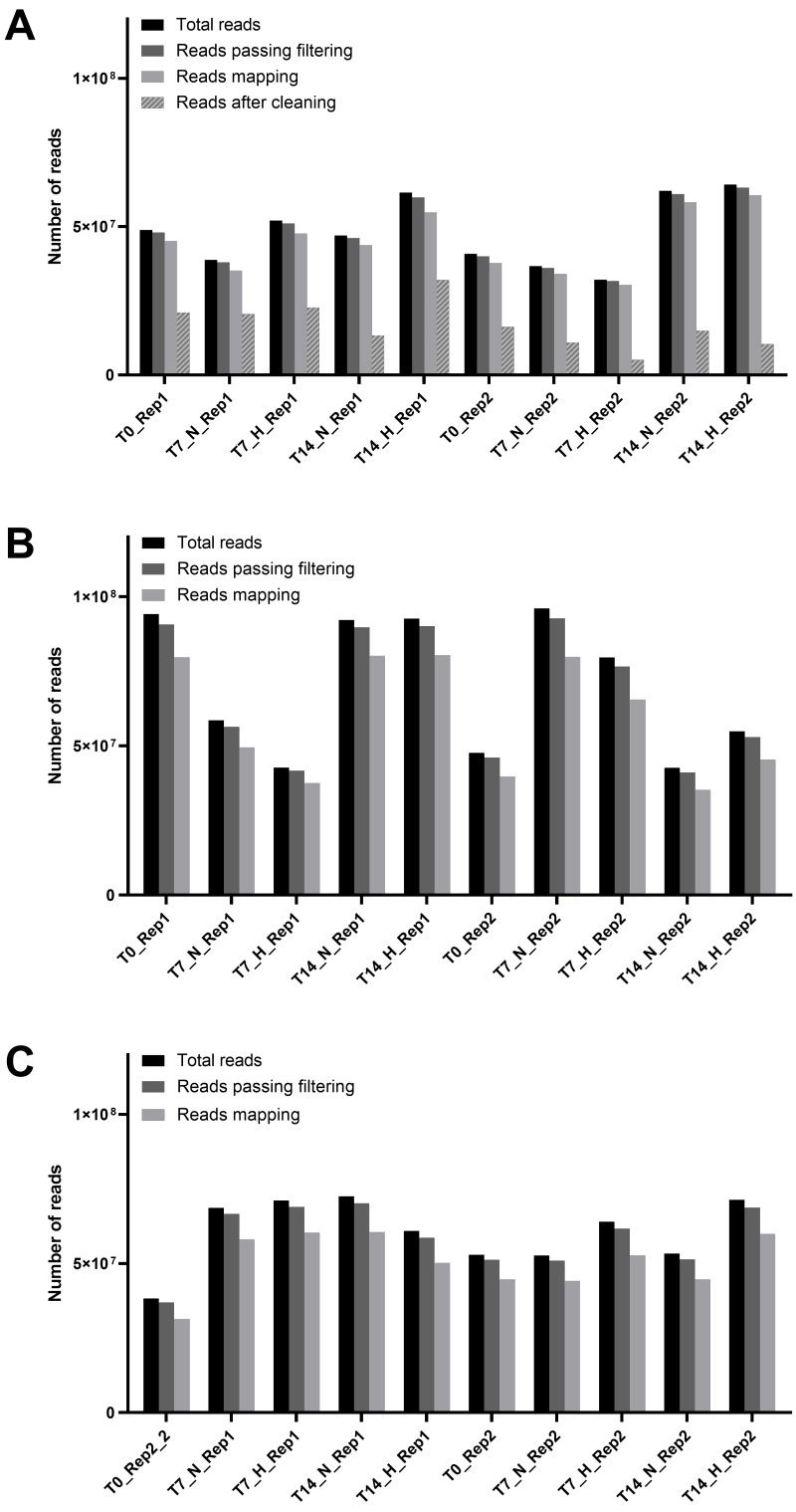

Figure 51: Assessment of quality control metrics in the CRISPR sequencing data. The number of reads obtained from sequencing, the number of reads passing quality control filters, and the number of reads that align to the TKOv3 CRISPR library are shown for **(A)** MDA-MB-231, **(B)** MDA-MB-468, and **(C)** MCF7 cells. An additional column shows the number of aligning reads after removal of the contaminating guides in the MDA-MB-231 samples.

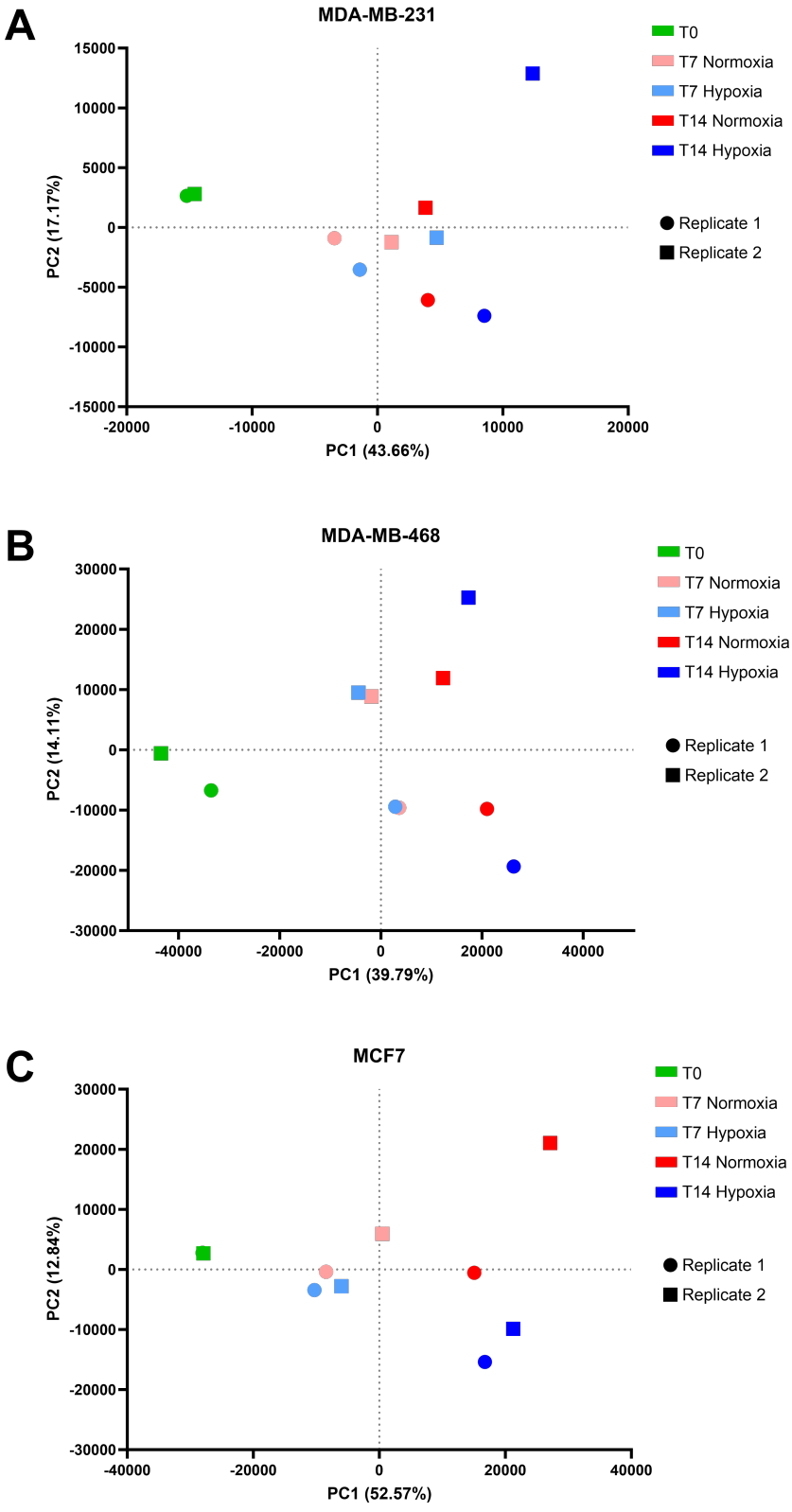

Figure 52: Principal component analysis of all TKOv3 CRISPR screens. Principal component analysis of CRISPR screens of MDA-MB-231, MDA-MB-468, and MCF7 cells. Samples separate along the first principal component in a time-dependent manner. PC1 = principal component 1. PC2 = principal component 2.

In all screens, the non-targeting control guides show enrichment over time (Figures S8 to S10). This phenomenon has been noted by other researchers [177] and likely occurs because non-targeting guides are the only ones which do not cause double-strand breaks in the DNA. In a population where double-strand breaks are the norm, this offers a survival advantage. For this reason, non-targeting guides were not used to create a null distribution.

To ensure the screen functioned as expected, the effect of guides targeting known essential genes was assessed. A curated list of known core essential genes (CEG2) was used [173]. When compared with T0, all timepoints and conditions show a depletion of guides which target core essential genes (CEGs) (Figures 53 to 55). There is also a high degree of overlap between genes that are significantly (adjusted p value $\leqslant 0.05$) negatively selected and known CEGs (Figure 56). MCF7 cells have a slower growth rate, and fewer CEGs are labelled as significant. This suggests that this screen would have benefited from a longer exposure time to compensate for the longer doubling time. Despite this, this positive control shows that the screens have functioned as expected and are able to detect survival disadvantages conferred by gene knockout. This also shows that enough power remains in the MDA-MB-231 screen to identify genes of interest despite the reduction in coverage caused by the contaminating guides.

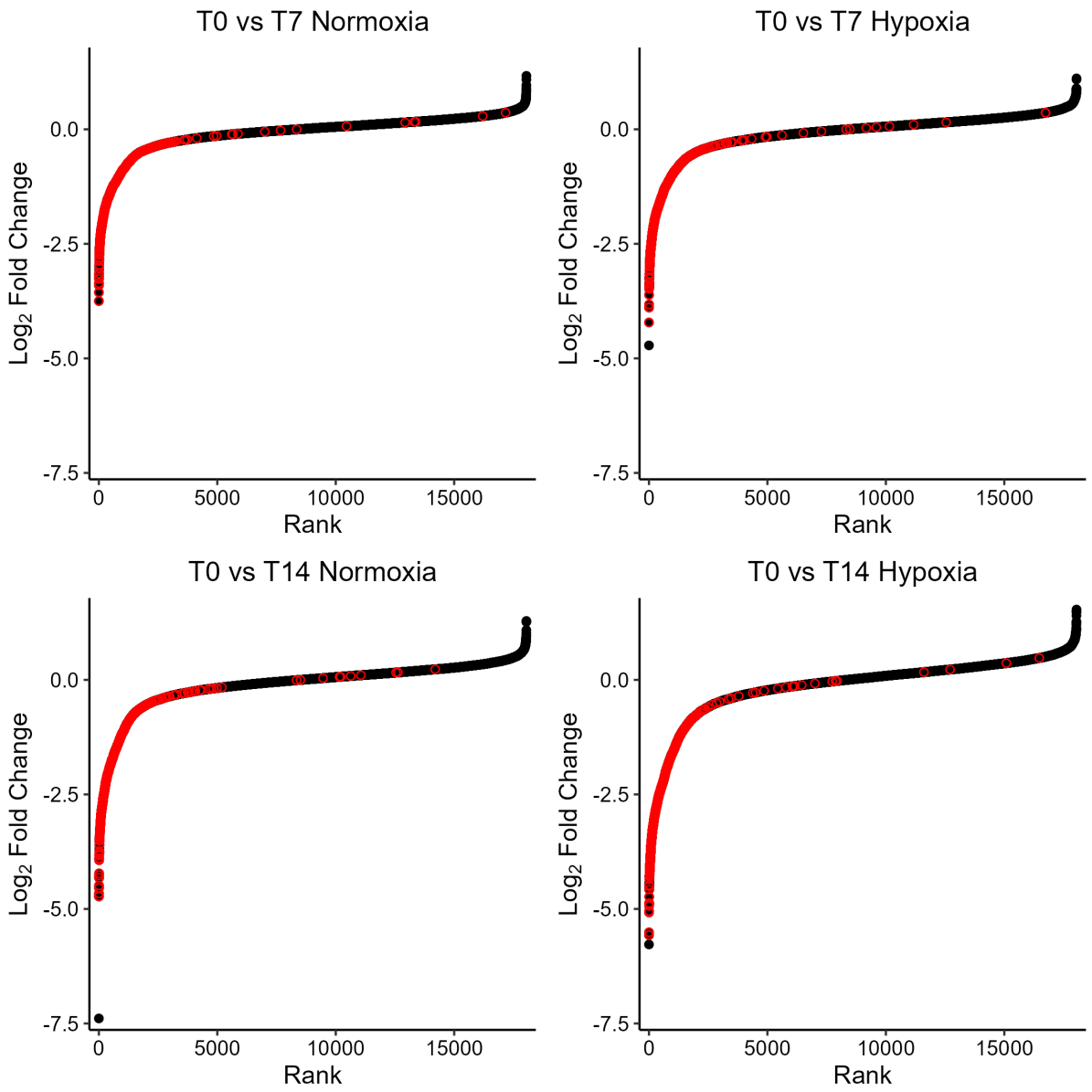

Figure 53: Log_2 fold change of CEGs - MDA-MB-231 screen. The log_2 fold change of guides targeting genes in the CEG2 core essential gene list (red) is generally negative, showing these guides confer a growth/survival disadvantage.

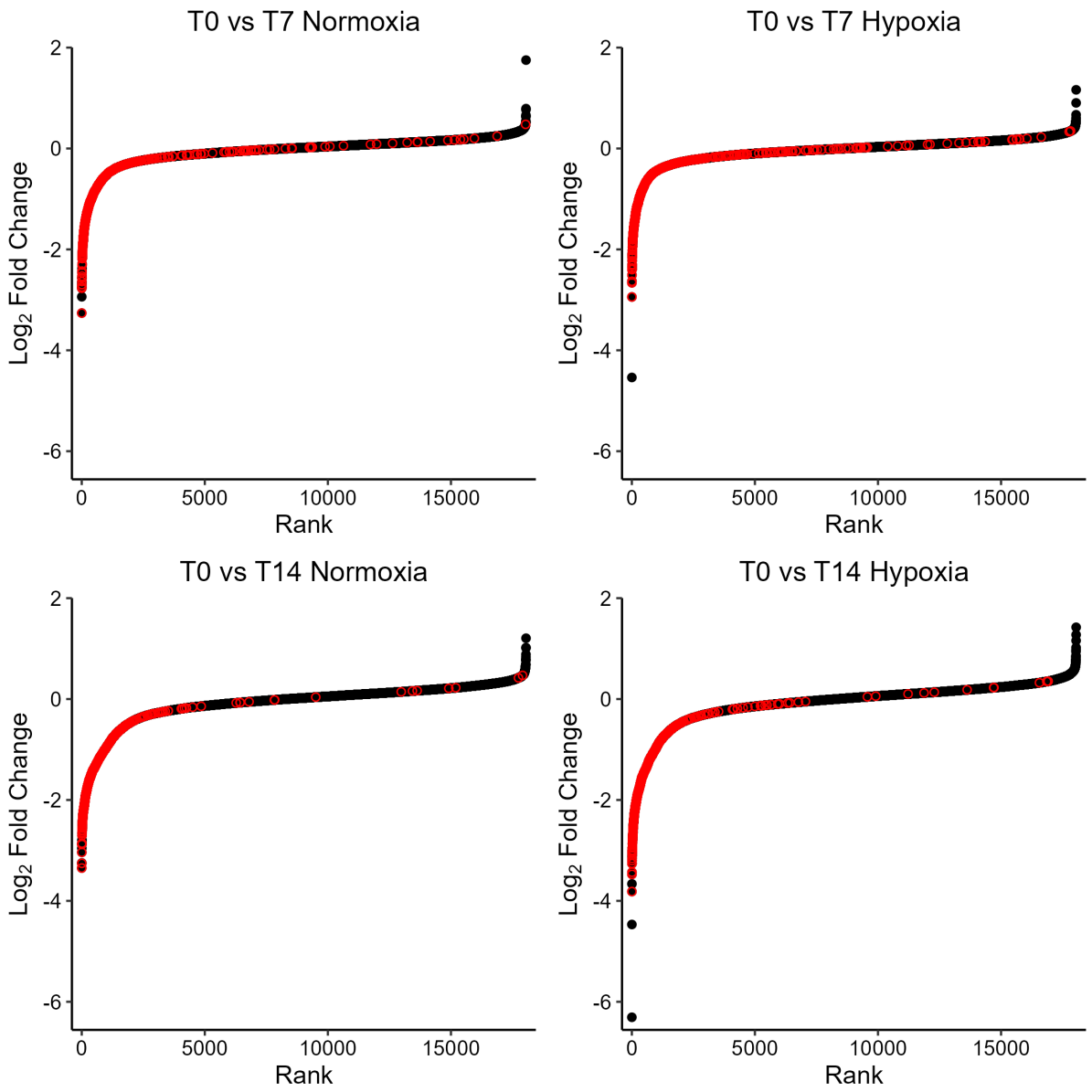

Figure 54: Log_2 fold change of CEGs - MDA-MB-468 screen. The log_2 fold change of guides targeting genes in the CEG2 core essential gene list (red) is generally negative, showing these guides confer a growth/survival disadvantage.

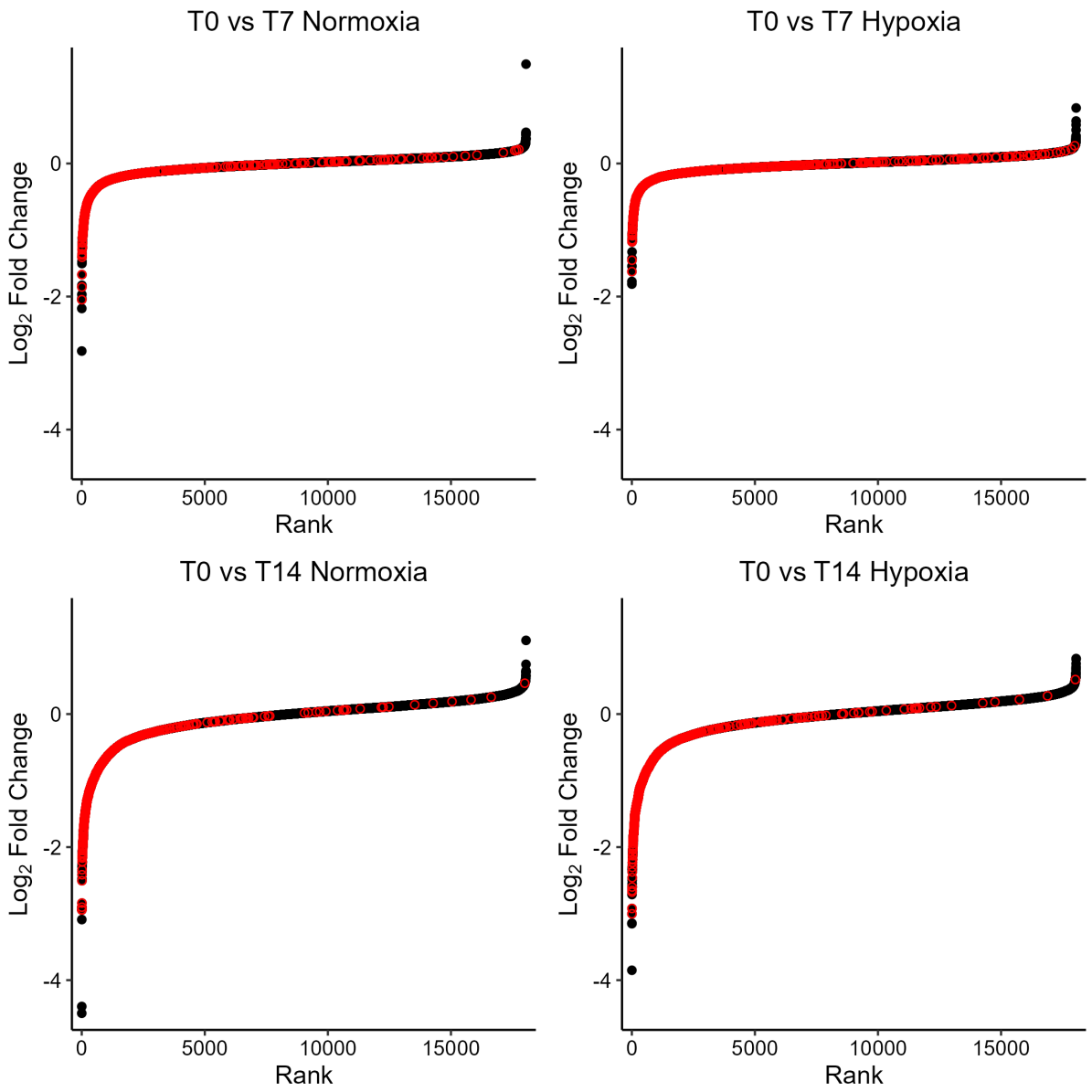

Figure 55: Log_2 fold change of CEGs - MCF7 screen. The log_2 fold change of guides targeting genes in the CEG2 core essential gene list (red) is generally negative by day 14, showing these guides confer a growth/survival disadvantage.

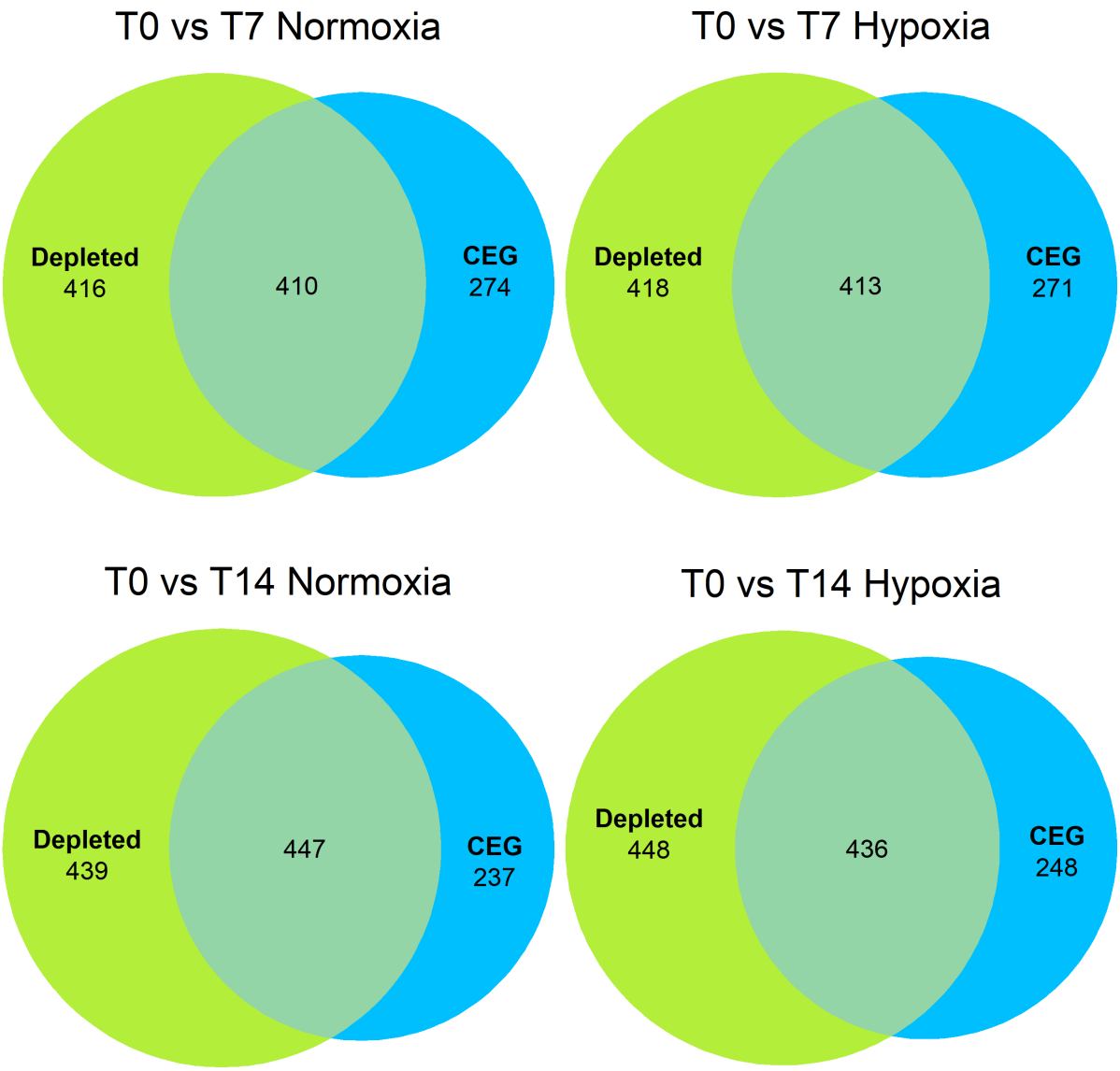

Figure 56: Overlap of CEGs and genes whose targeting guides are significantly depleted in MDA-MB-231 samples. The overlap of genes in the CEG2 list and genes whose knockout causes a significant (adjusted p value $\leqslant 0.05$) survival disadvantage. Around 2/3 of known CEGs are captured, even as early as T7.

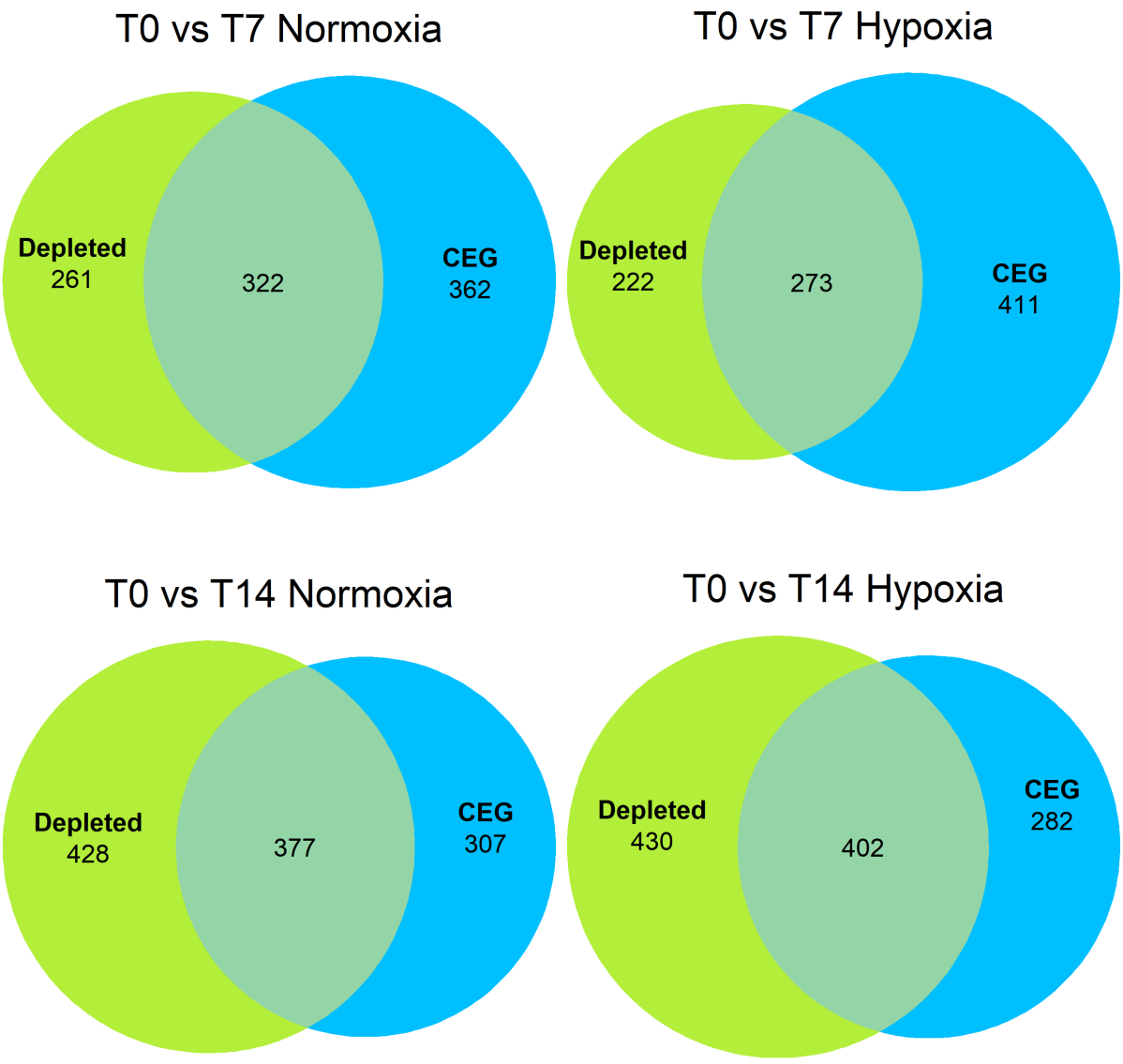

Figure 57: Overlap of CEGs and genes whose targeting guides are significantly depleted in MDA-MB-468 samples. The overlap of genes in the CEG2 list and genes whose knockout causes a significant (adjusted p value $\leqslant 0.05$) survival disadvantage. Around 1/2 of known CEGs are captured by T14.

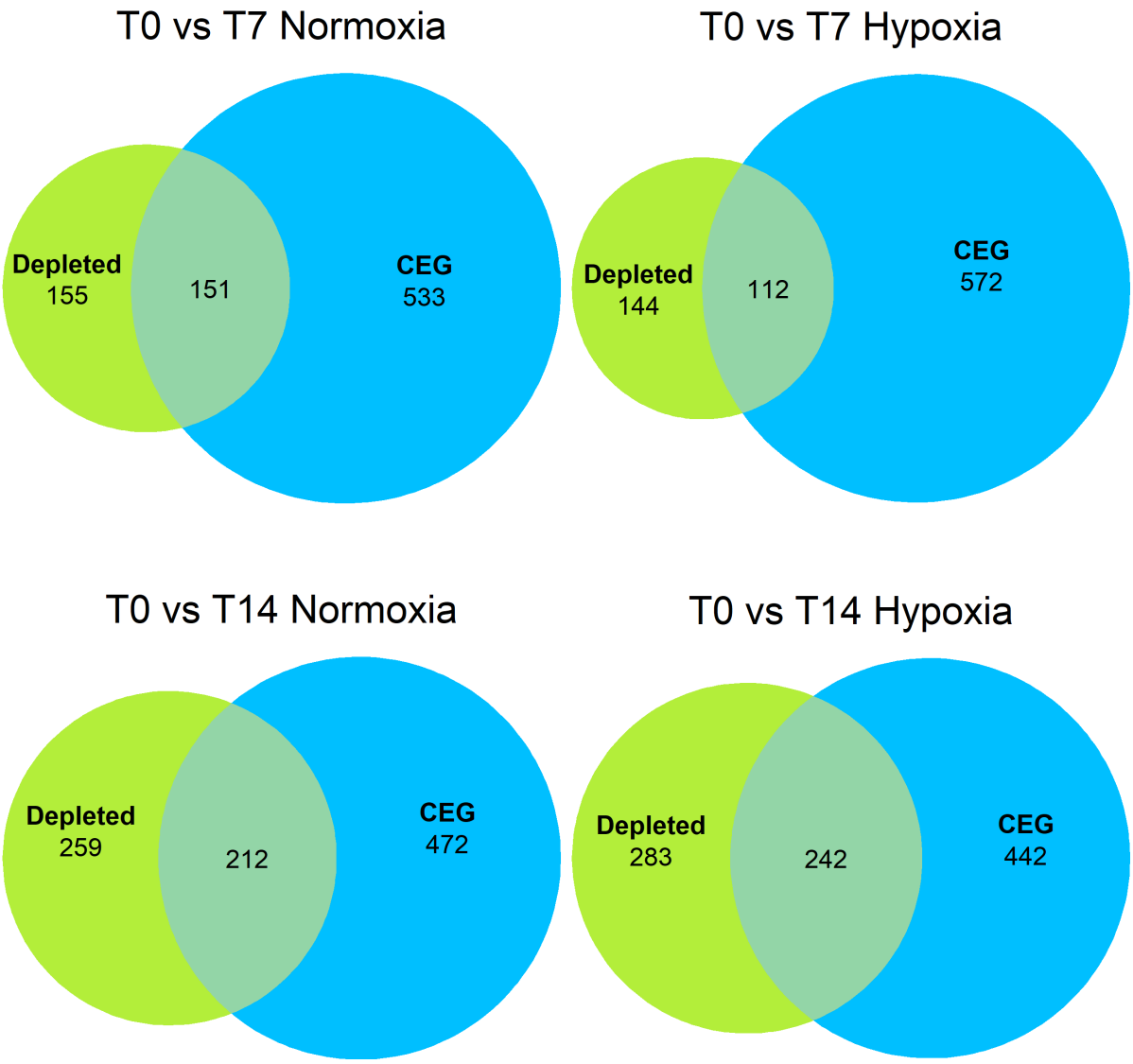

Figure 58: Overlap of CEGs and genes whose targeting guides are significantly depleted in MCF7 samples. The overlap of genes in the CEG2 list and genes whose knockout causes a significant (adjusted p value $\leqslant 0.05$) survival disadvantage. Around 1/3 of known CEGs are captured by T14.

5.3.3 HIF1/2 do not appear to be essential for survival under hypoxia

Hypoxia triggers huge transcriptional changes in the cell, but it is not known whether these changes are essential for cell survival. To identify genes which have a differential effect on cell growth or survival specifically under hypoxic conditions, samples collected at day 14 (T14) in normoxia and hypoxia were directly compared. I assessed the performance of master transcription factor genes, HIF1α (*HIF1A*), HIF2α (*EPAS1*), and HIF1β (*ARNT*) in this setting. In none of the cell lines were any of these genes essential under hypoxia (Figure 59). In fact, cells with these genes knocked out tended to be enriched in hypoxic samples when compared to normoxia. Of interest, HIF1 targeting guides are significantly (adjusted *p* value = 0.0498) more abundant in the MCF7 hypoxic population compared to the normoxic population after 14 days of culture (Figure 60). This cell line has the highest expression of HIF1α over the experimental time course (Figure 20). Similarly, MDA-MB-468 cells, which have the highest expression of HIF2α over the time course (Figure 21), showed significant enrichment of HIF2 targeting guides under hypoxia (T0 vs T14 Hypoxia: adjusted *p* value = 0.0035; T14 Normoxia vs T14 Hypoxia: adjusted *p* value = 0.00045). There is (non-significant) positive selection of *ARNT* and *HIF1A* knockout MDA-MB-231 cells under both normoxia and hypoxia. There were detectable levels of HIF1 in MDA-MB-231 cells even in normoxia (**??**), which may explain why the normoxic cells are also affected by HIF1 knockout. These data suggest that the effect of HIF knockout on survival is connected to the HIF gene which is active in the cells and, contrary to popular opinion, that HIF activity may not be necessary for survival under hypoxia.

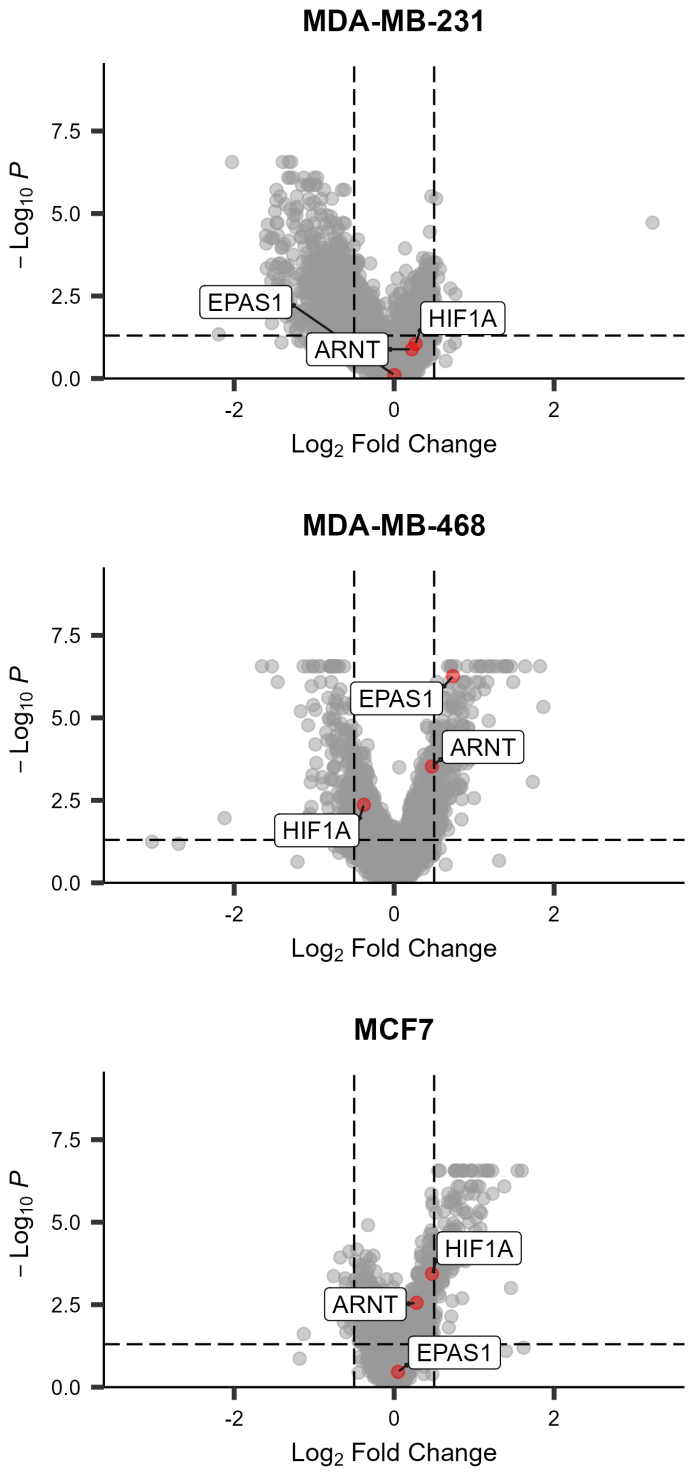

Figure 59: Selection of HIF subunit genes in T14 normoxia vs T14 hypoxia samples. Volcano plots showing the results of the T14 normoxia vs T14 hypoxia comparison. *HIF1A*, *EPAS1*, and *ARNT* are highlighted. Genes on the upper left of the volcano plots are more essential in hypoxia, while genes on the upper right are more essential in normoxia.

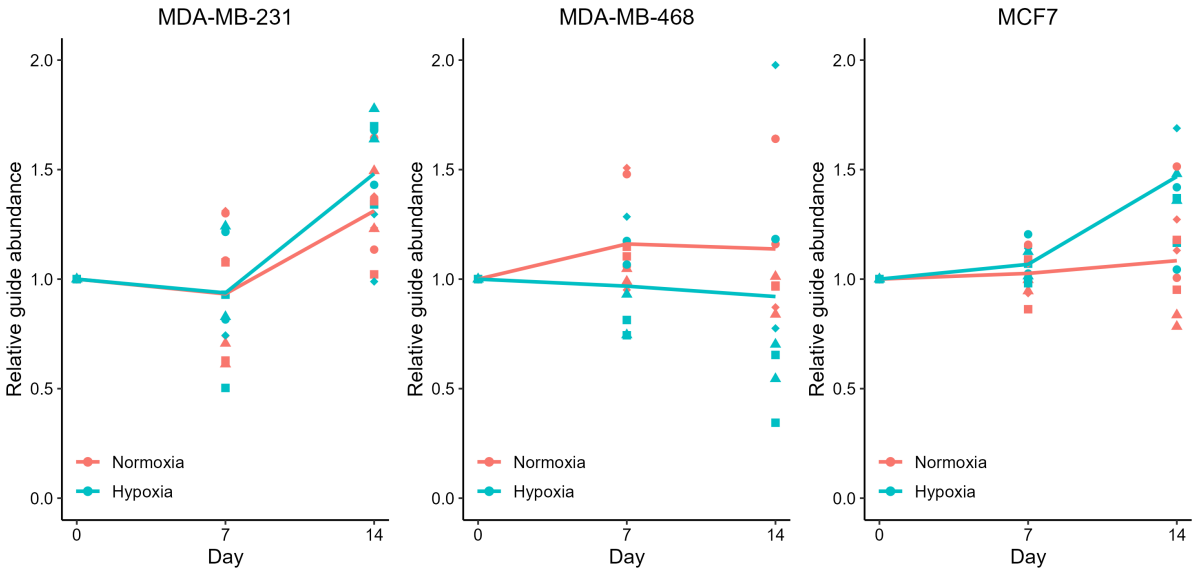

Figure 60: Change in abundance of *HIF1A* targeting guides over time. The abundance of each *HIF1A* (HIF1α) targeting guide relative to T0. Both replicates are displayed and the shapes of the symbols represent each unique guide. This allows guides which have little effect, and therefore may have low guide efficiency, to be identified. Points are connected from the mean at each timepoint.

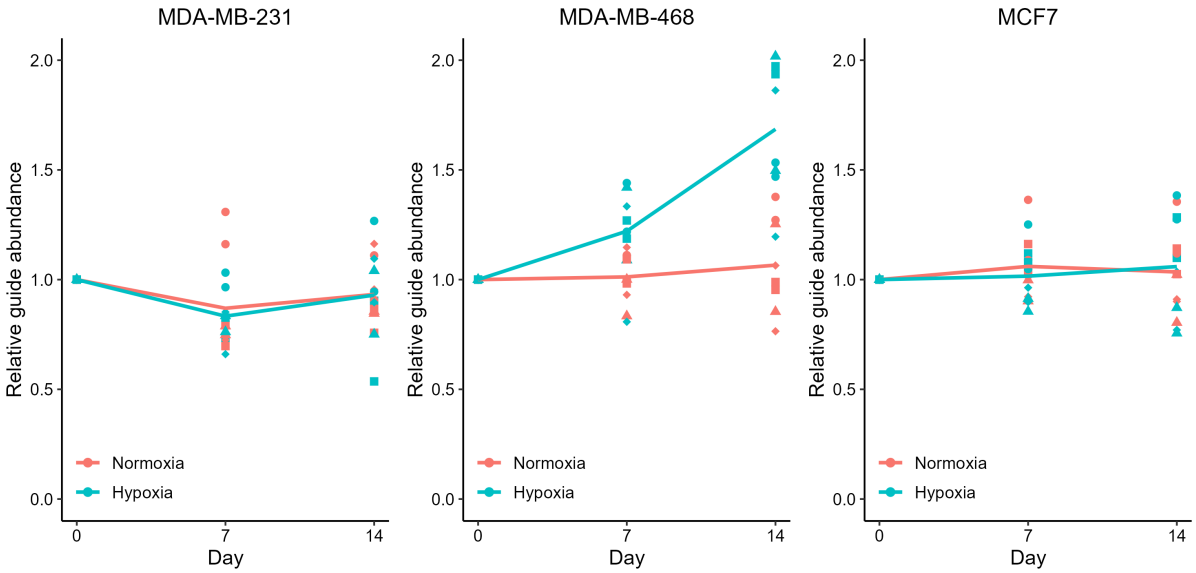

Figure 61: Change in abundance of *EPAS1* targeting guides over time. The abundance of each *EPAS1* (HIF2α) targeting guide relative to T0. Both replicates are displayed and the shapes of the symbols represent each unique guide. This allows guides which have little effect, and therefore may have low guide efficiency, to be identified. Points are connected from the mean at each timepoint.

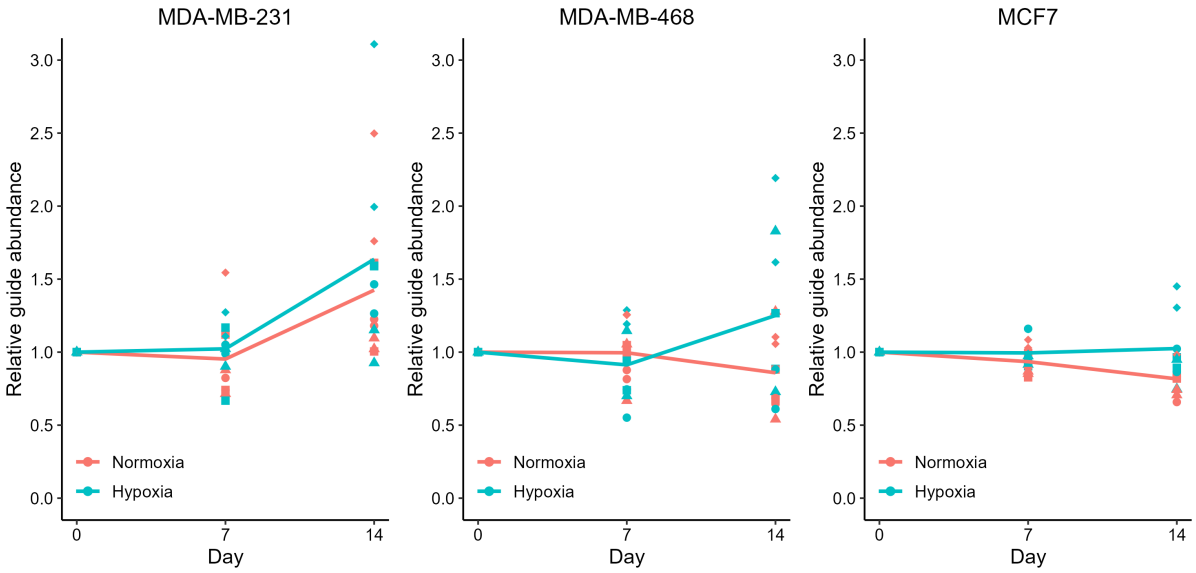

Figure 62: Change in abundance of *ARNT* targeting guides over time. The abundance of each *ARNT* (HIF1β) targeting guide relative to T0. Both replicates are displayed and the shapes of the symbols represent each unique guide. This allows guides which have little effect, and therefore may have low guide efficiency, to be identified. Points are connected from the mean at each timepoint.

5.3.4 Hypoxia response genes do not appear to be essential for survival under hypoxia

Though the HIF transcription factors did not appear to be essential under hypoxia, expression of some of their downstream targets may be. The high degree of overlap in activity of the HIF subtypes may mean that they can compensate for one another in this system (i.e. if HIF1 is knocked out, HIF2 can compensate, and vice versa). This means that knockout of a HIF target gene may have a greater effect than targeting the HIF subunit genes themselves. Given this, I asked whether HIF targets, or known hypoxia response genes, were more likely to be essential for cell survival under hypoxia. To test this, the distribution of:

1. Direct HIF targets (ChIP Atlas) [119]
2. A curated HIF regulon (DoRothEA) [120]
3. An established hypoxia gene signature (Buffa) [116]

in the ranked gene results from each CRISPR screen was assessed by hypergeometric test. The distribution of genes in these lists did not deviate from the norm in any cell line (p value $>$ 0.05) (Figures 63 to 65).

I also chose to investigate the distribution of the chronic hypoxia response genes identified in Section 4 (Table 2). The distribution of genes in this list also did not significantly deviate from the norm in any cell line (p value > 0.05). This means that, although the genes are upregulated under chronic hypoxia, they do not appear to be essential for the cell's survival.

Similarly, there is a lack of correlation between the $\log_2$ fold change of a gene in the RNA-seq data (Section 4) and its $\log_2$ fold change in the CRISPR screen (Figure 67).

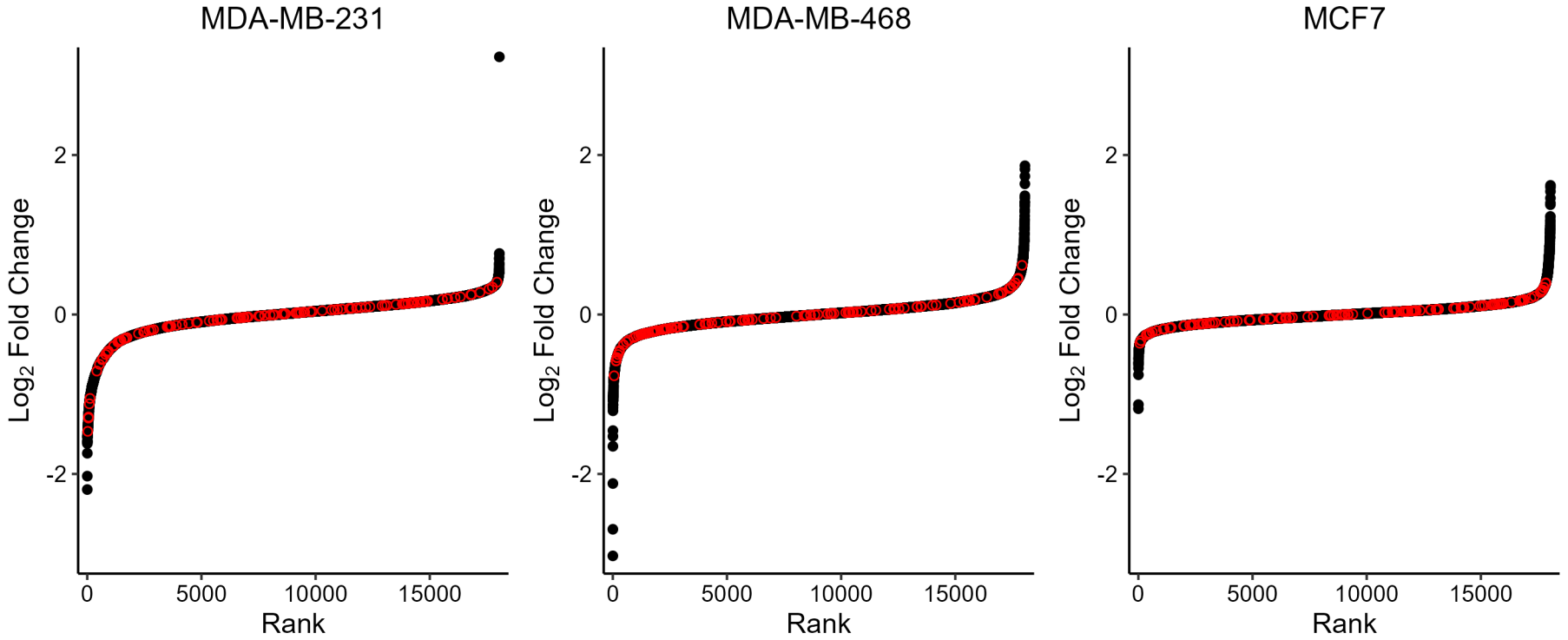

Figure 63: Distribution of HIF target genes (ChIP Atlas) in the T14 normoxia vs T14 hypoxia ranked gene results. HIF target genes predicted using ChIP Atlas are highlighted in red. They distribute evenly throughout the results as measured by hypergeometric test ($p > 0.05$), appearing not to be selectively essential under hypoxia.

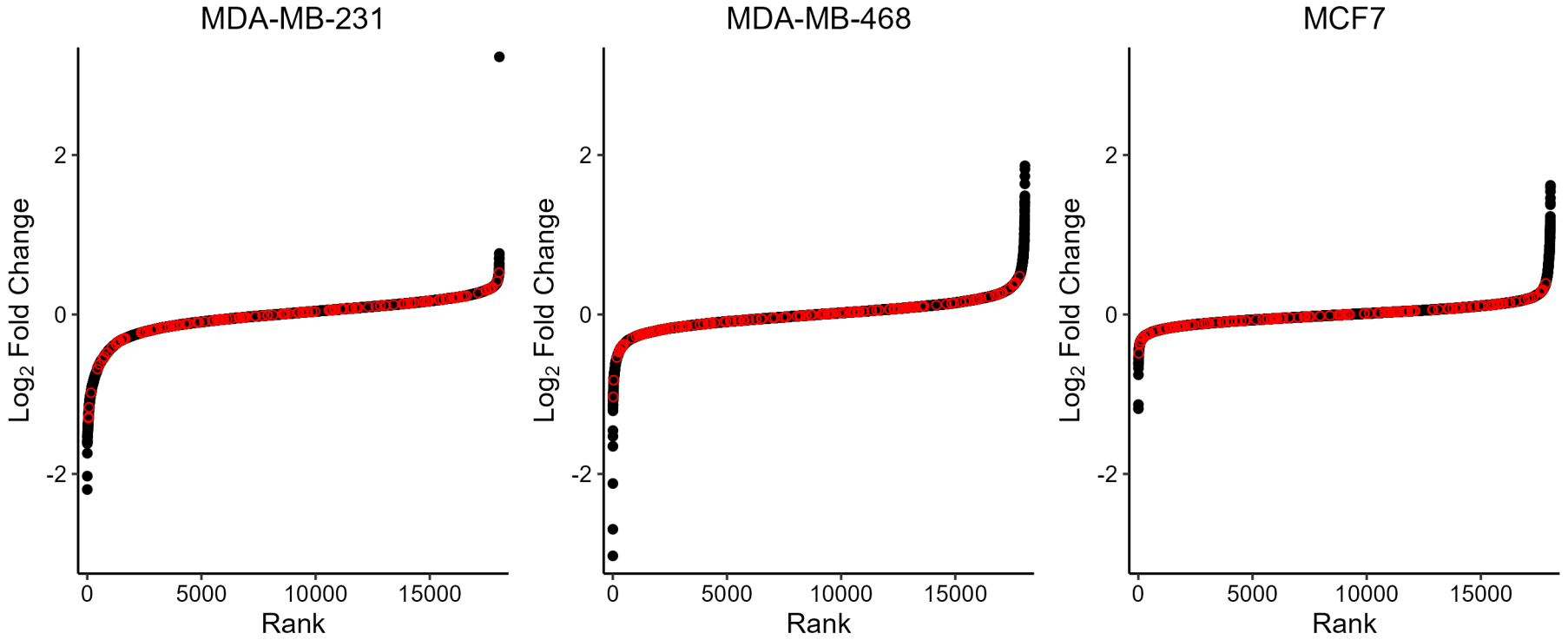

Figure 64: Distribution of HIF target genes (DoRothEA) in the T14 normoxia vs T14 hypoxia ranked gene results. HIF target genes from the curated DoRothEA regulon are highlighted in red. They distribute evenly throughout the results as measured by hypergeometric test ($p > 0.05$), appearing not to be selectively essential under hypoxia.

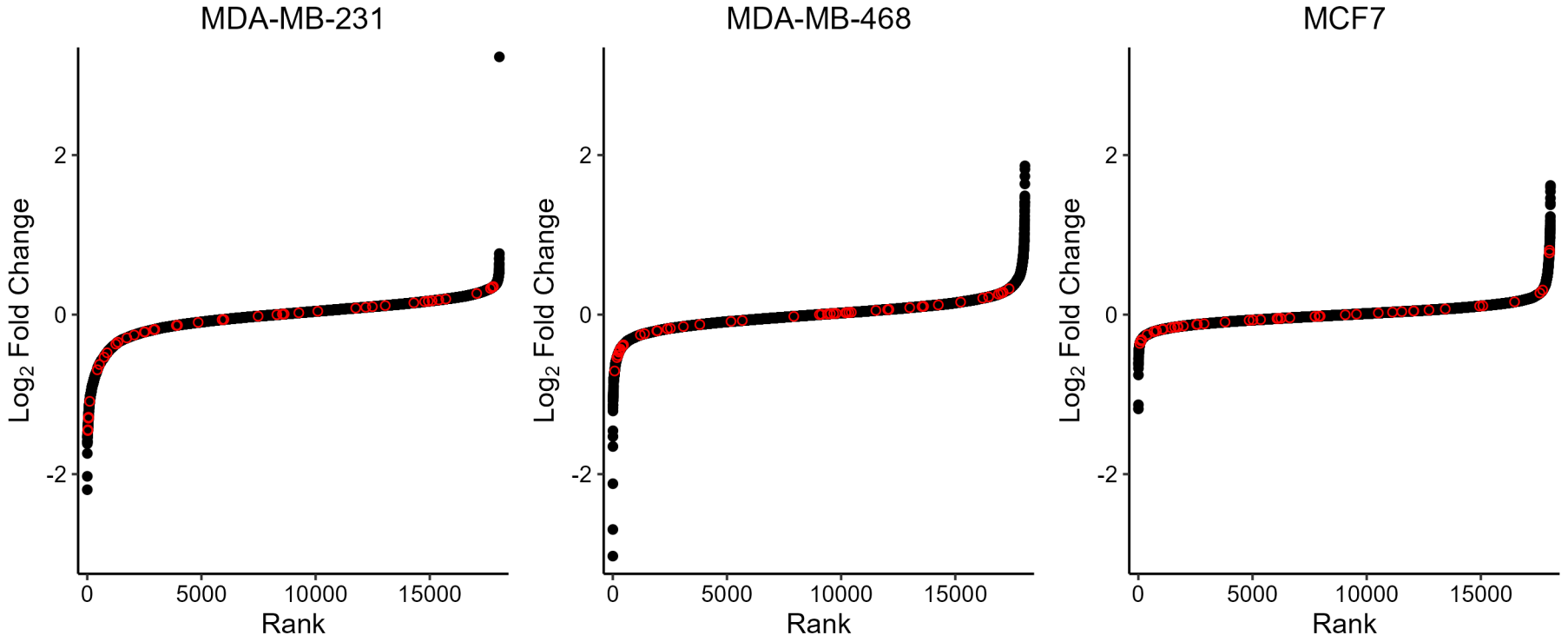

Figure 65: Distribution of hypoxia signature genes (Buffa) in the T14 normoxia vs T14 hypoxia ranked gene results. Genes from the Buffa hypoxia signature are highlighted in red. They distribute evenly throughout the results as measured by hypergeometric test ($p > 0.05$), appearing not to be selectively essential under hypoxia.

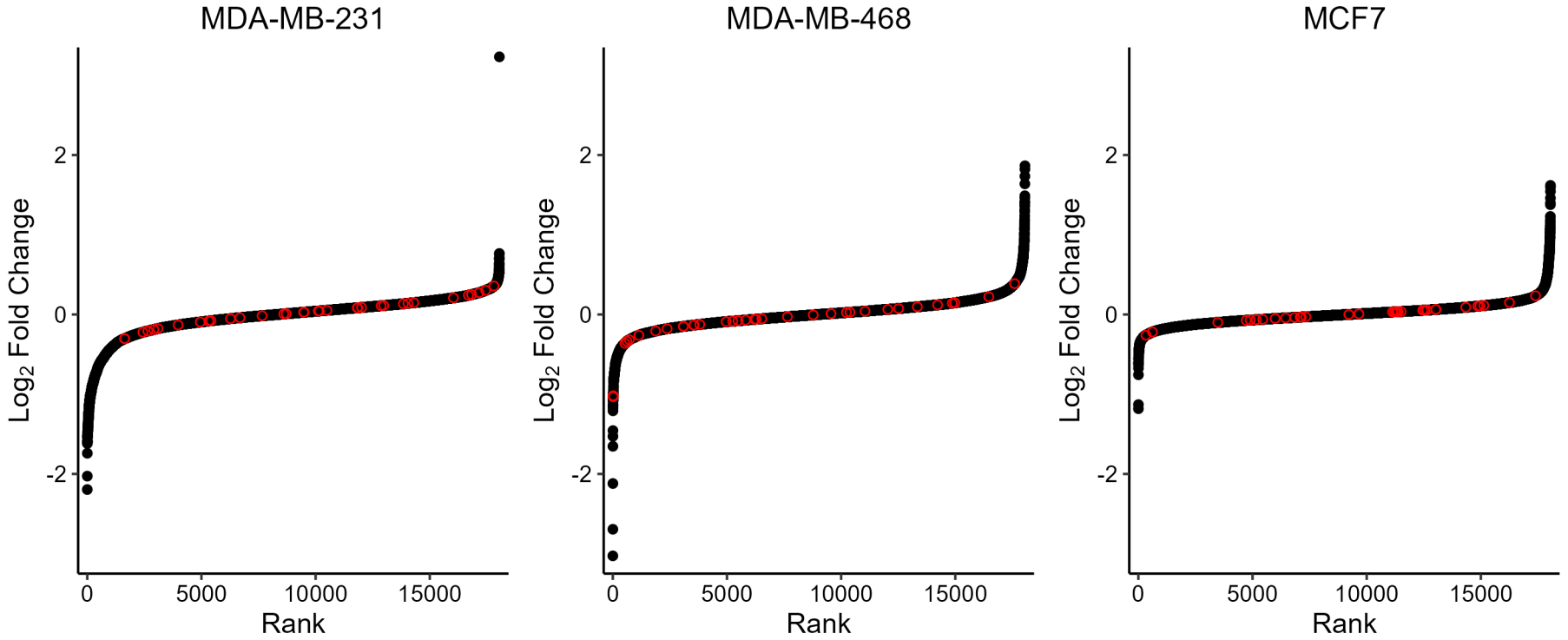

Figure 66: [Distribution of genes upregulated specifically under chronic hypoxia in the T14 normoxia vs T14 hypoxia ranked gene results. Genes from Table 2, which are upregulated specifically under chronic hypoxia, are highlighted in red. They distribute evenly throughout the results as measured by hypergeometric test ($p > 0.05$), appearing not to be selected for or against under hypoxia.

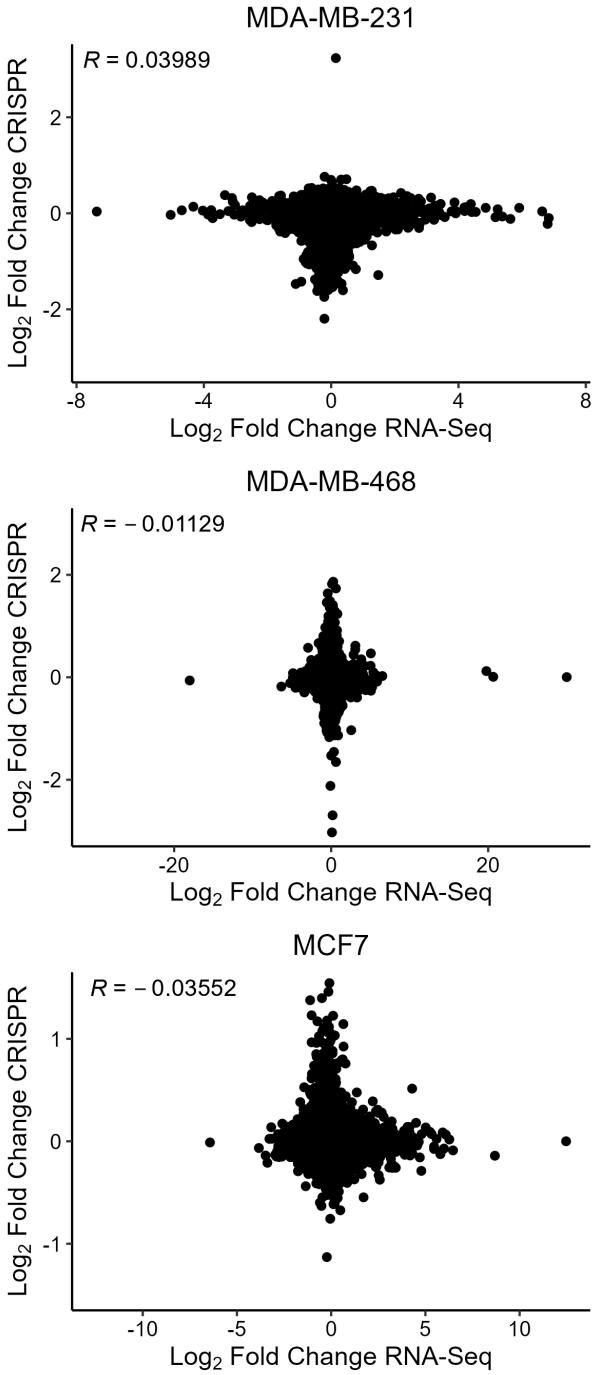

Figure 67: Correlation of log_2 fold changes in RNA-seq and CRISPR. There is no correlation between hypoxia-specific essentiality in the CRISPR screen and changes in gene expression under hypoxia in any of the cell lines tested. CRISPR log_2 fold changes are taken from the T14 normoxia vs T14 hypoxia comparison. RNA-seq log_2 fold changes represent gene expression changes after 14 days of hypoxia exposure.

5.3.5 There is little overlap in hypoxia specific essential genes between cell lines

In general, there was little overlap between the cell lines in genes that were essential for survival under hypoxia. This reflects the variation in the transcriptional response to chronic hypoxia seen between cell lines. There is a difference in the skew of the results between each cell line (Figure 68). For example, in MDA-MB-231 cells, gene knockouts were more likely to cause a selection disadvantage in hypoxia than in normoxia. In MCF7 cells, this trend is reversed. MDA-MB-468 cells have a more equal distribution of genes having each effect. The distribution of non-targeting guides in the T14 normoxia vs T14 hypoxia comparisons is similar between cell lines (Figure S11), meaning the differences in skew here are likely to reflect true biological differences in the cell lines.

To identify cell processes which were more/less essential under hypoxia in each cell line, gene set enrichment analysis (GSEA) was conducted using clusterProfiler (version 4.6.2) (Figure 69). Despite a general lack of overlap in the significant results, three pathways appear in all of the cell lines as being more essential in hypoxia. These are: spliceosome, nucleotide excision repair, and proteasome. The spliceosome and proteasome could both be linked to the hypoxic response, as HIF target genes are expressed through production of new mRNA molecules which must be processed using spliceosome machinery. Meanwhile, protein degradation may increase to provide amino acids for building these new proteins.

There are also some interesting differences in the data, particularly regarding pathways involved in mitochondrial function. In MCF7 and MDA-MB-468 cells, GSEA shows mitochondrial genes are selectively essential under normoxic conditions, with both oxidative phosphorylation and the tricarboxylic acid (TCA) cycle being significant results. Conversely, in MDA-MB-231 cells, genes involved in mitochondrial function are more essential in hypoxic cells, with both of the aforementioned pathways being significantly negatively enriched. There is an overlap in the individual genes which contribute to the enrichment in each cell line, meaning the same genes which are more essential in normoxia in MCF7 and MDA-MB-468 cells are more essential in hypoxia in MDA-MB-231 cells. This is particularly true for genes which encode subunits of complex I and complex III of the mitochondrial electron transport chain.

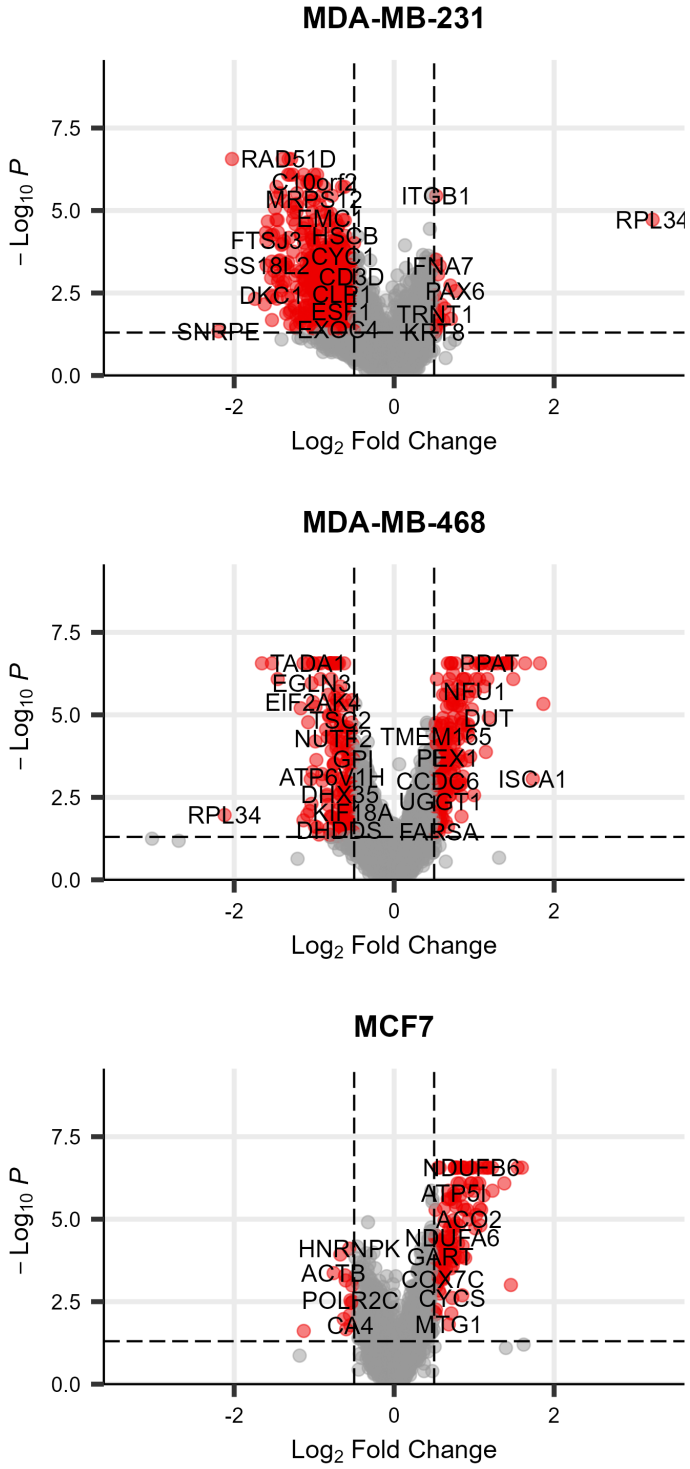

Figure 68: T14 normoxia vs T14 hypoxia CRISPR results. Results of the T14 normoxia vs T14 hypoxia comparison. Genes with a $\log_2$ fold change $\geqslant$ 0.5 or $\leqslant$ -0.5, and a p value $\leqslant$ 0.05 are displayed in red. Genes on the upper left of the volcano plots are more essential in hypoxia, while genes on the upper right are more essential in normoxia.

MDA-MB-231

KEGG Pathway	NES	Adjusted *p* value
Cytokine-cytokine receptor interaction	+1.35	0.01
Alcoholic liver disease	+1.50	0.02
JAK-STAT signalling pathway	+1.40	0.06
Retinol metabolism	+1.54	0.07

KEGG Pathway	NES	Adjusted *p* value
Spliceosome	-2.40	0.00
Oxidative phosphorylation	-2.49	0.00
Nucleotide excision repair	-2.22	0.00
Cell cycle	-1.96	0.00
Proteasome	-2.18	0.00
Citrate cycle (TCA cycle)	-1.71	0.05

MDA-MB-468

KEGG Pathway	NES	Adjusted *p* value
N-Glycan biosynthesis	+2.66	0.00
Chemical carcinogenesis - reactive oxygen species	+2.24	0.00
Oxidative phosphorylation	+2.01	0.00
Citrate cycle (TCA cycle)	+1.78	0.02

KEGG Pathway	NES	Adjusted *p* value
Cell cycle	-2.39	0.00
Proteasome	-2.48	0.00
Ribosome	-2.18	0.00
Nucleotide excision repair	-2.32	0.00
Spliceosome	-1.64	0.01

MCF7

KEGG Pathway	NES	Adjusted *p* value
Chemical carcinogenesis - reactive oxygen species	+2.78	0.00
Oxidative phosphorylation	+2.98	0.00
Citrate cycle (TCA cycle)	+2.56	0.00
Ribosome	+2.24	0.00

KEGG Pathway	NES	Adjusted *p* value
Spliceosome	-2.13	0.00
Nucleotide excision repair	-2.03	0.00
Proteasome	-1.87	0.01
HIF-1 signalling pathway	-1.57	0.04

Figure 69: GSEA of gene essentiality results between T14 normoxia and T14 hypoxia. Results for all genes were sorted by their $\log_2$ fold change and used to run GSEA using KEGG terms. Significance values were corrected for multiple testing using the Benjamini–Hochberg method. Selected results are shown. Pathways containing genes more essential in normoxia have a positive normalised enrichment score (NES) (red), and those containing genes more essential in hypoxia have a negative NES (blue).

5.3.6 *EIF2AK4* and *GPI* are selectively essential under hypoxia

Since I was interested in genes whose knockout causes a reduction in growth specifically in hypoxia, I looked for overlap in these genes between the cell lines. In MCF7 cells, no genes passed the adjusted *p* value threshold of 0.05. In MDA-MB-231 and MDA-MB-468 cells, there were 150 and 72 genes respectively which had an adjusted *p* value $\leqslant$ 0.05, and a $\log_2$ fold change $\leqslant$ -0.5. Of these genes, only two are overlapping: *EIF2AK4* and *GPI* (Figure 70). *EIF2AK4* encodes the eukaryotic translation initiation factor 2α kinase 4, otherwise known as GCN2. This gene activates the ISR through phosphorylation of eIF2α. *GPI* encodes glucose-6-phosphate isomerase, a glycolytic enzyme which catalyses the reversible isomerisation of glucose-6-phosphate and fructose-6-phosphate.

Despite not being significant in MCF7 cells, *EIF2AK4* is ranked 8th for hypoxia selective essentiality (adjusted *p* value = 0.33, $\log_2$ fold change -0.48). *GPI* is ranked 112th in MCF7 cells (adjusted *p* value = 0.65, $\log_2$ fold change -0.18). The effects of these gene knockouts are shown in Figures 71 and 72.

Although the hypoxic cells were always more sensitive to *EIF2AK4* knockout than their normoxic counterparts, there was a cell line difference in how sensitive the normoxic cells were. MDA-MB-231 cells were more sensitive to the knockout in normoxia than MDA-MB-468 and MCF7 cells, which appeared to be unaffected. *GPI* knockout did not appear to affect MDA-MB-468 cells in normoxia, whereas MDA-MB-231 and MCF7 cells were somewhat sensitive.

Results from a published hypoxia CRISPR screen were taken into consideration when choosing which of these hits to validate [118]. When comparing samples cultured in 21% and 1% O_2 after 15 days, *EIF2AK4* ranked as the 135th hit in their screen (adjusted *p* value = 0.17). This is despite the fact they use a different cell line (K562, a myelogenous leukaemia cell line) and a different genome-wide library. On the other hand, *GPI* ranked 16282nd out of 20112 genes in their study (adjusted *p* value = 1).

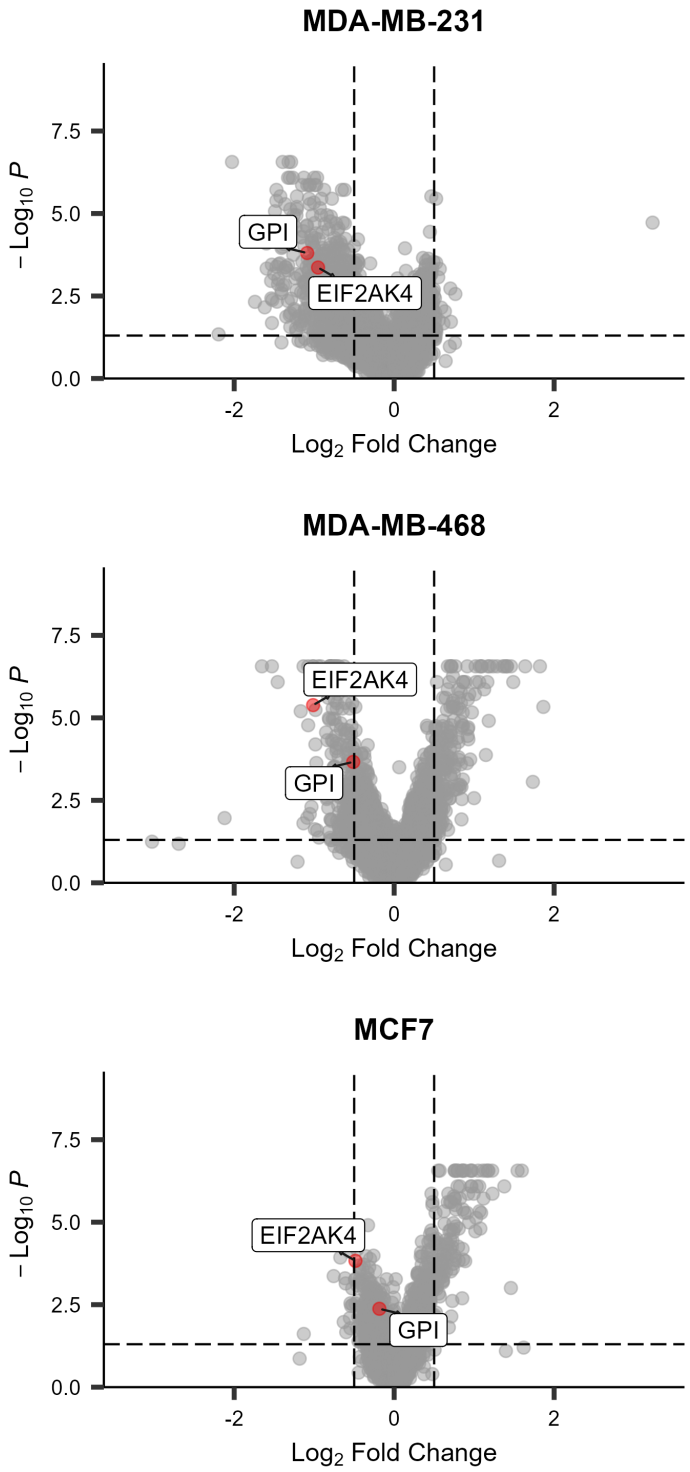

Figure 70: *EIF2AK4* and *GPI*. Results of the T14 normoxia vs T14 hypoxia comparison with *EIF2AK4* and *GPI* highlighted.

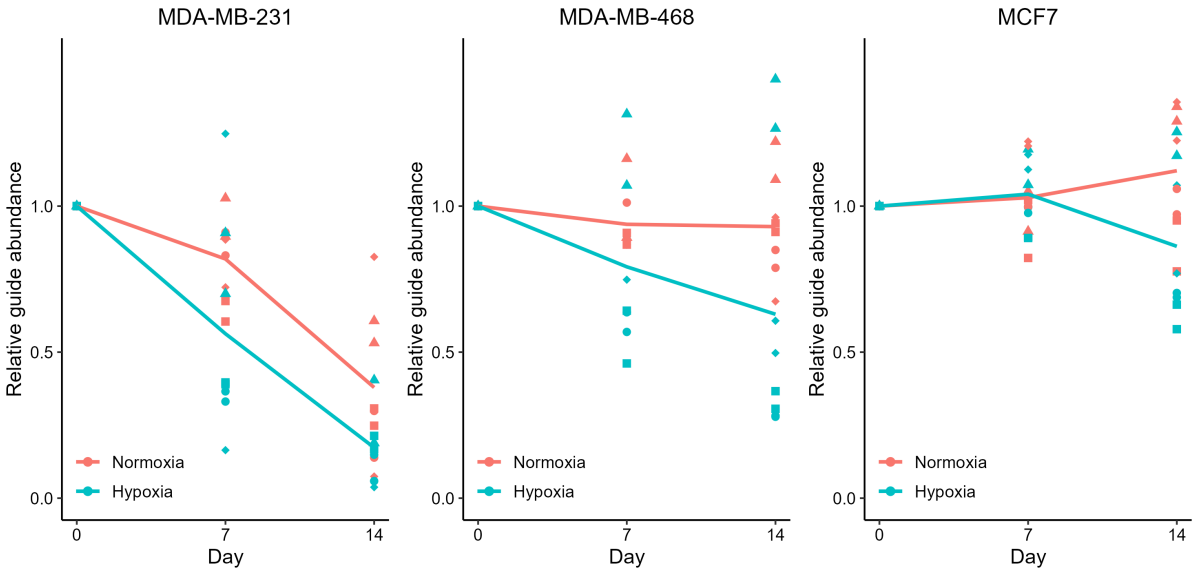

Figure 71: Depletion of *EIF2AK4* knockout cells over time. The abundance of each *EIF2AK4* targeting guide relative to T0. Both replicates are displayed, and the shapes of the symbols represent each unique guide. This allows guides which have little effect, and therefore may have low guide efficiency, to be identified. Points are connected from the mean at each timepoint.

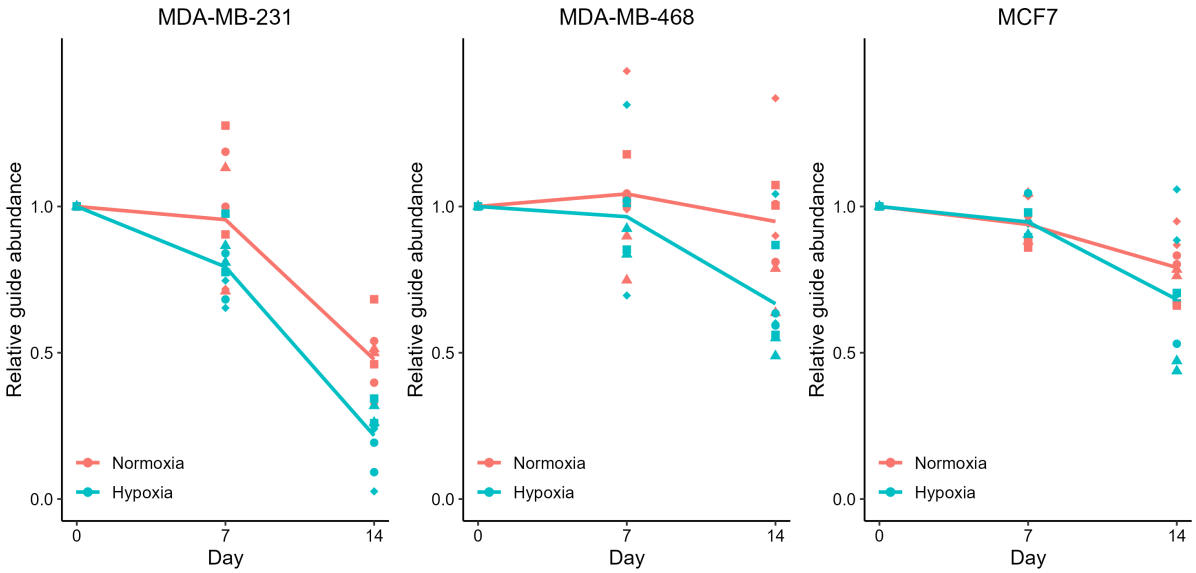

Figure 72: Depletion of *GPI* knockout cells over time. The abundance of each *GPI* targeting guide relative to T0. Both replicates are displayed, and the shapes of the symbols represent each unique guide. This allows guides which have little effect, and therefore may have low guide efficiency, to be identified. Points are connected from the mean at each timepoint.

Because *EIF2AK4* was significantly more essential in hypoxia in the two TNBC cell lines, had a trend towards the same result in MCF7 cells, was somewhat supported by an independent published CRISPR screen, and has a plausible biological mechanism, this gene was chosen for validation. As there are three other eIF2α kinases that activate the ISR pathway (*EIF2AK1*, *EIF2AK2* and *EIF2AK3*), the effect of their knockout on hypoxia specific growth/survival was assessed. Despite activating the same pathway, none of these genes showed any significant effect on hypoxic survival in the CRISPR screen (Figure 73).

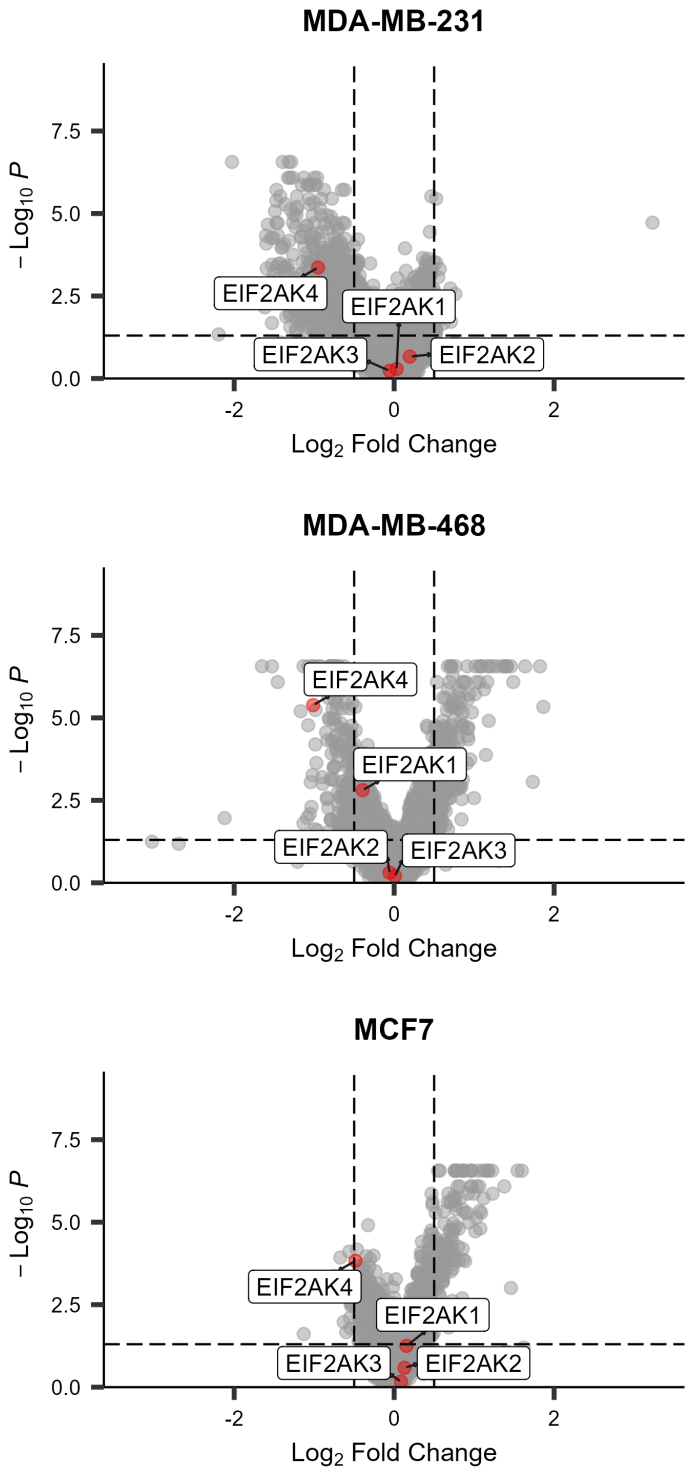

Figure 73: Selection of the eIF2α kinase genes in T14 normoxia vs T14 hypoxia. Results of the T14 normoxia vs T14 hypoxia comparison with genes encoding all eIF2α kinases highlighted (*EIF2AK1*, *EIF2AK2*, *EIF2AK3*, and *EIF2AK4*).

5.3.7 GCN2 was not differentially expressed under hypoxia

EIF2AK4 is more commonly known as GCN2 and will be referred to as this for the rest of the chapter. The first step in validating GCN2 as essential for survival under hypoxia was to establish that it is typically expressed in the cell lines used. Analysis of RNA-seq data (Section 4) showed that GCN2 was well expressed in all cell lines (median reads per million: MDA-MB-231 = 62, MDA-MB-468 = 64, HCC1806 = 187, MCF7 = 50). This is around the 80th percentile of all genes expressed in the cells, meaning GCN2 is highly expressed in these breast cancer cell lines. GCN2 was not significantly differentially expressed in any of the hypoxia conditions compared to normoxia. However, GCN2 activity is predominantly controlled at the post-translational level through phosphorylation [57]. This means GCN2 could more active under hypoxia, but it would not be reflected at the transcript level.

5.3.8 GCN2 is activated under chronic hypoxia

Hypoxia is not known to trigger activation of GCN2, with *EIF2AK3*, otherwise known as PERK, normally associated with the hypoxia driven ISR [55, 56]. Thus, a key step in validating GCN2 as a hypoxia specific survival gene was to determine whether or not it is active under hypoxia. Western blot analysis confirmed expression of GCN2 under all conditions, with some cell line differences observed (Figure 74).

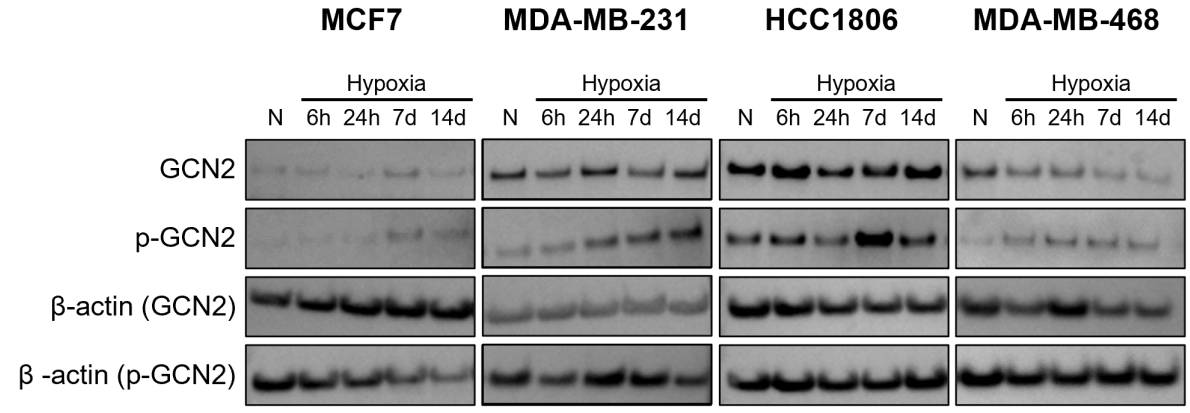

Figure 74: GCN2 and p-GCN2 kinetics under chronic hypoxia. Western blots showing GCN2 and p-GCN2. GCN2 and p-GCN2 were run on different gels, but come from the same protein preparation. Using two gels meant a separate β-actin control was produced for each. This western blot is one of three biological replicates.

MDA-MB-231 and HCC1806 cells had higher baseline levels of GCN2 than MCF7 and MDA-MB-468 cells. Western blot also showed presence of the active form of GCN2 (p-GCN2), which is phosphorylated at the T899 residue, in all cell lines. Phosphorylation of GCN2 was significantly increased under chronic hypoxia in MDA-MB-468 cells (Figure 75). There was a trend towards increased phosphorylation under hypoxia in MCF7 and MDA-MB-231 cells, but this was not significant. HCC1806 express GCN2 at a high level and the relative amount of p-GCN2 appears stable between normoxia and hypoxia.

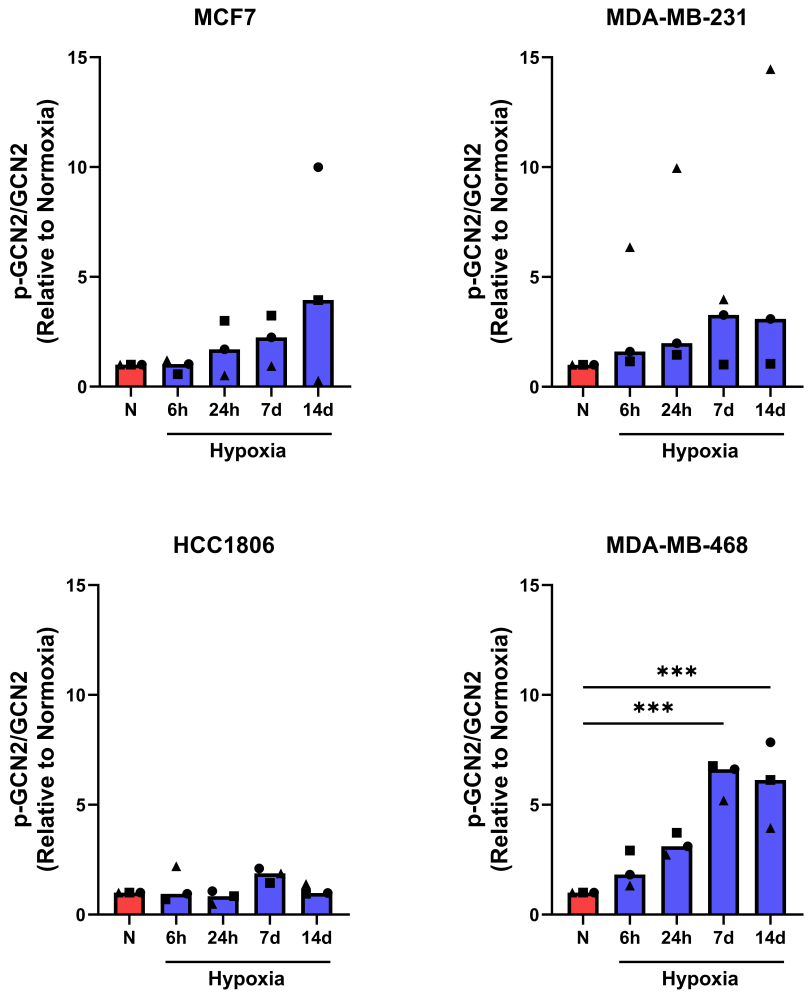

Figure 75: Relative abundance of p-GCN2 under hypoxia. The abundance of p-GCN2 compared to GCN2 after different hypoxia exposure times *(n = 3)*. Each replicate is represented by a different shape. Results are normalised to the normoxia control. Significance was tested using repeated measures one-way ANOVA and *p* values were corrected for multiple testing using Dunnett's method (*** $p \leqslant 0.001$).

5.3.9 Hypoxic cells are more sensitive to chemical inhibition of GCN2

To test whether hypoxic cells were more reliant on GCN2 activity than their normoxic counterparts, cells were treated with a chemical inhibitor of GCN2 (GCN2i) [174]. Western blot analysis confirmed that 0.5 µM GCN2i prevented phosphorylation of GCN2. This value aligns with concentrations used by Nakamura et al. [174], the developers of the inhibitor. In all cell lines, p-GCN2 was reduced to non-detectable levels after treatment with GCN2i, confirming the efficacy of the inhibitor (Figure 76).

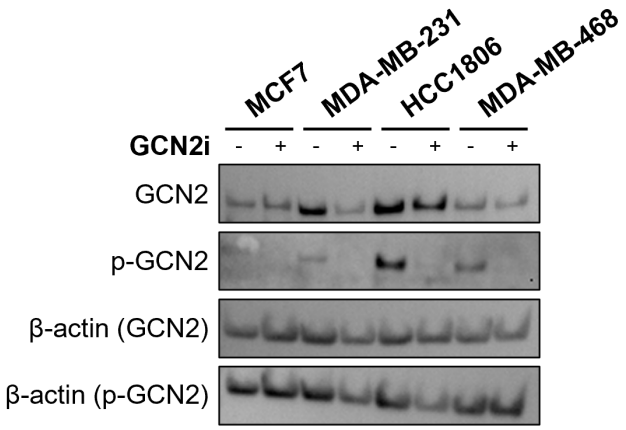

Figure 76: Levels of GCN2 phosphorylation after treatment with GCN2i. Western blots showing GCN2 and p-GCN2 with and without GCN2i treatment (0.5 µM). Cells were treated and cultured in hypoxia for 6 days. GCN2 and p-GCN2 were run on different gels, but using the same protein preparation. Using two gels means a separate β-actin control is produced for each. This figure shows one of two biological replicates.

Next, cells were treated with GCN2i and cultured in normoxia or hypoxia for 6 days. The change in cell confluency between day 0 and day 6 was measured using the Incucyte S3. MDA-MB-231 and HCC1806 cells demonstrated reduced cell growth when treated with GCN2i under both normoxic and hypoxic conditions (Figure 77). This reflects the data seen in the CRISPR screen for MDA-MB-231 cells. The growth of MCF7 and MDA-MB-468 cells was not affected by GCN2i under normoxia, but a small reduction in growth was seen with GCN2i treatment under hypoxia that reached statistical significance in MCF7 cells (Figure 77). When normalising to normoxia, all cell lines appear to be somewhat more sensitive to GCN2i treatment under hypoxia, although this is only statistically significant in MDA-MB-231 cells (Figure 78).

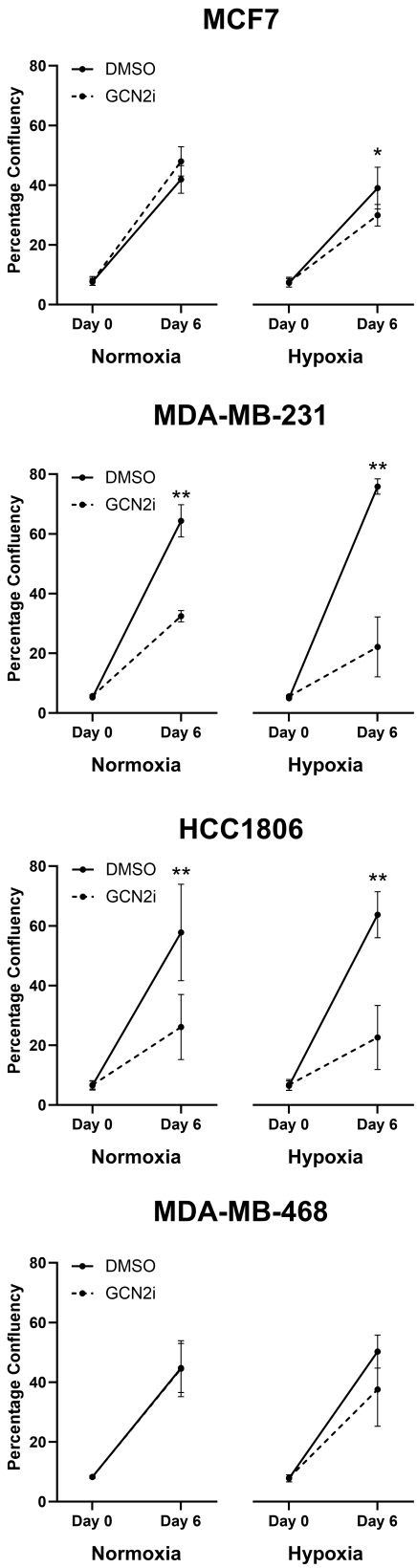

Figure 77: Growth in normoxia and hypoxia with GCN2i treatment. Percentage confluency at day 0 and day 6 after exposure to 0.5 µM GCN2i in normoxia and hypoxia *(n = 3)*. Significance was tested at day 6 using paired t-tests (* $p \leqslant 0.05$, ** $p \leqslant 0.01$). Error bars show standard deviation.

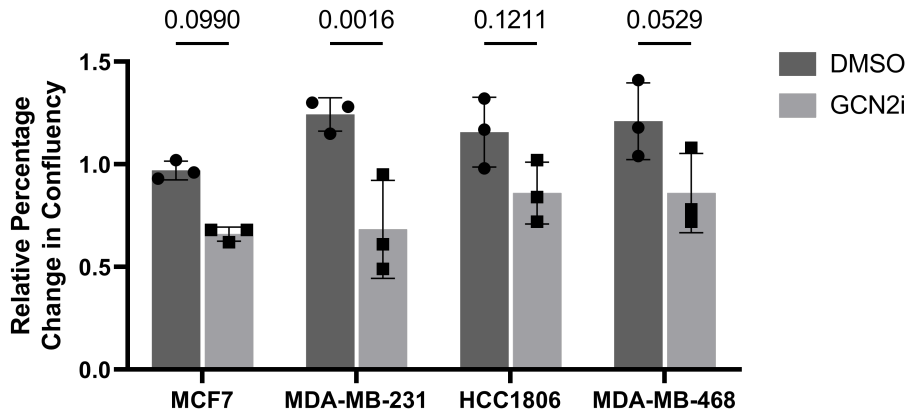

Figure 78: Percentage change in confluency relative to normoxia after GCN2i treatment. Percentage change in confluency between day 0 and day 6 after exposure to 0.5 µM GCN2i. Data were normalised to their normoxic counterpart. Significance was calculated using two-way ANOVA and corrected for multiple testing using Sidak's method *(n = 3)*. Error bars show standard deviation.

5.3.10 GCN2 knockdown did not cause a significant reduction in growth or survival under hypoxia

Given the promising results using GCN2i, the cell lines were engineered to express short hairpin RNAs (shRNAs) targeting GCN2. Two shRNAs were used for each cell line (sh850 and sh851). Multiple MOIs were tested (Figure S12). MOI 2 was identified as optimum as it did not reduce growth in normal cell culture but still achieved knockdown (Figure 79).

The knockdown cells did not show a significant decrease in growth in either normoxia or hypoxia (Figure 80) and therefore did not cause a hypoxia specific decrease in growth (Figure 81). In fact, HCC1806 cells appeared to grow comparatively faster in hypoxia when GCN2 was knocked down, and this reached statistical significance in the sh851 knockdown.

As there was no significant decrease in growth under hypoxia in GCN2 knockdown cells, the viability of the cells at the end of the experiment was measured. In accordance with the growth curves, there were no significant differences in the percentage of dead cells between control and GCN2 knockdown cells (Figure 82).

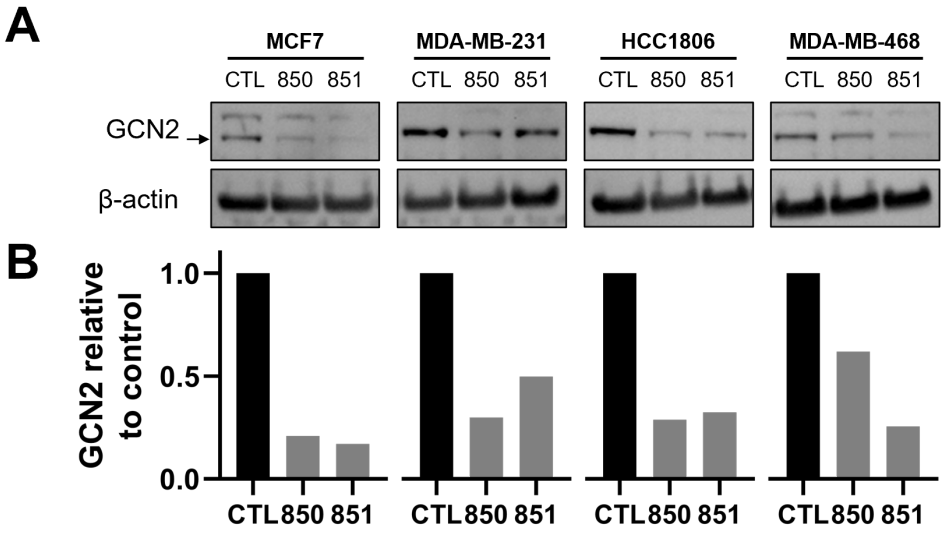

Figure 79: Expression of GCN2 in GCN2 shRNA knockdown cells. (A) Western blots showing GCN2 in control and shRNA knockdown cells (MOI 2). sh851 appeared to be more effective than sh850 at knocking down the protein. **(B)** Quantification of GCN2 relative to β-actin *(n = 1)*. CTL = control. 850 = cells transduced with sh850. 851 = cells transduced with sh851.

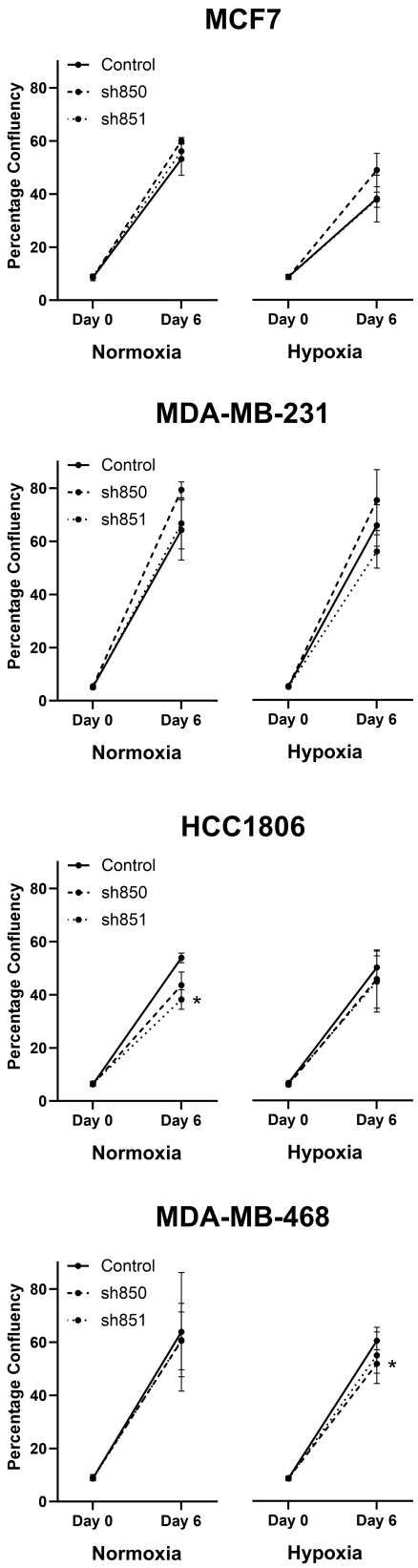

Figure 80: Growth of GCN2 knockdown cells in normoxia and hypoxia. Only minor differences in growth are observed between knockdown and control cells. Significance was tested at day 6 using paired t-tests *(n = 3)* (* $p \leqslant 0.05$). Error bars show standard deviation.

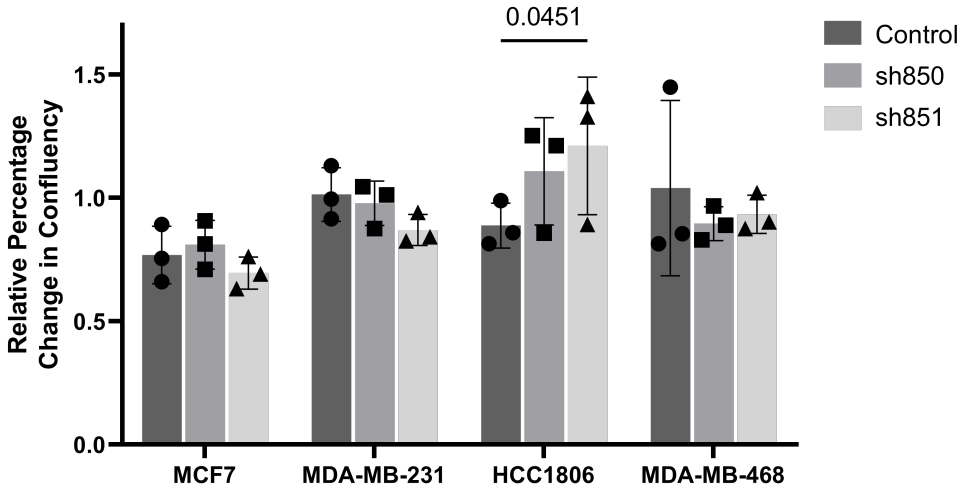

Figure 81: Percentage change in confluency relative to normoxia in GCN2 knockdown cells. Percentage change in confluency between day 0 and day 6 in GCN2 knockdown cells. Data are normalised to their normoxic counterpart. Significance was calculated using two-way ANOVA and corrected for multiple testing using Sidak's method *(n = 3)*. Error bars show standard deviation.

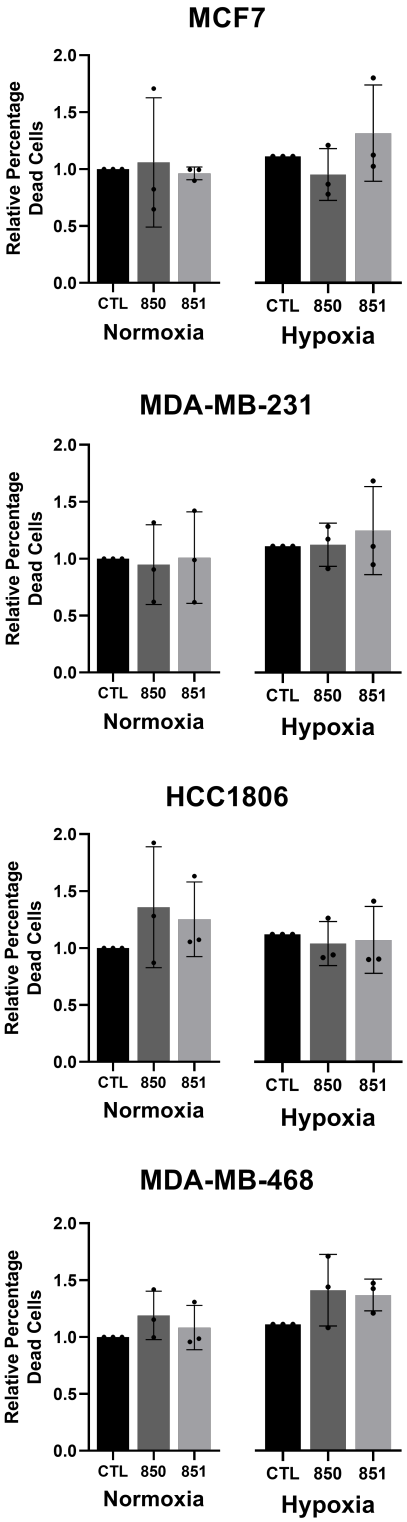

Figure 82: Relative percentage of dead cells in GCN2 knockdown cell lines. The percentage of dead cells at day 6 in GCN2 knockdown cells (850 = cells transduced using sh850, 851 = cells transduced using sh851) relative to control (CTL). There were no significant differences between control and knockdown cells in any comparison when tested using one-way ANOVA and corrected for multiple testing using Dunnett's method *(n = 3)*. Error bars show standard deviation.

5.4 Discussion

The data here suggest that hypoxia response genes, though upregulated, are not essential for cell survival under hypoxia. This is in agreement with two published hypoxia CRISPR screens which found no relationship between HIF targets and their ranking in the results [118, 178]. Knockout of HIF subunits also did not result in a significant reduction in growth/survival under hypoxia. Though surprising, this also agrees with previous research. Jarman et al. [111] found that MCF7 cells were not sensitive to HIF2 knockdown, while Thoren et al. [179] showed that HIF1 knockdown did not reduce growth or survival under hypoxia in small cell lung carcinoma cells. In addition to this, the chronic hypoxia response genes identified in Section 4 were also not essential for survival under hypoxia in this study. Though the results are in agreement with previous studies, the lack of essentiality seen here is not sufficient to conclude that the HIF proteins or their response genes are not required. This is because the ability of the guides to knockout their target genes was not validated in this study or during library development. Thus, poor guide efficiency could be the cause of the negative results observed.

Results of the CRISPR screen instead indicated GCN2 as a promising hypoxia specific essential gene. Despite not being upregulated transcriptionally, my data show an increase in GCN2 phosphorylation, and therefore activation, under hypoxic conditions. The effect was cell line dependent however, with HCC1806 cells having a high baseline level of GCN2 phosphorylation and not showing an increase under hypoxia. Though the ISR is known to be activated by hypoxia, the literature usually reports PERK as being the instigator under these conditions [63, 180, 181]. Early studies linking PERK and hypoxia use extremely low levels of oxygen (0% O_2 [180]; 0.02% O_2 [181]). PERK signalling has been shown to aid tolerance to hypoxia, but again these experiments used very low levels of oxygen ($\leqslant$ 0.2% O_2) [63, 182]. Here, I show that at least some of the ISR triggered by hypoxia comes through GCN2 signalling. This supports research by Liu et al. [51] who showed that eIF2α was phosphorylated under hypoxia even in PERK knockout cells, and that GCN2 knockout cells had reduced levels of phosphorylated eIF2α under hypoxia. Furnish et al. [183] showed correlation of p-GCN2 and HIF1α in xenografts, indirectly linking GCN2 activation and hypoxia. On the other hand, these data disagree with research by Ye et al. [55], who showed no accumulation of eIF2α in their PERK knockout cells under hypoxia. Differences in experimental setup could explain these contradictions. Ye et al. [55] used 0.2% O_2, which is more extreme than the

0.7% O_2 used by Liu et al. [51] and the 1% O_2 used in my study. Both of these studies used much shorter timepoints for assessing eIF2α phosphorylation during hypoxia exposure, with their maximum exposure times reaching only 6-8 hours. In my study, p-GCN2 did not significantly accumulate until 7 and 14 days of hypoxia exposure. Additionally, both of these studies looked at phosphorylated eIF2α as a proxy for GCN2 activity in either PERK or GCN2 knockout cells. Observation of p-GCN2 allows us to directly monitor its active levels under the chosen experimental conditions. It is possible that in the study by Ye et al. [55], GCN2 is being activated under hypoxia, but it has not yet caused a detectable increase in phosphorylated eIF2α. Further work is needed to elucidate the role of GCN2 under hypoxia, which may be relevant under less extreme levels of hypoxia than PERK.

My data also show that CRISPR knockout or chemical inhibition of GCN2 reduces cell growth in a hypoxia dependent manner (Figures 70 and 78). The CRISPR data are in agreement with a published hypoxia CRISPR screen [118] where GCN2 was ranked as the 135th hit (adjusted *p* value = 0.17) when comparing normoxia and 1% O_2. In the case of GCN2 chemical inhibition, all cell lines appeared to be somewhat more sensitive under hypoxic conditions. MDA-MB-231 and HCC1806 cells were also sensitive to GCN2 inhibition in normoxia (Figure 77). This is in accordance with the CRISPR data, where normoxic MDA-MB-231 cells were sensitive to GCN2 knockout, though to a lesser extent than the hypoxic cells (Figure 71). Interestingly, the sensitivity of cells to GCN2 perturbation in normoxia seems to correlate with GCN2 expression and p-GCN2 levels at baseline. GCN2 is more highly expressed in MDA-MB-231 and HCC1806 cells (Figure 74), and these cells are the most sensitive to GCN2i treatment under normoxic conditions. These cells also show some degree of GCN2 phosphorylation in normoxia. Conversely, MCF7 and MDA-MB-468 cells, which have lower baseline levels of GCN2 and less p-GCN2 under normoxia, have a more hypoxia driven sensitivity to GCN2i.

GCN2 knockdown did not produce the same results as knockout or chemical inhibition, with GCN2 knockdown only causing a reduction in survival under hypoxia in MDA-MB-468 cells using sh850 (Figure 80). The differences in response could be explained by the different levels of perturbation achieved by each strategy. The CRISPR screen is expected to completely knockout GCN2 in the cells expressing GCN2 targeting sgRNAs. This would completely abolish activity of GCN2. On the other hand, knockdown does not cause full depletion of the protein, and western blots of the knockdowns show residual GCN2 is present (Figure 79).

It is possible that the knockdowns did not deplete GCN2 enough to have an effect on growth over a 6 day time period. The levels of p-GCN2 were not tested in the knockdown cells. Because only a portion of GCN2 is phosphorylated under hypoxia, even though GCN2 levels had dropped, the amount of p-GCN2 in the cell could have remained stable. Given that cells were more sensitive to GCN2i, which did abolish p-GCN2 (Figure 76), this could explain why GCN2 knockdown did not reduce growth under hypoxia. Additionally, GCN2 has a short half-life which is thought to be mediated via ubiquitin ligation and proteasomal degradation [57]. However, p-GCN2 is protected from this pathway [57]. This means a small amount of remaining GCN2 caused by incomplete knockdown could, through this mechanism, still cause enough p-GCN2 accumulation to propagate downstream effects.

GCN2 has been linked to prognosis [184, 185], and several studies have noted increased expression of GCN2 in tumours compared to normal tissue [55, 184, 186]. A recent review suggested that "GCN2 may be aberrantly overexpressed and is responsible for keeping cells on 'life-support' while tumours undergo significant nutritional stress during tumourigenesis" [57]. This references amino acid starvation, the canonical activator of GCN2. However, hypoxia is likely to coincide with amino acid starvation *in vivo* due to poor vasculature, which would starve tumour cells of essential amino acids as well as oxygen. These factors, along with the results presented in this chapter, put forward GCN2 as a possible target in cancer therapy. GCN2 knockout mice are developmentally normal and are healthy unless fed a diet restricted in essential amino acids [55, 187], meaning targeting this protein would be unlikely to have severe effects on healthy tissue. In accordance, the developers of the GCN2 inhibitor found that GCN2i did not reduce viability unless paired with amino acid starvation [174]. This agrees with research by Liu et al. [51] whose GCN2 knockout cells had the same viability as wild-type cells under normoxia. This is in contrast to ATF4 knockout mice who display severe developmental abnormalities [188], suggesting that drugs targeting ATF4, a downstream activator of the ISR, would have more severe side effects. Systematic analysis of DepMap data showed that 13% of cancer cell lines are dependent on GCN2 [189], whereas very few are dependent on the other eIF2α kinases. This indicates some cell type specificity in terms of GCN2 dependency, which is also reflected in my data. It would be interesting to know how many of the cell lines in DepMap would be dependent on GCN2 if the CRISPR screens were not conducted at atmospheric oxygen levels.

A confounding variable in the CRISPR screens presented here is the reduction in growth rate for cells cultured under hypoxia. Cells which proliferate faster go through more population doublings during the screen and will therefore show stronger selection effects. Although this remains an important point to be aware of, this would not affect the genes of interest in our study. This is because we are looking for guides which are depleted more in the slower growing hypoxic population than in the normoxic population. This means the true magnitude of the effect of genes identified as selectively essential under hypoxia is likely to be greater than that reported.

A question this research did not address is whether the GCN2 activation observed during hypoxia was indirectly caused by amino acid starvation. Although this was not tested, the cells in hypoxia were cultured in an excess of media during the screen (50 mL per T175 flask) and the media was changed at least every 48 hours. This makes it unlikely that amino acids could be depleted, especially since the cells had a lower growth rate under hypoxia. Additionally, Liu et al. [51] saw enhanced phosphorylation of eIF2α after only 4 hours of hypoxia exposure in their PERK knockout cells. It is unlikely that amino acids could be depleted to such an extent within 4 hours as to activate GCN2. However, further work would need to be conducted in my experimental setting to show whether depletion of any amino acids occurred. Mass spectrometry of media collected just before a media change would allow amino acid levels to be determined.

Future work should focus on elucidating the mechanism by which GCN2 could protect cells under hypoxia. Ye et al. [55] showed that GCN2 knockout cells had lower levels of LC3-II during glutamine starvation, indicating lower rates of autophagy. Hypoxia is one of the main stimulators of autophagy, and hypoxic cells have been shown to be more sensitive to autophagy inhibitors [190]. MDA-MB-231 cells which had been adapted to chronic hypoxia demonstrated increased levels of autophagy, indicating this process as an important survival mechanism [109]. If GCN2 knockout cells are unable to switch on autophagy, this could explain the growth defects seen in hypoxia. Liu et al. [51] found that GCN2 knockout reduced the number of cells in G1 arrest under hypoxic conditions. Entering G1 arrest protects cells from entering apoptosis and can explain the acquirement of other features such as chemotherapy resistance. It also prevents cells from overshooting their metabolic capacity. In concordance, the researchers found that GCN2 knockouts had higher levels of apoptosis under hypoxia. Thus, GCN2 activity

may aid survival under hypoxia by pushing cells towards G1 arrest.

GCN2 has also been shown to mitigate oxidative stress, which is known to be increased under hypoxic conditions [62, 191]. Wang et al. [62] showed that GCN2 was phosphorylated when cells were exposed to oligomycin, an ATP synthase inhibitor that disturbs mitochondrial function, and that the GCN2 phosphorylation was ROS dependent. Active GCN2 upregulates xCT (cystine/glutamate transporter) through ATF4. xCT imports the cystine required for the production of glutathione, an important antioxidant which helps to counter the oxidative stress. In my system, GCN2 may be activated by hypoxia induced ROS. However, analysis of the RNA-seq data from Section 4 showed that, despite increased ATF4 activity, xCT expression was not upregulated under hypoxia in any of the cell lines.

One unusual result from this study is the high abundance of mitochondria associated genes which are selectively essential under hypoxia in MDA-MB-231 cells. Both MCF7 and MDA-MB-468 cells displayed greater reliance on mitochondrial genes under normoxia, which is in line with published data [118, 178]. This type of cell line specificity for CRISPR screens has been noted by other researchers. Both Thomas et al. [178] and Bao et al. [192] found differences between the cell lines that they studied, and Thomas et al. [178] noted the lack of overlap between their study and previous screens.

There is a small amount of literature supporting the unexpected results in MDA-MB-231 cells. Bao et al. [192] conducted a genome-wide hypoxia CRISPR screen and through GSEA found that oxidative phosphorylation genes were selectively essential under hypoxia (7 days at 1% O_2). The main hit from their screen is protein tyrosine phosphatase mitochondrial 1 (PTPMT1). This is also a significant hit when comparing T14 normoxia and T14 hypoxia in my MDA-MB-231 screen (adjusted p value = 0.0015, 9th ranked hit). However, this was not the case in MDA-MB-468 or MCF7 cells. Data from Bao et al. [192] suggest that knockout of PTPMT1 increases ROS accumulation during hypoxia by causing a deficiency in cardiolipin, a mitochondrial lipid . This results in a poorly organised mitochondrial membrane which allows for more electron leakage and results in oxidative stress. They validate their findings by treating several cell lines with a PTPMT1 inhibitor. MDA-MB-231 cells were part of this panel and were found to be more sensitive to PTPMT1 inhibition under hypoxia. Although Thomas et al.'s [178] overall findings showed that silencing of mitochondrial complex II gene *SDHC* improved

growth under hypoxia, one of the four cell lines they tested demonstrated the reverse effect. When *SDHC* was knocked down in the colon carcinoma cell line HCT116, cell growth under hypoxia was reduced. This cell line had a higher basal rate of oxidative phosphorylation than the other cell lines tested, which could be the cause of the differences seen. Taken together, these findings strongly suggest that some cell lines do have a greater reliance on effective mitochondrial function under hypoxia, contrary to popular belief. However, it is clear that this is not relevant in all systems. More work is needed to identify common characteristics between cells which are sensitive to mitochondrial perturbation under hypoxic conditions, as well as to untangle the role of GCN2 in adaptation to hypoxia.

6 A microRNA-ome wide CRISPR screen

6.1 Introduction

There are over 2000 mature miRNAs in the human genome, and it is predicted that over 60% of mRNAs are under miRNA control [44]. This means that altering miRNA expression can have widespread effects on a cell or organism. Unsurprisingly, many miRNAs are essential in development and for key processes such as EMT [82, 193, 194]. There are also miRNAs which are up or downregulated in tumours compared to normal tissue and are known to drive the hallmarks of cancer [195–197]. As such, many miRNAs have prognostic capacity [198]. There is also evidence of miRNA dependency in some cancer types. For example, knockdown of miR-21, and knockout of miR-23b and miR-27b, reduced the growth of breast cancer cells both *in vitro* and *in vivo* [199–201].

Many miRNAs are differentially expressed under hypoxia [40], and there is evidence that manipulating miRNA expression can affect cell survival specifically under these conditions. For example, miR-210 is strongly upregulated under hypoxia [43]. Inhibition of miR-210 increased apoptosis under hypoxia in MDA-MB-231 cells and reduced their tumour forming capacity *in vivo* [202]. This is in agreement with a study by Favaro et al. [203] who showed that miR-210 antagonism increased apoptosis in breast and colorectal cell lines under hypoxia. Similarly, knockdown of miR-210 increased the radiosensitivity of hepatocarcinoma cells under hypoxia [204], while inhibition of miR-155 increased the radiosensitivity of lung cancer cells cultured under hypoxia [205]. miR-155 was also associated with chemoresistance in breast cancer which was conferred through stimulation of EMT [206]. Despite evidence from small scale studies, the effect of miRNA inhibition under hypoxia has not been systematically investigated.

It has been proven that CRISPR-Cas9 can be used to knockout miRNAs with a high level of specificity. This drastically improves upon previous methods of miRNA inhibition, which involve miRNA sponging or the use of anti-miR oligonucleotides [207]. miRNA sponges contain multiple miRNA binding sites and thus compete with target mRNAs for binding, while anti-miRs are short oligonucleotides complementary to the miRNA which inhibit their function through steric interference [207]. Because there can be high levels of sequence homology within a miRNA family, these tools offer less specificity than targeted knockouts. They are also generally

more short-lived, and knockdown is more likely to be incomplete. CRISPR-Cas9 mutates the genomic sequence of the miRNA causing changes to the RNA secondary structure. This can disrupt miRNA biogenesis by altering Drosha or Dicer binding sites, preventing production of the mature miRNA (Figure 83). Even small, 1-2 bp insertion/deletion events can impede this process [207].

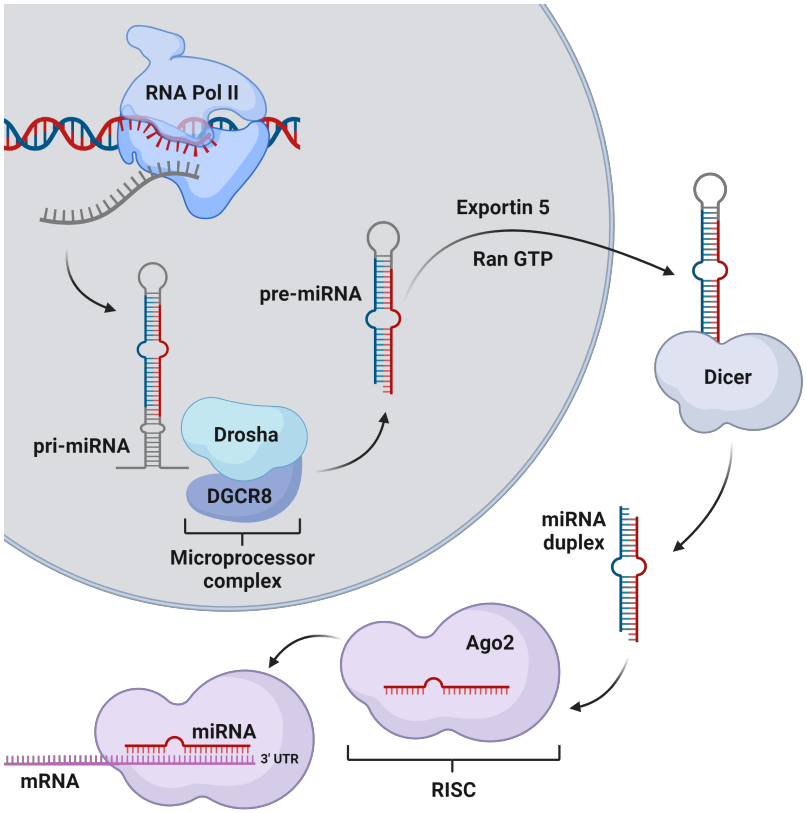

Figure 83: Canonical miRNA biogenesis and function. Primary miRNA (pri-miRNA) is transcribed by RNA polymerase II. This is processed by the microprocessor complex, comprised of Drosha and DGCR8, leaving an approximately 70 nucleotide hairpin structure known as the precursor miRNA (pre-miRNA) [97]. Pre-miRNA is exported from the nucleus and processed by Dicer to remove the hairpin loop [41, 97]. One strand from the mature miRNA duplex is loaded into the RNA-induced silencing complex (RISC). As each strand is comprised of a different sequence, its target mRNAs differ. For some miRNAs, both the -3p and -5p arms persist to equal levels, but in most cases one of the strands dominates. RISC binds to specific mRNA molecules depending on base pairing between the miRNA and the 3' untranslated region (UTR) of the mRNA [41]. Perfect base pairing is only required in the seed region of the miRNA (nucleotides 2-8) [42]. When the RISC binds to an mRNA molecule in this manner, it either suppresses translation and prevents the protein product being formed, or it destabilises the mRNA through deadenylation and decapping resulting in mRNA degradation [42]. Figure created using BioRender.

To my knowledge, there is only one commercially available library which targets exclusively miRNAs and it was designed by Kurata and Lin in 2018 [97]. Although the GeCKO v2 library targets miRNAs, it does this alongside the protein-coding genome meaning the library is very large [208]. Having a smaller, targeted library allows greater coverage to be achieved while keeping cell numbers manageable. Additionally, the miRNA targeting sgRNAs in the GeCKO v2 library are predicted to have more off target effects, making results more difficult to interpret [97]. Kurata and Lin's library (pLX-miR) targets fewer miRNA genes as their quality filters deem some of the miRNA genes un-targetable. In total, the pLX-miR library targets 1594 out of 1881 primary miRNAs using at least 4-5 sgRNAs per miRNA gene. miRNAs whose knockout reduces growth in this TNBC model will constitute possible drug targets [197, 209]. Unlike most proteins, miRNAs can be delivered exogenously as therapeutics meaning miRNAs whose knockout increases growth (i.e. tumour suppressor miRNAs) are also of interest [210, 211]. Alongside the CRISPR screens conducted in Section 5, a miRNA-ome wide CRISPR screen was conducted on MDA-MB-231 cells under normoxia and hypoxia. This screen aimed to identify miRNAs which confer a survival advantage or disadvantage in TNBC and any miRNAs which are selectively essential under hypoxia.

6.2 Methods

6.2.1 Amplification of the pLX-miR plasmid library

Plasmid amplification and lentivirus production steps were carried out by the Virus Screening Facility, Weatherall Institute of Molecular Medicine, Oxford. The Lin Human miRNA CRISPR Knockout Library (pLX-miR) was a gift from Ren-Jang Lin (Addgene, 112200). 2 µL of the pLX-miR plasmid was added to 25 µL of Lucigen Endura electrocompetent cells in a cuvette. The mix was electroporated at 1800 V and 975 µL of SOC (super optimal broth with catabolite repression) added. The solution was transferred to an Eppendorf tube and any remaining cells washed from the cuvette using a further 1 mL of SOC. The cell solution was incubated at 37℃ for 1 hour with shaking at 700 rpm. 300 µL of the solution was diluted into 30 mL of lysogeny broth and 533 µL of this was added to thirty 15 cm 100 µg/mL ampicillin plates. The bacteria were incubated at 30℃ for 18 hours. Serial dilutions were performed to enable colony counting. 200 µL of 1:100, 1:1000 or 1:10000 cell dilutions were prepared and added to 10 cm ampicillin plates in quadruplicate. The 1:100 and 1:1000 serial dilution plates were too confluent to count. The 1:10000 plates produced the following colony numbers: 761, 711, 582, and 239. This means an average of 573.25 colonies were produced per 200 µL of 1:10000 dilution, which scales up to 57,325 colonies per 200 µL for the 1:100 dilution. This equates to 287 colonies per µL of the 1:100 dilution. 15.99 mL of cells was used in total (533 µL x 30), meaning a total of 4,589,130 colonies were produced across the 30 plates. As there are 8382 guides in the pLX-miR library, the coverage is 547-fold. This is significantly above the 50-fold coverage recommended by the pLX-miR library amplification protocol, ensuring good representation of the guides [212]. The remaining thirty plates were processed in batches of five. Each plate was scraped, washed with lysogeny broth and the solutions pooled into one centrifuge tube. The solution was centrifuged at 4000 *g* for 10 minutes and the supernatant discarded. Each pellet was processed on a QIAGEN maxi column according to the manufacturer's instructions. The eluted DNA was combined to give a total of 102 µg DNA in 300 µL of Milli-Q water. The remaining SOC solution was diluted 1:2 in 40% glycerol and stored at -80℃.

6.2.2 Production of the pLX-miR lentiviral library

1.2×10^7 HEK293T cells were seeded into a 15 cm plate in DMEM (supplemented with 2 mM L-glutamine, non-essential amino acids, P/S, and 10% FBS) and incubated overnight. 13.3 µg

of the pLX-miR plasmid was added to 1.5 mL of Opti-MEM along with the packaging plasmids (10 µg of psPAX2 and 6.7 µg of pMD2.G). psPAX2 and pMD2.G were gifts from Didier Trono (Addgene, 12260 and 12259). 30 µL of PEIpro transfection reagent was added to a second aliquot of 1.5 mL Opti-MEM and incubated at room temperature for 5 minutes. This solution was added to the plasmid mix and incubated at room temperature for a further 15 minutes. This transfection mixture was added drop by drop to the HEK293T cells and incubated for 5-6 hours at 37℃. The media containing the transfection mix was then removed and replaced with fresh DMEM. Media was collected in two harvests at 48 hours and 72 hours post transfection. The pooled media was centrifuged at 2000 *g* for 5 minutes to remove cell debris. The supernatant was passed through a 0.45 µm cellulose acetate filter and aliquots stored at -80℃. A p24 ELISA was performed using the QuickTiter Lentivirus Titer Kit which indicated a titre of $8x10^7$ transducing units (TU)/mL.

6.2.3 Determination of polybrene sensitivity

Cell Counting Kit 8 (CCK-8), a plate-based cell survival assay, was used to determine whether MDA-MB-231 cells are sensitive to polybrene, a reagent used to increase the efficacy of lentivirus infection. $5x10^3$ cells were seeded per well of a 96 well plate in 100 µL of culture media and incubated overnight. Media was removed from each well and replaced with 100 µL of culture media containing polybrene at various concentrations (1-16 µg/mL) in triplicate. 100 µL of culture media containing 16 µg/mL of polybrene was used as a blank. After 24 hours of incubation with polybrene, 10 µL of CCK-8 was added to each well (except blanks) and incubated at 37℃ for two hours. The absorbance at 450 nm was measured using a POLARstar Omega microplate reader. The media containing CCK-8 was then aspirated and replaced with 100 µL of fresh culture media, and the plate returned to the incubator. The CCK-8 assay was repeated on the following two days, providing cell count readings at 24, 48, and 72 hours following polybrene exposure. For analysis, the absorbance of the blank wells was subtracted from the experimental wells and technical replicates averaged. This assay was repeated to give three biological replicates.

6.2.4 Determination of optimum antibiotic concentrations for selection

Antibiotic kill curves were performed to determine the optimum concentration of G418 required to select Cas9 expressing MDA-MB-231 cells generated using commercially available Cas9

Lenti Particles (Merck, LVCAS9NEO-1EA), and the optimum concentration of blasticidin S (hydrochloride) required to select cells transduced with the pLX-miR library. $5x10^3$ cells were seeded per well of a 96 well plate in 100 µL of culture media and incubated overnight. Media was removed from each well and replaced with 100 µL of culture media containing the relevant antibiotic at various concentrations (G418: 200-1500 µg/mL; blasticidin 2-25 µg/mL) in triplicate. 100 µL of culture media containing the highest respective antibiotic concentration was used as a blank. After 24 hours of incubation, 10 µL of CCK-8 was added to each well (except blanks) and incubated at 37℃ for two hours. The absorbance at 450 nm was measured using a POLARstar Omega microplate reader. The media containing CCK-8 was then aspirated and replaced with 100 µL of culture media containing the appropriate antibiotic concentration and the plate returned to the incubator. The CCK-8 assay was repeated on the following two days, providing cell count readings for 24, 48, and 72 hours of treatment with each antibiotic. For analysis, the absorbance of the blank wells was subtracted from the experimental wells and technical replicates averaged. This assay was repeated to give three biological replicates.

6.2.5 Viral titrations

Viral titrations were performed to determine the functional titre of the pLX-miR lentivirus and the commercially available Cas9 Lenti Particles which were used to generate Cas9 expressing MDA-MB-231 cells. $1x10^4$ cells were seeded per well of a 24 well plate in 1 mL of culture media. After overnight incubation, serial dilutions of each virus were prepared in culture media containing 8 µg/mL of polybrene. Media was aspirated from each well and replaced with 500 µL of the appropriate virus dilution. After 24 hours, the virus was removed and fresh culture media was added to the wells. 48 hours after infection, antibiotic selection began. Cells transduced using the pLX-miR virus were selected using 15 µg/mL of blasticidin, and cells transduced using the Cas9 Lenti Particles were selected using 1 mg/mL of G418. 1 mL of culture media containing the appropriate antibiotic was added to relevant wells and replenished every 24 hours. Untreated wells received 1 mL of fresh media every 24 hours. After 72 hours of selection, cells were quantified using the CCK-8 assay according to the manufacturer's instructions. For analysis, blank readings were subtracted from the experimental readings.

For the pLX-miR lentivirus, further titrations were conducted in T175 flasks to confirm the titre

under the exact transduction conditions which were to be used in the final screen. $2x10^6$ cells were seeded into each of five T175 flasks in 25 mL of culture media and incubated overnight. Cells were infected with various volumes of pLX-miR viral solution in the presence of 8 µg/mL polybrene. The control flask was exposed to polybrene without the virus. The final volume of media during the infection stage was 20 mL. After 24 hours, the media containing the virus was removed and 30 mL of fresh media applied. After 24 hours of recovery, cells were selected using 40 µg/mL of blasticidin in 30 mL of media. The media was replenished after 48 hours, and cells were trypsinised and counted after 96 hours of blasticidin selection. The cell number in each flask was calculated as a percentage of the cell number in the control flask. This experiment was conducted in duplicate, but only one replicate was chosen to calculate the viral titre due to a greater R^2 value and lack of outliers.

6.2.6 Generation of Cas9+ MDA-MB-231 cells

To generate MDA-MB-231 cells which stably express the Cas9 enzyme, $1.1x10^5$ cells were seeded into a 6 well plate and incubated overnight. Cells were then transduced using Cas9 Lenti Particles at an MOI of 0.3. An MOI < 0.7 is recommended in the literature for generating Cas9 expressing cells [213]. MDA-MB-231 cells are easily transduced [214, 215], so a low MOI was used to prevent cells receiving multiple copies of the Cas9 enzyme, as this can inhibit cell growth [216]. 1 mL of culture media containing 8 µg/mL of polybrene was used for the transduction. After 24 hours, the virus was removed and cells allowed to recover for 24 hours. Cells were then selected using 1 mg/mL of G418 and passaged into a T75 flask when confluent. Cells infected with the Cas9 Lenti Particles also express the *neoR* gene (aminoglycoside 3'-phosphotransferase) which provides resistance to neomycin, otherwise known as G418. A control well which was treated with G418 but not exposed to the virus was used to indicate when selection was complete. Selection was complete after 7 days, but the Cas9+ MDA-MB-231 cells were maintained in media containing G418 for their remaining time in culture. Multiple cell stocks were frozen in a solution of 90% FBS and 10% DMSO after approximately one month in culture. These cells tested mycoplasma negative on the day of freezing using the MycoAlert Mycoplasma Detection Assay according to manufacturer's instructions.

6.2.7 Confirmation of Cas9 expression

To confirm Cas9 expression, samples of approximately 1 million cells were collected from transduced and non-transduced MDA-MB-231 cells. Cell pellets were re-suspended in 150 µL of RIPA buffer treated with cOmplete Mini Proteinase inhibitor cocktail and incubated on ice for 30 minutes with occasional vortexing. The solutions were centrifuged at 18000 *g* for 15 minutes at 4℃ and the supernatant containing the protein lysate was collected. To quantify the protein, the Pierce BCA Protein Assay Kit was used, and a standard curve was generated using the Bio-Rad Bovine Serum Albumin Standard Set. The protein was diluted 1:5 to bring the concentration into the range of the assay. 40 µg of protein was added to 4 µL of NuPAGE Sample Reducing Agent (10X) and 10 µL of NuPAGE LDS Sample Buffer (4X). The volume was made up to 40 µL using RIPA buffer. Samples were incubated at 95℃ for 10 minutes. The samples and 40 µL of Rainbow Molecular Weight Marker were loaded into each well of a NuPAGE 4-12% Bis-Tris Gel. The gel was run in NuPAGE MOPS SDS running buffer for 10 minutes at 80 V, then for 1 hour at 120 V. The protein was transferred using the semi-dry method (Section 2.3.7). The membrane was blocked using 5% BSA (Section 2.3.9) then probed with a mouse monoclonal Cas9 antibody diluted 1:1000 in blocking buffer. The secondary antibody (goat anti-mouse IgG HRP) was diluted 1:3000 in blocking buffer. Full details for probing the membrane are found in Section 2.3.9. The protein was visualised using the Pierce ECL Plus Western Blotting Substrate Kit and the ImageQuant LAS 4000 mini according to manufacturer's instructions. The membrane was then trimmed to remove the area containing Cas9. The remaining membrane was probed for β-actin. The primary antibody to β-actin was diluted 1:1000 in blocking buffer and incubated with the membrane on a rocking platform at room temperature for 1 hour. The membrane was washed as described in Section 2.3.9. The secondary antibody (goat anti-mouse IgG HRP) was diluted 1:2000 in blocking buffer and incubated for 1 hour on a rocking platform at room temperature. The protein was visualised as described for the Cas9 protein above.

6.2.8 Growth curve of Cas9+ MDA-MB-231 cells

$3x10^5$ Cas9+ or control MDA-MB-231 cells were seeded into each of four T25 flasks in a total of 4 mL culture media (i.e. eight flasks in total). Cas9+ and control flasks were harvested after 24, 48, 72, and 96 hours and the cell count determined using a LUNA II automated cell

counter. A media change took place at 48 hours. This experiment was repeated to produce three biological replicates.

6.2.9 CRISPR screen

The screen was carried out in triplicate, with separate infections for each replicate. This is important as the infection step is the main source of bias in CRISPR screens, so if all replicates are produced from the same infection the same bias exists across all replicates. Cas9+ MDA-MB-231 cells were thawed one week prior to beginning the screening protocol. During this time, cells were cultured with 1 mg/mL of G418 to maintain Cas9 expression. P/S was only added to the media for the last two weeks of each experiment (i.e. at T0). Figure 84 summarises the timeline of the CRISPR screen.

6.2.10 Infection and selection of cells

According to the schedule in Figure 84, $2x10^6$ Cas9+ MDA-MB-231 cells were seeded into each of sixteen T175 flasks in 30 mL of media. The following day, the culture media was removed from fifteen flasks and replaced with 20 mL of media containing 8 µg/mL polybrene and 18.65 µL of pLX-miR lentivirus. This equates to $6x10^5$ TU per flask, and an MOI of 0.3. The remaining flask received 20 mL of media containing 8 µg/mL polybrene but no virus. This served as the "no infection" control. Cells were incubated overnight to allow infection to occur. The following day, the virus containing media was removed and replaced with 30 mL of media. Cells were again incubated overnight, allowing time for integration of the viral genome and expression of those genes to occur. Media was then replaced with 30 mL of media containing 40 µg/mL blasticidin. The pLX-miR lentivirus genome contains the blasticidin S deaminase gene, which confers resistance to blasticidin. This means that cells which have been successfully infected and have integrated the pLX-miR lentivirus genome will survive blasticidin exposure, while all other cells die. Cells were incubated with blasticidin until all cells in the "no infection" control flask had died. In all replicates, this occurred on day 4 of culture. Media was replenished every 48 hours.

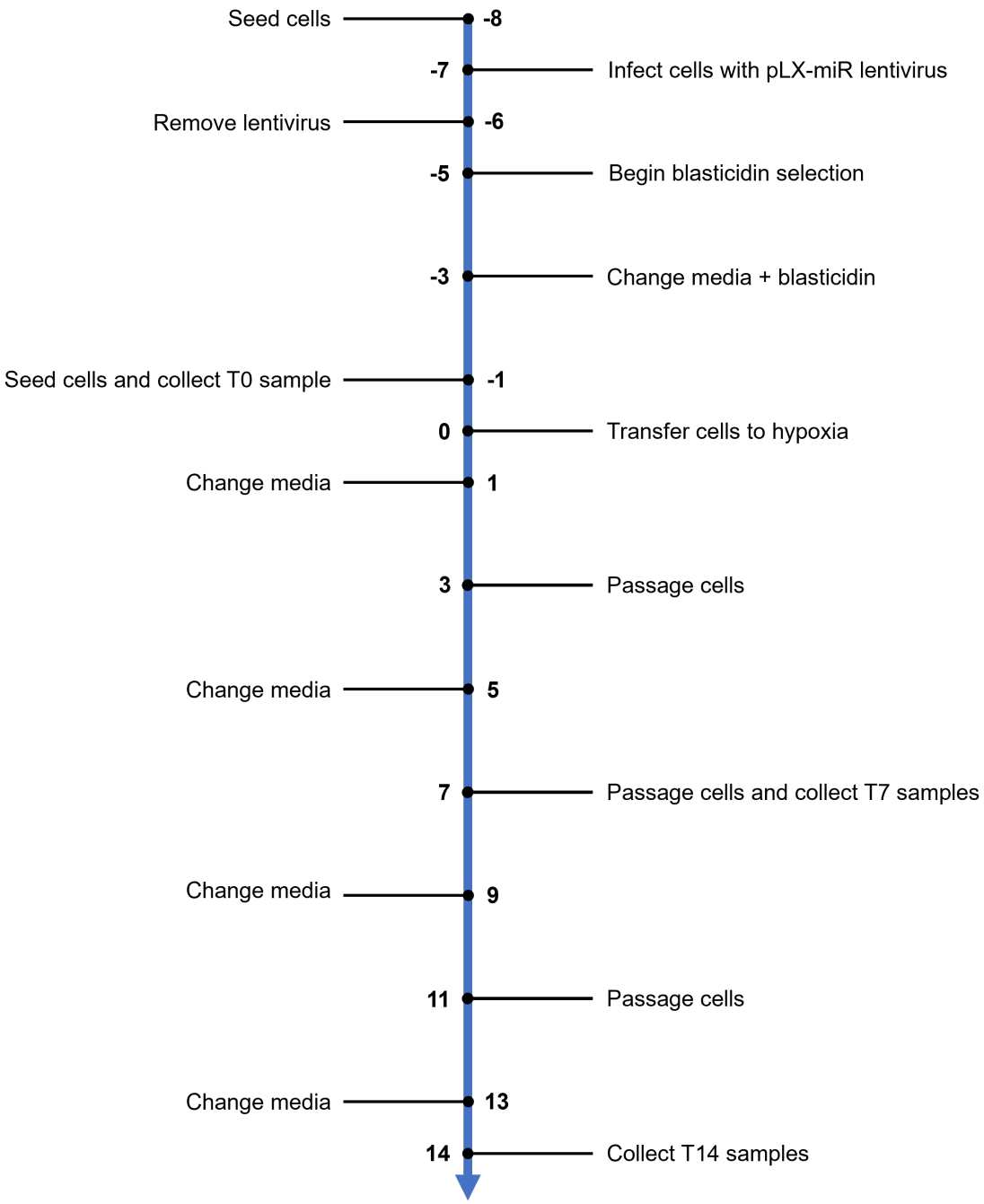

Figure 84: CRISPR screen timeline.

6.2.11 Maintenance of cells

After blasticidin selection was complete, cells from the fifteen infected flasks were harvested. Flasks were rinsed using PBS, and incubated with trypsin until the cells had detached. Media was added and the solutions from each flask were pooled into five 50 mL centrifuge tubes. The cell solutions were centrifuged at 275 *g* for 5 minutes. Each of the five cell pellets were

re-suspended in 10 mL of media and the solutions pooled to make one cell solution. An aliquot of cells was added to an equal volume of trypan blue solution. Live and dead cells were quantified using the LUNA II automated cell counter. Two cell counts were taken to ensure consistency. An aliquot of 6.29x10^6 cells was taken and processed according to Section 6.2.12. This is the T0 reference sample, and 6.29x10^6 cells represents 750x coverage. From the remaining cells, 2.5x10^6 cells were seeded into each of seven T175 flasks in 50 mL of media now supplemented with P/S. Three flasks were required to maintain coverage, and this must be maintained in each arm of the experiment (i.e. three flasks designated for normoxia and three flasks designated for hypoxia). The remaining flask was cultured alongside the hypoxia samples but was removed intermittently from the hypoxia chamber to allow observation of the cells. This flask was discarded after use so cells which had been reoxygenated were not part of the samples collected. Cells were incubated overnight to allow attachment to the flask and the following day, four of the flasks were transferred to the hypoxia chamber and were subsequently cultured at 1% O_2. The schedule shown in Figure 84 was followed for the remaining 14 days of the experiment.

On days which required a media change, the culture media was discarded and replaced with 50 mL of media supplemented with P/S. Cells in the normoxia arm were visually inspected, along with one flask from the hypoxia arm. On days which required cell passaging, cells were harvested according to the method described above. All cell handling for the hypoxia arm of the experiment occurred inside the hypoxia chamber. Cell counts were recorded and 2.5x10^6 cells were seeded per flask (three flasks in normoxia, four flasks in hypoxia). On days which required harvesting, 6.29x10^6 cells were collected and processed as per Section 6.2.12.

6.2.12 Collection of cells

6.29x10^6 cells were collected per sample and centrifuged at 275 *g* for 5 minutes. The supernatant was discarded and the cells re-suspended in 629 µL of PBS to produce a 1x10^7 cells/mL solution. The cell solution was then processed using the QIAGEN Blood & Cell Culture DNA Midi Kit. 629 µL of chilled Buffer C1 and 1887 µL of ice-cold distilled water were added to the cell solution, mixed by inversion, and incubated on ice for 10 minutes. The solution was centrifuged at 1300 *g* for 15 minutes at 4℃. The supernatant was discarded and the pellet re-suspended in 1 mL chilled Buffer C1 and 3 mL ice-cold distilled water. The solution was again

centrifuged at 1300 *g* for 15 minutes at 4℃. The supernatant was removed, and the remaining nuclear pellet was stored at -80℃ until DNA extraction could be completed. Nuclear pellets were stored instead of cell pellets as this increases the DNA yield according to the QIAGEN Genomic DNA Handbook [217].

6.2.13 Extraction of DNA

The QIAGEN Blood & Cell Culture DNA Midi Kit was used to extract DNA. Manufacturer's instructions were followed. The frozen nuclear pellets from Section 6.2.12 were thawed on ice then re-suspended in 5 mL of Buffer G2. The nuclear pellets were re-suspended by vortexing at maximum speed. Samples were vortexed for a minimum of 30 seconds, but vortexing time was extended as necessary until the pellet was no longer visible to prevent clogging of the QIAGEN tip. To each 5 mL solution, 95 µL of QIAGEN protease stock solution was added and incubated at 50℃ for 1 hour. Protease was prepared by reconstituting the lyophilised enzyme in 1.4 mL distilled water. During this time, each QIAGEN genomic tip was loaded with 4 mL of Buffer QBT which was allowed to run through into a waste container. After the incubation period, each sample was vortexed for 10 seconds at maximum speed and quickly loaded into the QIAGEN genomic tip. After the sample had fully entered the column, the QIAGEN genomic tip was washed twice using 7.5 mL of Buffer QC. Finally, the genomic DNA was eluted from the column using 5 mL of Buffer QF which was pre-warmed to 50℃. 3.5 mL of room-temperature isopropanol was added to the eluted DNA and the solution centrifuged at 7000 *g* for 15 minutes at 4℃. The supernatant was discarded, and the remaining DNA pellet washed using 2 mL of chilled 70% ethanol. The sample was vortexed briefly and centrifuged at 7000 *g* for 10 minutes at 4℃. The supernatant was removed, and the remaining DNA pellet allowed to air dry for 10 minutes. Finally, the DNA pellet was re-suspended in 300 µL of TE buffer (0.1 mM EDTA · Na2, 10 mM Tris, pH 8.0) and incubated at 55℃ for 2 hours to dissolve. DNA was quantified using a NanoDrop Spectrophotometer. 1 µL of TE buffer was used to blank the instrument, and 1 µL samples of DNA were used for quantification. Two readings were taken per sample to ensure measurement consistency and the concentration was recorded (ng/µL), along with the A260/A280 and A260/A230 ratios.

6.2.14 Amplification of guide sequences

Guide sequences were amplified from the extracted genomic DNA using a nested PCR strategy. This was used by Kurata and Lin when developing the pLX-miR library [97] and was originally adapted from Shalem et al. [218]. The primers (sequences in Table 8) used for both PCR reactions were reconstituted using an appropriate volume of DEPC treated water to bring the molar concentration to 20 µM, except for the R1 primer which was diluted to 10 µM.

Table 8: CRISPR primer sequences - pLX-miR

Name	Published Name	Sequence
F1	lentiC-dseq-2-F1	GCCTATTTCCCATGATTCCTTCA
R1	LX-dseq-R1	AGCGCTAGCTAATGCCAACTT
F2	lentiC-dseq-F2	CTACACGACGCTCTTCCGATCTTCTTGTGGAAAGGACGAAACACCG
R2	pLX-dseq-R2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCACCGACTCGG TGCCACTTTT

For the first PCR step (PCR1), the following reagents were used per reaction:

25 µL	NEBNext High-Fidelity 2X PCR Master Mix
1.25 µL	Forward F1 primer (20 µM)
2.5 µL	Reverse R1 primer (10 µM)
1 µg	Genomic DNA
__ µL	DEPC treated water to make the total reaction volume 50 µL

To maintain 750x coverage of the library, genomic DNA from $6.29x10^6$ cells must be amplified. Assuming 6.6 pg DNA per cell [218] means 41.5 µg of DNA is required to maintain representation. Therefore, 42 reactions, each containing 1 µg of DNA, were carried out per sample. PCR1 was run on a Bio-Rad C1000 Touch Thermal Cycler using the following settings:

Table 9: PCR1 settings - pLX-miR

Step	Temperature	Time
Initial denaturing	98℃	30 seconds
Denaturation*	98℃	10 seconds
Annealing*	60℃	30 seconds
Extension*	72℃	20 seconds
Final extension	72℃	2 minutes
Hold	4℃	Infinite

*30 cycles

10 µL was taken from each reaction and pooled to form one 420 µL solution. This was used as the DNA template for the second round of PCR (PCR2). 10 µL of this pooled solution was run on an agarose gel to confirm presence of a DNA product of the correct size (Section 6.2.15). A negative control containing DNA from un-transduced MDA-MB-231 cells, and a negative control containing no DNA at all, were included to confirm that the appearance of a band is specific to the CRISPR experiment. The following reagent volumes were then used for each PCR2 reaction:

25 µL	NEBNext High-Fidelity 2X PCR Master Mix
1.25 µL	Forward F2 primer (20 µM)
1.25 µL	Reverse R2 primer (20 µM)
2 µL	Pooled product from PCR1
__ µL	DEPC treated water to make the total reaction volume 50 µL

Five reactions were carried out per sample. The annealing temperature of these primers is high ($>$72℃), so the following 2-step protocol was used:

Table 10: PCR2 settings - pLX-miR

Step	Temperature	Time
Initial denaturing	98℃	30 seconds
Denaturation*	98℃	10 seconds
Annealing and extension*	72℃	20 seconds
Final extension	72℃	2 minutes
Hold	4℃	Infinite

*22 cycles

The products of the five reactions were pooled to form a final 250 µL solution. Again, a 10 µL sample of this was taken to run on an agarose gel to confirm presence of the DNA product at the correct size (Section 6.2.15). A negative control containing DNA from PCR1 of un-transduced MDA-MB-231 cells, and a negative control containing no DNA, were included to confirm that the appearance of a band is specific to the CRISPR experiment. 80 µL of PCR product was purified using a QIAquick PCR Purification Kit according to manufacturer's instructions. This removed free nucleotides and primers from the PCR products. The DNA was eluted in Buffer EB (10 mM Tris-Cl, pH 8.5) from the QIAquick PCR Purification Kit.

The final product was quantified using a Qubit 4 Fluorometer using the Qubit dsDNA High Sensitivity Assay Kit according to the manufacturer's instructions. The fluorescence reading is temperature sensitive, so all samples were incubated at room temperature for at least 5 minutes to ensure temperature equilibration before reading. 1 µL of sample was added to 199 µL of working reagent. For the standards, 10 µL was added to 190 µL of working reagent. A260/A280 and A260/A230 ratios were quantified using a NanoDrop Spectrophotometer. 1 µL of each sample was measured, with 1 µL of Buffer EB used to blank the instrument.

6.2.15 Agarose gel

PCR products were visualised using agarose gels both in the process of PCR optimisation, and to confirm the presence of bands of the correct size after PCR amplification during the CRISPR screen. 3% agarose gels were prepared using TAE buffer (40 mM Tris, 20 mM acetate, 1 mM EDTA) with 1:10000 SYBR Safe DNA Gel Stain. The expected products of PCR1 and PCR2 are 391 bp and 175 bp respectively. These are both small products so a higher than usual percentage of agarose was used to maximise their separation when running through the gel. 2 µL of DNA Gel Loading Dye (6X) was added to either 10 µL of sample or 10 µL of the Quick-Load Purple 100 bp DNA Ladder. Samples were loaded into the gel and the casing filled with TAE buffer. The gel was run at 80 V for 10 minutes, followed by 1 hour at 120 V. The gel was imaged using either a BioRad GelDoc, or an Invitrogen iBright 1500 to visualise the DNA bands. Exposure time was determined automatically by the instrument.

6.2.16 Amplicon sequencing

The purified products of PCR2 were loaded into a full skirted 96 well PCR plate. Indexing and sequencing were carried out by the Oxford Genomics Centre, Oxford. Material was quantified using Qubit and the size profile analysed on the TapeStation using the dsDNA High Sensitivity Assay. Material was normalised to 2.5 ng. Amplicon libraries were prepared without fragmentation or size selection. Libraries were prepared using an Apollo next-generation sequencing library preparation system using PrepX reagents and standard Illumina multiplexing adapters. Dual index amplification (18 cycles) was completed using the NEBNext High-Fidelity 2X PCR Master Mix and in-house unique dual indexing primers (based on Lamble et al. [115]). Individual libraries were normalised using Qubit and the size profile analysed on the TapeStation. Individual libraries were normalised and pooled together accordingly. The

pooled library was diluted to approximately 10 nM for storage. The 10 nM library was denatured and further diluted prior to loading on the sequencer along with 20% PhiX complexity spike. Paired-end sequencing was performed using the NextSeq500 platform (v2.5 kits, 150 cycles).

6.2.17 Processing of reads

FASTQ files were downloaded by file transfer protocol and md5sums were checked to ensure files were not corrupted. Technical replicates from sequencing were merged at this stage. FastQC (version 0.11.9) was used to assess the quality of the reads. At this stage, read 2 was shown to have high sequence duplication (85%) so was removed from the analysis. Cutadapt (version 2.10) was used to trim the adapter sequences and remove reads that were not 16-21 nucleotides long after trimming. The options used for cutadapt were: `--discard-untrimmed -q 20 -m 16 -M 21 -g TCTTGNGGAAAGGACGAAACACC...GTTTTAGAGCTAGAAATAGCAAGTTAAAATAA`. The pLX-miR library contains some guides which target multiple miRNAs. For instances where a guide simply targeted different copies of the same miRNA (e.g. hsa-mir-181b-1 and hsa-mir-181b-2, which produce the same mature miRNA but are found at different genomic loci) the suffix was trimmed and the resulting duplicated rows removed using a custom Python script (Section 9.5). This resolved 238 of the 427 non-uniquely targeting guides. The remaining non-uniquely targeting guides were renamed with a numerical suffix but maintained in the library file. Numerical suffixes were manually converted to alphabetical. Duplicate guide sequences are not tolerated by the MAGeCK alignment program, a published program designed specifically for CRISPR analysis [219], so instead Bowtie (version 1.2.3) was used to align the reads. A FASTA file was generated from the modified pLX-miR library file using a custom R script (Section 9.5). This was used to generate a Bowtie index. Reads were aligned against the index using Bowtie with the following options: `-a -v 0 -m 3`. These options allow a read to map against three guide sequences before it is removed. This value was chosen because it is the maximum number of times a non-uniquely targeting guide appears in the library. No mismatches were allowed during alignment so mapping is highly stringent. The alignment files were converted to binary, sorted, and indexed using SAMtools (version 1.8). The reads mapping to each guide were quantified using the idxstats function, also from the SAMtools package. To ensure this alignment and quantification method performed well, it was checked against results produced by MAGeCK (version 0.5.9.2). All non-uniquely targeting guides were removed from the

library file and sequences were aligned using MAGeCK using the `--trim-5 0,1` option. After removal of non-uniquely targeting guides from the Bowtie output, the mean Pearson correlation coefficient of counts produced by the two methods was 0.996 (Figure S13), indicating no issues with using duplicate tolerant Bowtie for mapping.

6.2.18 Analysis

Differential analysis was conducted using MAGeCK [219] using the output of the Bowtie alignment. The MAGeCK RRA (robust ranking algorithm) normalises the raw read counts to account for the effect of library size then fits a negative binomial model to test whether the counts vary between control and treated conditions [219]. The guides are ranked according to their significance and the RRA assesses whether the distribution of the ranking deviates from the null. Although there are non-targeting negative control guides in this library, which under the MAGeCK algorithm can be used to generate a null distribution, their distribution was positively skewed making them an unsuitable control. This positive skewing has been noted by other researchers [177, 220]. The ranking of the guides acts to combine the significance of all guides targeting a particular gene to output a single significance value for each gene. Significance values are adjusted for multiple testing using the Benjamini-Hochberg method [219, 221]. According to a recent review article [222], the MAGeCK RRA method is the most robust when gene effects are small and when guide efficiencies are low. The practicalities of targeting miRNAs using CRISPR mean these limitations are relevant to this study, justifying the use of this method.

6.2.19 miRNA target prediction

The targets of miRNAs were predicted using a modified version of the `getPredictedTargets` function in miRNAtap (version 1.32.0) [223]. miRNAtap searches five databases for the predicted targets of a given miRNA and integrates the results to produce a final aggregate score for each predicted target using the geometric mean. The five databases searched are: DIANA [224], miRanda [225], PicTar [226], TargetScan [227], and miRDB [228]. Full code for the modified function is found in Section 9.5. Firstly, the function was modified so miRNA names longer than traditional miRNAs were not incorrectly trimmed. Secondly, multiple changes were made to allow the minimum required sources of a miRNA target to be strictly enforced. Thirdly, edits were made so results from all sources would be output as default.

Lists of predicted targets were filtered so that only genes expressed in MDA-MB-231 cells were retained. Any gene with a mean average <1 count per sample in the RNA-seq data (Section 4) was removed from the predicted target list.

6.3 Results

6.3.1 Sensitivity to polybrene

The first step of preparing for the CRISPR screen was to ensure that MDA-MB-231 cells are not sensitive to polybrene. Polybrene is used to increase the efficacy of lentivirus infection but is toxic to some cell types. Cells were exposed to various concentrations of polybrene for 24 hours and the cell number monitored using the CCK-8 assay for the following 72 hours. The results showed that MDA-MB-231 cells were not sensitive to polybrene (Figure 85), so the widely recommended concentration of 8 µg/mL was used in all transductions going forward.

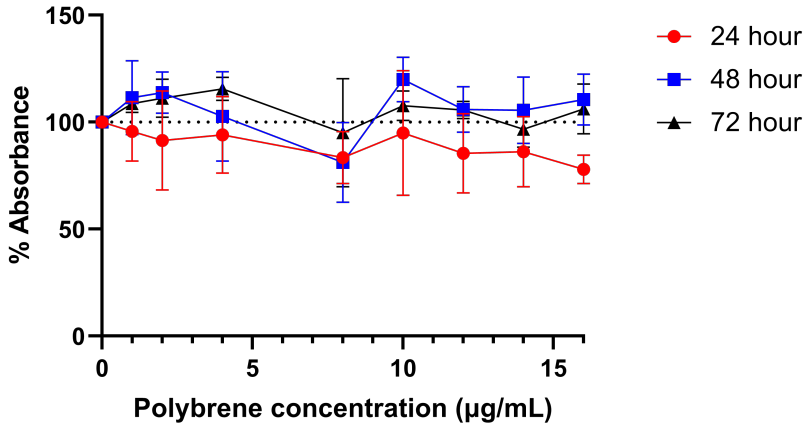

Figure 85: Polybrene sensitivity. MDA-MB-231 cells were exposed to concentrations of polybrene ranging from 1-16 µg/mL and the absorbance, which is representative of cell number, compared to that of an untreated control. Points show the mean and error bars show standard deviation *(n = 3)*.

6.3.2 Kill curves of G418 and blasticidin

The next step was to evaluate the sensitivity of MDA-MB-231 cells to the antibiotics to be used for selection. To generate Cas9 expressing cells, cells must be infected with Cas9 Lenti Particles. The selection marker for this vector is G418 resistance, therefore the concentration of G418 required to kill 100% of MDA-MB-231 cells was determined. This ensured a pure population of Cas9 expressing cells could be obtained (i.e. any cells that were not successfully transduced were killed by the G418 in the media). 1 mg/mL was identified as the optimum G418 concentration (Figure 86A).

The pLX-miR lentiviruses contain the guide sequence, along with a gene conferring resistance to blasticidin. The same experiment was conducted to determine the optimum concentration of blasticidin required to select the infected cells. Initially, 15 µg/mL was identified as the optimum blasticidin concentration (Figure 86B), and this concentration was used for the library titrations. However, this was found to be insufficient in a pilot study of the screen and the concentration was increased to 40 µg/mL. This demonstrates the importance of testing the required parameters in the exact setting to be used for the experiment.

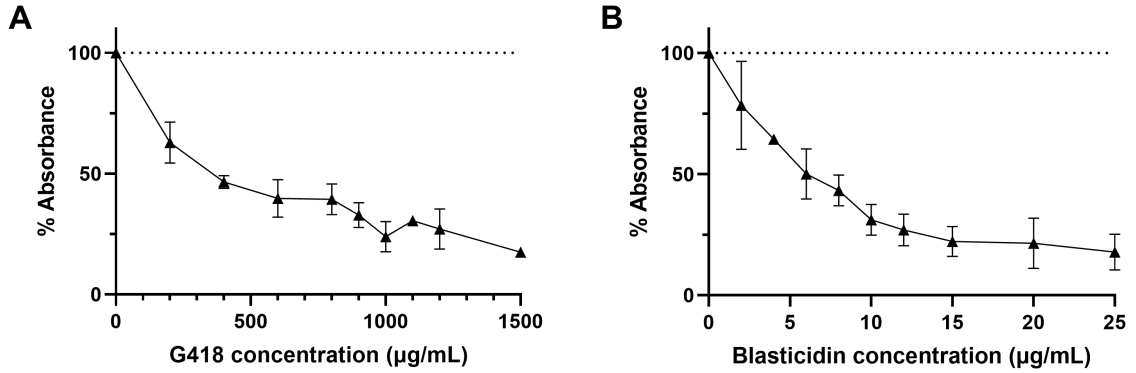

Figure 86: G418 and blasticidin kill curves. (A) MDA-MB-231 cells were exposed to concentrations of G418 ranging from 200-1500 µg/mL and the absorbance in the CCK-8 assay, which is representative of cell number, compared to that of an untreated control. There is a reduction in cell number even at the lowest concentration tested. The drop in cell number plateaus around 1000 µg/mL (0-1000 µg/mL: *n = 3*. 1100-1500 µg/mL: *n = 2*). Points show the mean and error bars show standard deviation. **(B)** MDA-MB-231 cells were exposed to concentrations of blasticidin ranging from 2-25 µg/mL and the absorbance in the CCK-8 assay, which is representative of cell number, compared to that of an untreated control. The reduction in cell number plateaus around 15 µg/mL (0-10 µg/mL: *n = 3*. 12-25 µg/mL: *n = 2*). Points show the mean and error bars show standard deviation.

6.3.3 Library titration of Cas9 Lenti Particles

Next, the viral titre of the Cas9 Lenti Particles was determined (Figure 87).

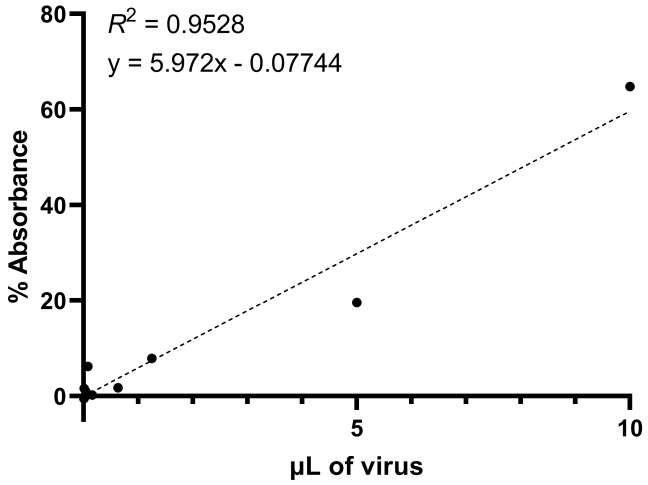

Figure 87: Cas9 Lenti Particle titration. MDA-MB-231 cells were infected with varying volumes of Cas9 Lenti Particles. Infected cells were selected using 1 mg/mL of G418 and the absorbance in the CCK-8 assay, which is representative of cell number, was calculated as a percentage of untreated controls. Simple linear regression was used to calculate the TU/mL of the lentiviral stock. Points shown are the mean of two biological replicates.

The functional titre of the virus was estimated to be 714,456 TU/mL using the results of linear regression:

$$\begin{aligned} y &= 5.972x - 0.07744 \\ x &= \frac{y + 0.07744}{5.972} \end{aligned}$$

An MOI of 0.3 results in 25% of cells being infected (i.e. 25% of cells remain alive in this assay). If $y = 25$:

$$\begin{aligned} x &= \frac{25 + 0.07744}{5.972} \\ x &= 4.199\ (\mu L\ of\ virus\ used) \\ TU/mL &= number\ of\ cells\ at\ transduction \times \frac{MOI}{mL\ of\ lentivirus\ used} \\ TU/mL &= 10000 \times \frac{0.3}{0.004199} \\ TU/mL &= 714456 \end{aligned}$$

6.3.4 Library titration of pLX-miR lentivirus

Similarly, the viral titre of the pLX-miR lentivirus was determined through titrations. The initial results of this are shown in Figure 88A.

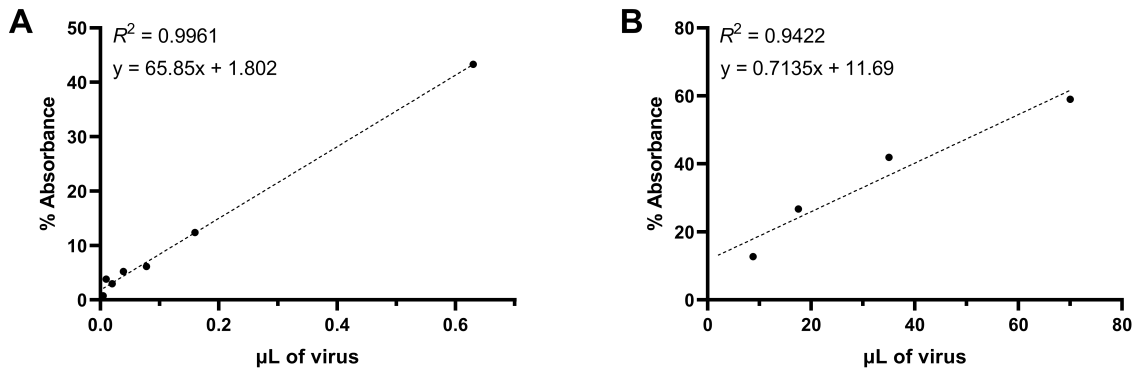

Figure 88: pLX-miR lentivirus titration. (A) MDA-MB-231 cells were infected with varying volumes of pLX-miR lentiviral stock solution in a 24 well plate. Infected cells were selected using 15 µg/mL of blasticidin and the absorbance in the CCK-8 assay, which is representative of cell number, was calculated as a percentage of untreated controls. Points shown are the mean of three biological replicates. **(B)** MDA-MB-231 cells were infected with varying volumes of pLX-miR lentiviral stock solution in a T175 flask. Infected cells were selected using 40 µg/mL of blasticidin and the cell number calculated as a percentage of the control flask *(n = 1)*.

The viral titre was estimated to be 8,515,820 TU/mL. The accuracy of this was then assessed in a T175 flask system to more be more representative of the final experimental setup. The estimated viral titre increased significantly in this experiment to 32,168,295 TU/mL (Figure 88B).

6.3.5 Generation of Cas9+ cells

Cas9 expressing MDA-MB-231 cells were generated by transducing the cells using Cas9 Lenti Particles. The viral genome integrates into the host DNA and therefore produces stable expression of the gene of interest. Cells were infected with the Lenti Particles and selected for one week using G418. Samples of cells were taken to confirm Cas9 expression (Figure 89).

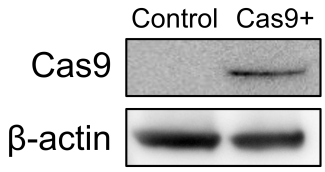

Figure 89: Cas9 expression. Western blot showing expression of Cas9 in transduced cells *(n = 2)*.

6.3.6 Growth of Cas9+ cells

High expression of the Cas9 enzyme can slow cell growth [216]. To ensure a substantial reduction in growth was not seen in my cells, growth curve experiments were conducted on Cas9+ MDA-MB-231 cells (Figure 90). Although there is a slight trend towards reduced growth in the Cas9+ cells, the slopes are not significantly different (p = 0.4647). This population of cells was therefore approved for use in the CRISPR screen.

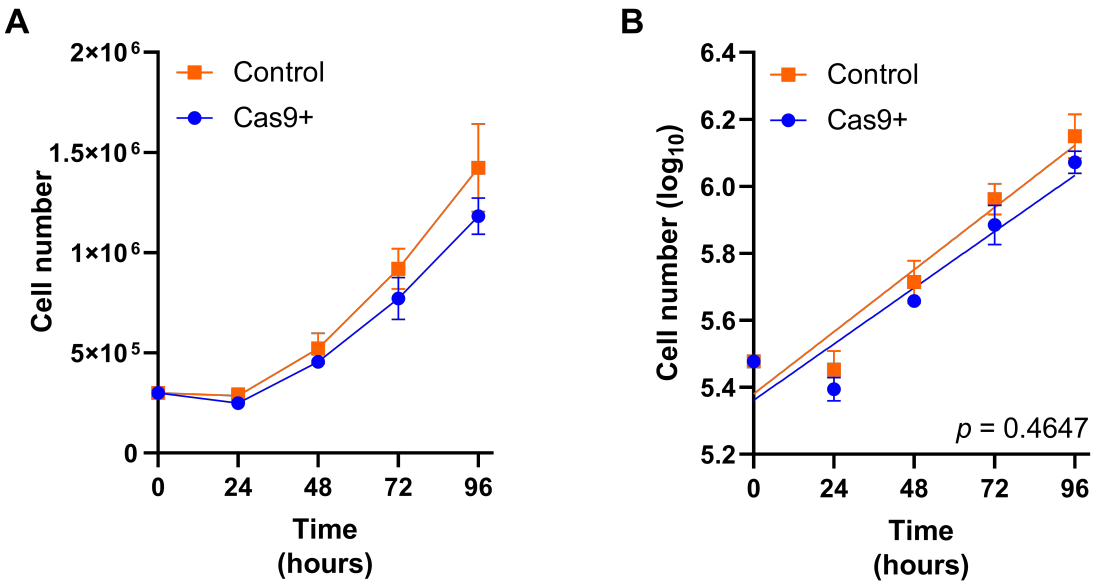

Figure 90: Cas9 growth curve. Growth curves of control MDA-MB-231 cells and MDA-MB-231 cells expressing Cas9. Points show the mean average of *n = 3* and error bars show standard deviation. The slopes generated by linear regression are not significantly different (p = 0.4647).

6.3.7 PCR optimisation

The pLX-miR library uses a nested PCR approach for guide amplification. A nested PCR strategy increases the specificity and sensitivity of the reaction [229]. The first PCR step (PCR1) amplifies the integrated guide region from the genomic DNA. The second round of PCR (PCR2) adds the Illumina adapters to the product of PCR1. Each of these PCR steps required optimisation. Initially, the optimum temperature for primer annealing was determined using a temperature gradient. For PCR1, the predicted melting temperature for the primers was 66℃ (New England BioLabs Tm Calculator, version 1.15.0). PCR reactions were run using annealing temperatures from 60℃ to 76℃ to encompass this value. The sharpest band was produced at 60℃ (Figure 91A).

Similarly, PCR2 reactions were run using a gradient of annealing temperatures from 60℃ to 76.3℃. The melting temperature of these primers was predicted to be very high, so the recommended procedure was a 2-step PCR combining the annealing and extension steps, both at 72℃. Using a temperature of 72℃ for PCR2 produced a band and allowed the 2-step protocol to be used going forward (Figure 91B). The selected temperatures for both PCR reactions allow amplification of the DNA without causing secondary bands to appear. Secondary bands would indicate that the annealing temperature was too low and non-specific primer annealing was occurring.

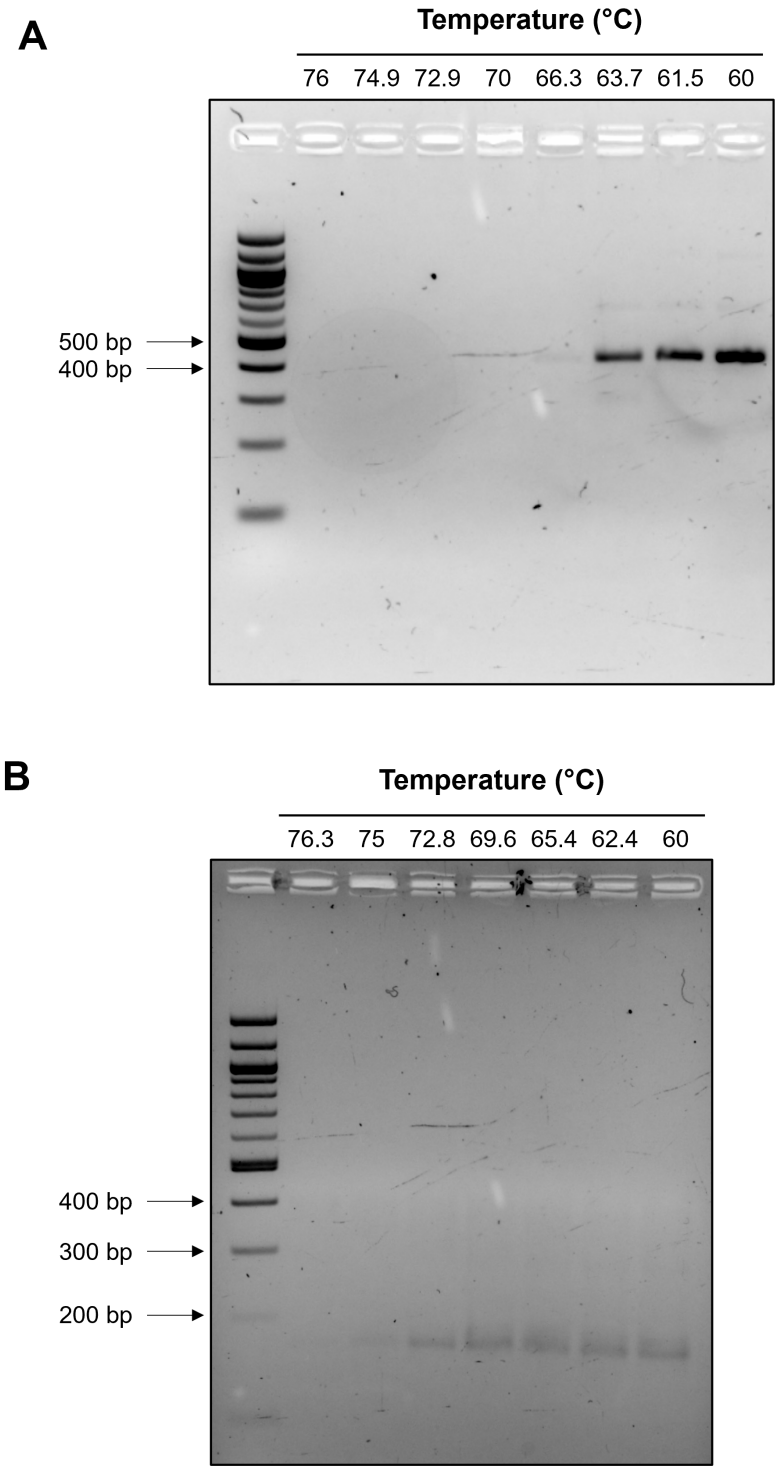

Figure 91: Optimisation of PCR annealing temperature. PCR reactions were conducted using a gradient of annealing temperatures and the resulting PCR products were run on an agarose gel. **(A)** The 391 bp PCR1 product is visible when an annealing temperature of 63.7℃ or lower is used. **(B)** The 175 bp product of PCR2 is lightly visible at all temperatures apart from the highest one tested (76.3℃).

Next, the appropriate amount of DNA to add to the PCR1 reaction was determined. The addition of too much DNA can inhibit the PCR reaction. However, the need to amplify DNA from enough cells to maintain representation means a lot of DNA must be amplified (i.e. the $6.29x10^6$ cells required for representation is equivalent to 41.5 µg of genomic DNA). This means that optimally each reaction should have the most DNA possible without inhibiting the reaction. Figure 92 shows that 1 µg of DNA could be loaded per PCR tube while still producing a strong band, whereas 2 µg of DNA produced a fainter band indicating that the PCR reaction was being inhibited.

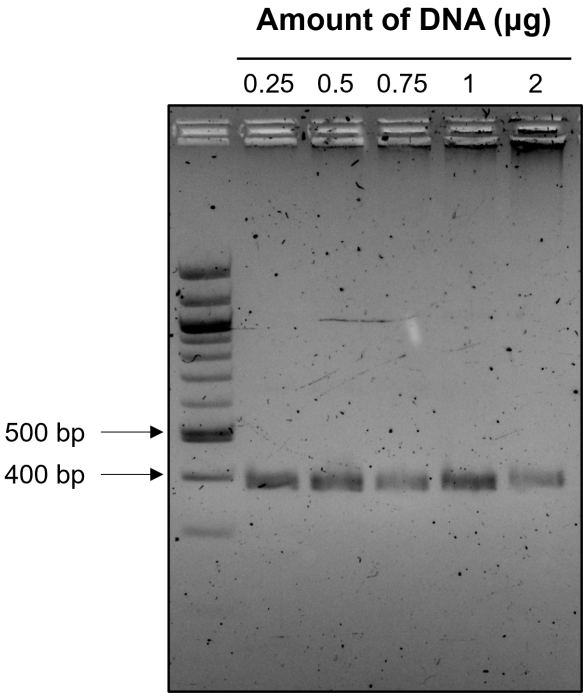

Figure 92: Optimisation of DNA loading in PCR1. PCR1 reactions were conducted using a different amount of genomic DNA as input and the resulting PCR products were run on an agarose gel. The bands are clearly visible using any amount of DNA as input, but the band at 2 µg is less intense than that at 1 µg.

Finally, an appropriate number of PCR cycles for each step was determined. PCR1 failed to produce a visible band when 25 cycles were used. Increasing this to 30 produced a clear band (data not shown). For PCR2, the number of cycles required to produce a visible band was higher than that suggested by Kurata and Lin [97], with even 15 cycles producing only a faint band (Figure 93).

Initially, a 20 cycle PCR was selected as this produced a clear band on the agarose gel. However, the final concentration of the PCR2 product only just met the minimum concentration required for sequencing by the Oxford Genomics Centre, so the number of cycles was increased to 22 to ensure enough product was available. Increasing the amount of PCR1 product added to PCR2 from 1 µL to 2 µL was also tested, but this was found to have little effect (Figure 93). 2 µL was used going forward since this did not affect the reaction but allowed for better maintenance of representation.

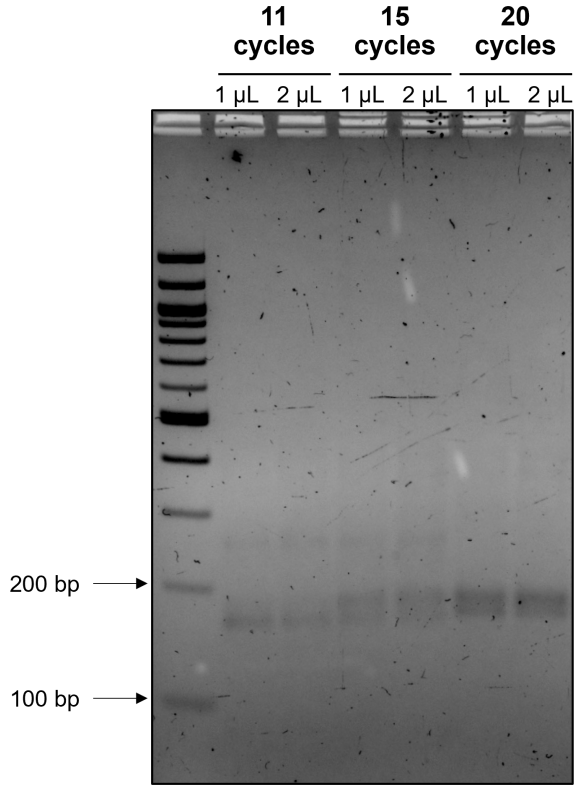

Figure 93: Optimisation of PCR cycles. PCR2 reactions were conducted using different volumes of PCR1 product and a different number of PCR cycles. The products were run on an agarose gel. The volume of PCR1 product has little effect on band intensity. 20 cycles produced the clearest band.

Figure 94 shows the final results of the PCR optimisation. DNA from transduced cells produced clear bands at the correct size for both PCR1 and PCR2 reactions. The negative controls (reactions without DNA and reactions using the DNA of un-transduced MDA-MB-231 cells) did not produce bands after either the PCR1 or the PCR2 reaction. This proved that the bands seen were specific to cells infected with the pLX-miR library, rather than off-target amplification from another part of the genome.

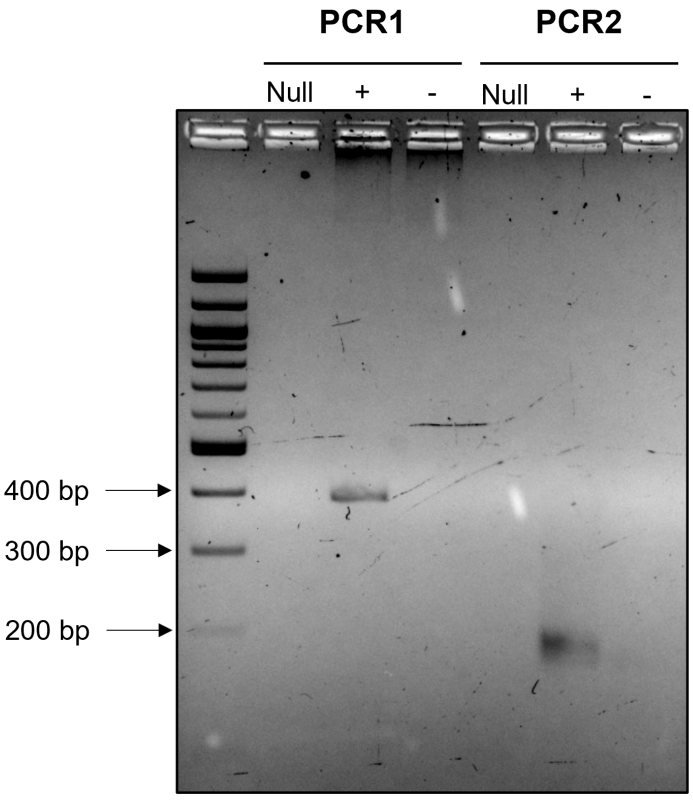

Figure 94: Optimised PCR. PCR reactions were conducted using the optimised conditions and bands were observed at the correct position in the gel for both PCR1 (391 bp product) and PCR2 (175 bp product). To ensure the bands seen were specific to transduced cells (i.e. cells containing the guide sequence), negative controls were used. Lanes labelled “null” did not contain any DNA in the PCR reaction. Genomic DNA from cells transduced with the pLX-miR library was used for the PCR reactions in the lanes labelled “+”. The lanes labelled “-” contained DNA from un-transduced MDA-MB-231 cells.

6.3.8 Quality control

Three replicates of the CRISPR screen were carried out. The doubling time was stable across the three replicates and, as expected, cells cultured under hypoxia had a reduced growth rate (Figure 95). The minimum number of population doublings that occurred was 7.05 (hypoxia; replicate 2) and the maximum was 9.52 (normoxia; replicate 3).

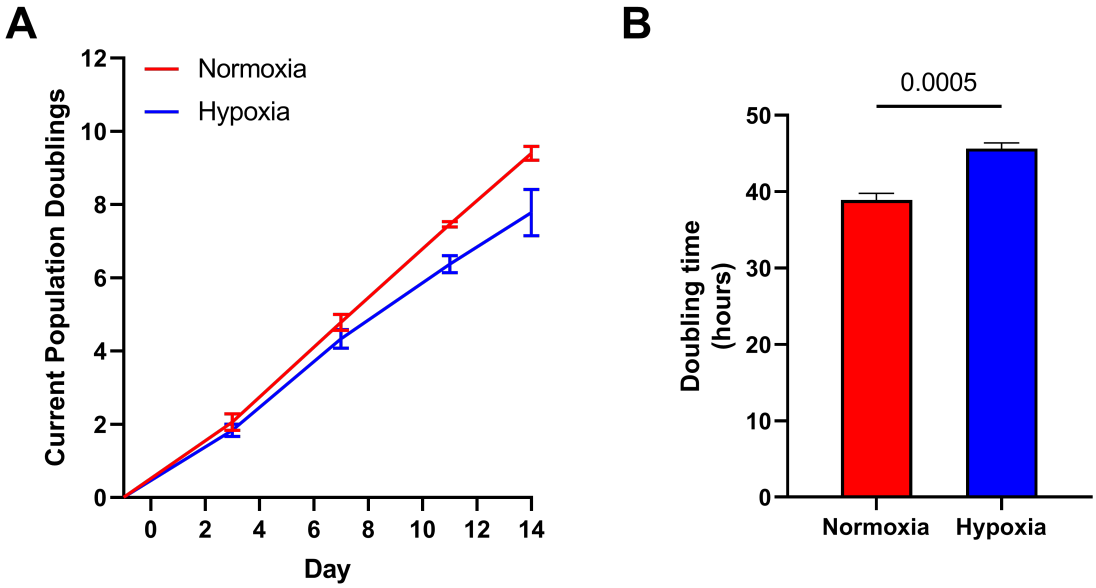

Figure 95: Cell growth during the pLX-miR CRISPR screen. Growth metrics of MDA-MB-231 cells during the CRISPR screen under normoxic and hypoxic conditions *(n = 3)*. Mean and standard deviation are shown. An unpaired t-test showed that growth was significantly reduced under hypoxia ($p = 0.0005$).

The concentration and quality of DNA extracted was consistent between samples (Table S5). Gels produced after the PCR1 and PCR2 reactions showed clear bands at the correct size for all samples and a lack of bands in the negative control lanes (Figure 96). The products of PCR2 were cleaned up and quantified. Again, consistent results were produced across all samples (Table S6). Together, these results indicate that successful DNA extraction and sgRNA amplification has taken place.

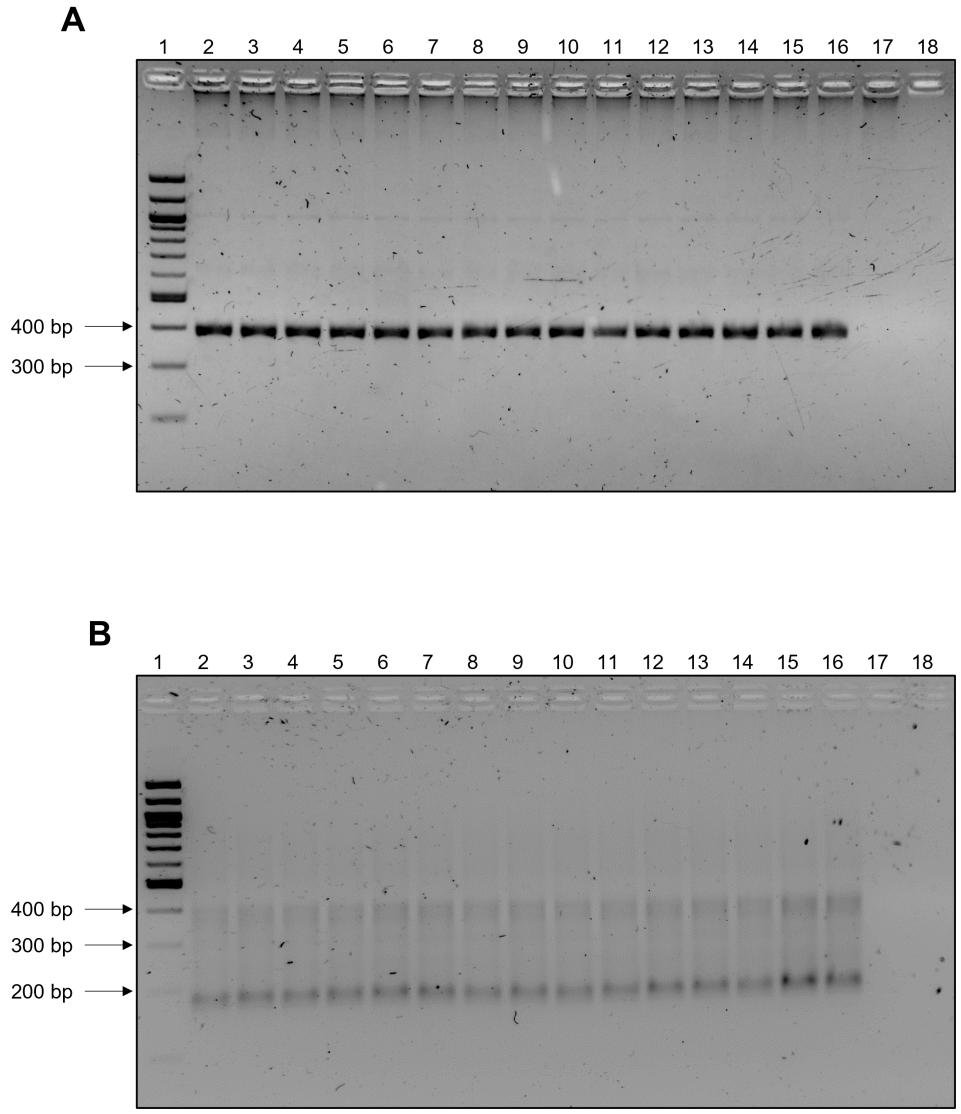

Figure 96: PCR products from the pLX-miR CRISPR screen. PCR products from the CRISPR screen on agarose gel. Contents of the lanes are: 1, ladder; 2, T0 Rep1; 3, T7 Normoxia Rep1; 4, T7 Hypoxia Rep1; 5, T14 Normoxia Rep1; 6, T14 Hypoxia Rep1; 7, T0 Rep2; 8, T7 Normoxia Rep2; 9, T7 Hypoxia Rep2; 10, T14 Normoxia Rep2; 11, T14 Hypoxia Rep2; 12, T0 Rep3; 13, T7 Normoxia Rep3; 14, T7 Hypoxia Rep3; 15, T14 Normoxia Rep3; 16, T14 Hypoxia Rep3; 17, negative control (un-transduced cells); 18, no DNA. **(A)** The products of the PCR1 reaction run for 30 cycles, with an annealing temperature of 60℃, and 1 µg of DNA per reaction tube. A single, clear band is found in all samples demonstrating successful PCR reactions. Both negative control lanes are empty. **(B)** The products of the PCR2 reaction run for 22 cycles, with an annealing temperature of 72℃, and 2 µL of PCR1 product per reaction tube. Clear bands are found across all samples demonstrating successful PCR reactions. A second band can be seen which matches the molecular weight of the PCR1 product. Both negative control lanes are empty.

After sequencing, various measures were assessed for quality control. The number of reads and reads passing filtering was consistent between samples, and all samples had enough mapped reads to give twice the desired coverage (Figure 97). Only 0.1% bases failed the quality check in all samples. The mapping rate was also consistent between samples, averaging 78%.

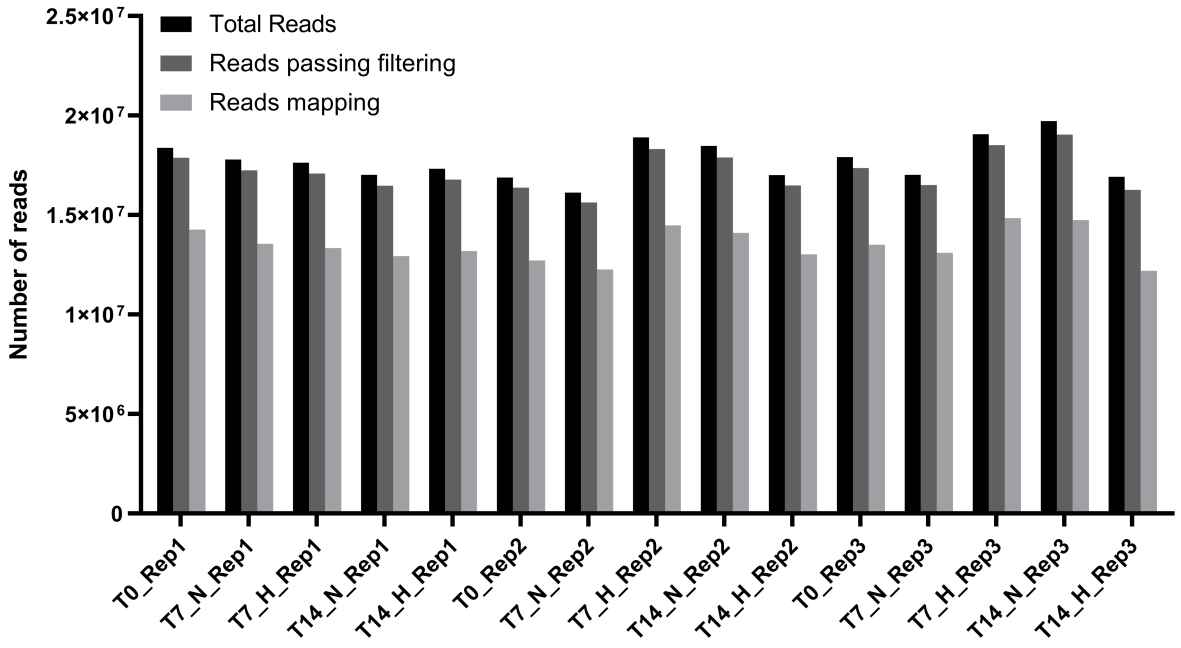

Figure 97: pLX-miR CRISPR guide mapping results. For each sample, the total number of sequenced reads, the number of reads passing filter parameters, and the number of reads which could be mapped to the library annotation file is shown.

The Gini index (reported by MAGeCK) is used to indicate consistency in the read count distribution. A low Gini index (0.2-0.3) is expected in negative selection screens [176]. The samples have Gini indices ranging from 0.05 to 0.07, indicating little to no bias in guide synthesis, viral transduction, or the PCR process [175]. The number of guides with zero counts (i.e. missing guides) is $< 0.25\%$ across all samples, with the maximum number of missing guides in a sample being 18. This indicates very good representation of the library across all samples and replicates. Plots of the guide distribution show consistency between samples and no extreme outliers (Figure 98).

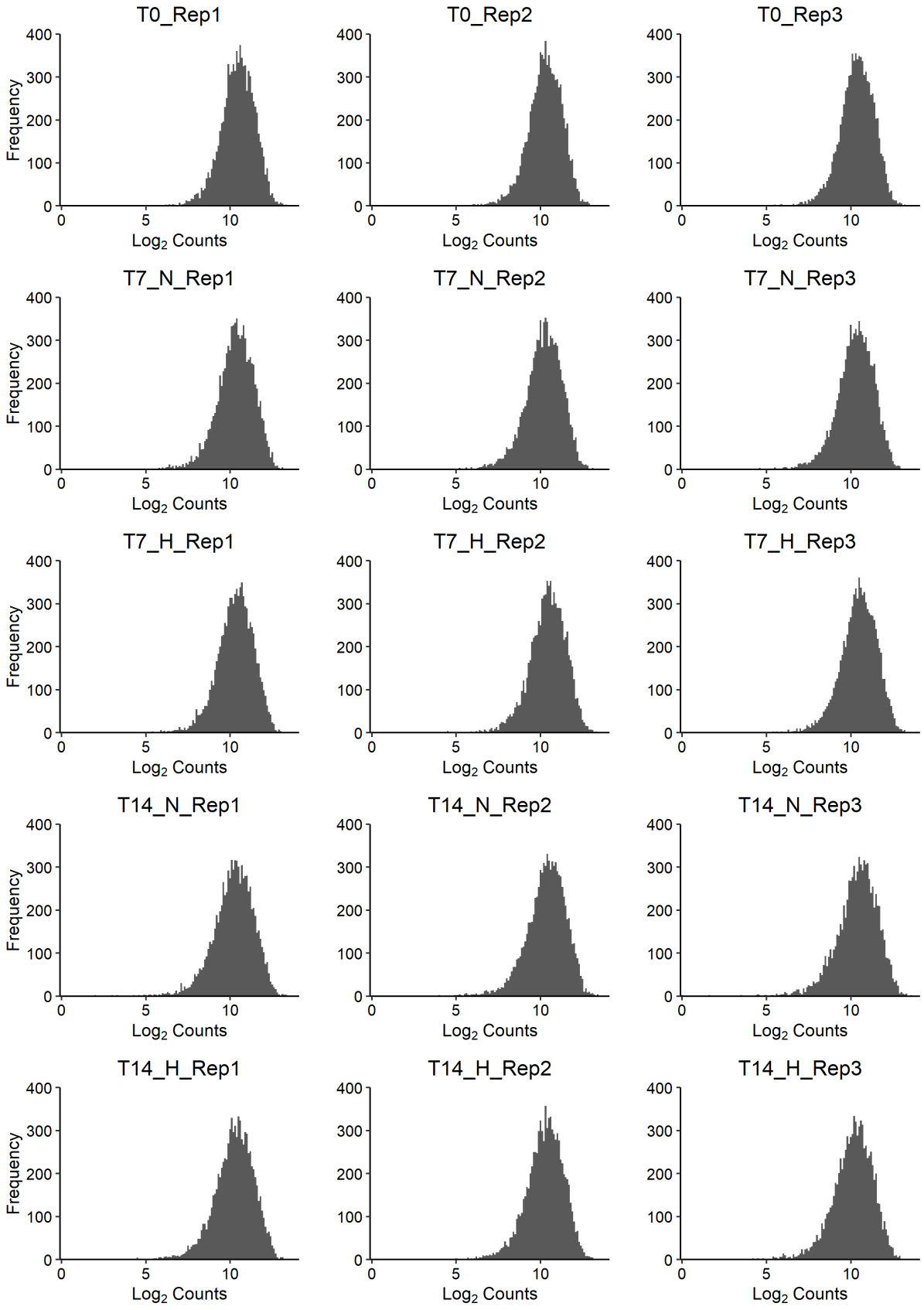

Figure 98: pLX-miR CRISPR guide distribution. The distribution of guides is consistent between samples and no guides are over-represented.

PCA was used to identify sources of variation between the samples (Figure 99). Samples separated along the first principal component over time. This shows that the profiles of the samples are changing during the selection process as expected. All three of the T0 samples cluster closely together, indicating good virus preparation with roughly equal representation of the guides at the start of each biological replicate. Principal component 2 is driven by the biological replicate. In particular, replicate 3 shows more variation compared to replicates 1 and 2, which cluster more closely at the bottom of the plot. This shows there is some batch variation in the data. The normoxia and hypoxia samples cluster very closely together within each replicate and timepoint pairing. This implies that similar selection processes are happening despite the difference in oxygen levels.

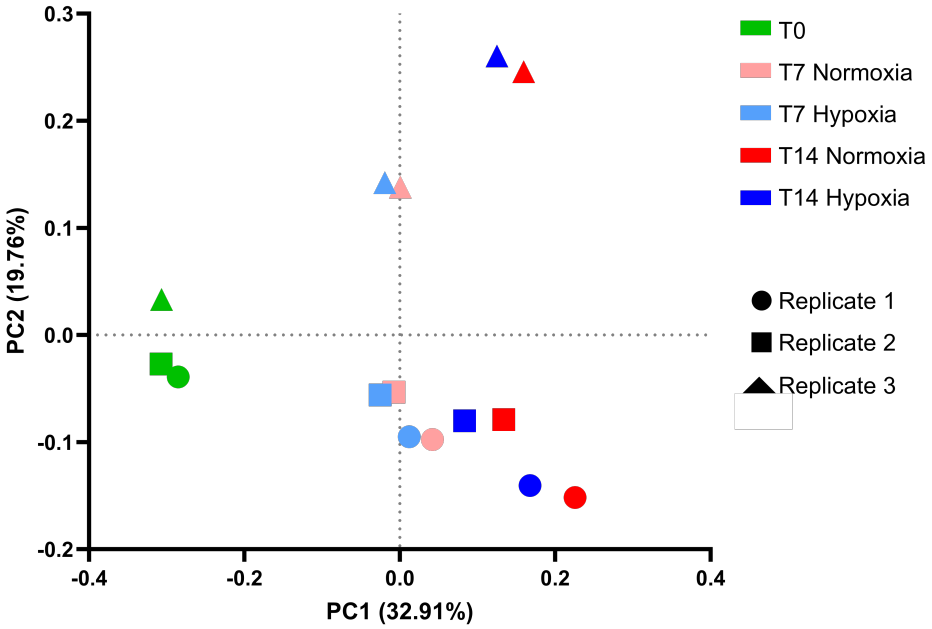

Figure 99: Principal component analysis of samples from the pLX-miR CRISPR screen. Principal component analysis of the fifteen CRISPR samples. Samples are colour coded based on their experimental condition, and the shapes of the symbols represent the replicate number. PC1 = principal component 1. PC2 = principal component 2.

In this experiment, the non-targeting control guides have a $\log_2$ fold change that is positively skewed over time (Figure S14). Less than 15% of non-targeting guides have a $\log_2$ fold change below zero (Table 11), compared to the expected value of 50%. This phenomenon has been described by other researchers [177, 220], and is thought to be caused by the lack

of DNA damage occurring in cells which receive non-targeting guides. The Cas9 enzyme does not initiate double strand breaks in cells expressing non-targeting sgRNAs, which provides a survival advantage in a population where double strand breaks are the norm. The positive skew meant that the non-targeting guides were unsuitable for use in generating a null distribution.

Table 11: Log_2 fold change of non-targeting control guides

Comparison	Non-targeting guides with a log_2 fold change	
	Less than 0	Greater than 0
T0 vs T7 Normoxia	13.1%	86.9%
T0 vs T7 Hypoxia	12.3%	87.7%
T0 vs T14 Normoxia	8.9%	91.1%
T0 vs T14 Hypoxia	11.2%	88.8%

No positive control guides (i.e. guides targeting known essential genes) are included in the pLX-miR library. There are miRNAs which are known to be essential in biological processes (e.g. embryonic development [193]), but there are no confirmed “essential” miRNAs which are relevant for survival or proliferation in a cell line context. In the absence of these preferred positive controls, the screen’s efficacy can be assessed by observing whether selection occurs over time. In general, miRNAs that are called as significant hits after 7 days of selection (T7) remain as significant hits after 14 days (T14), and there are more unique hits at T14 than at T7 (Figure 100). Of the hits that are in common to T7 and T14, 11/12 in normoxia and 15/17 in hypoxia have greater log_2 fold changes at T14, indicating greater selection over time.

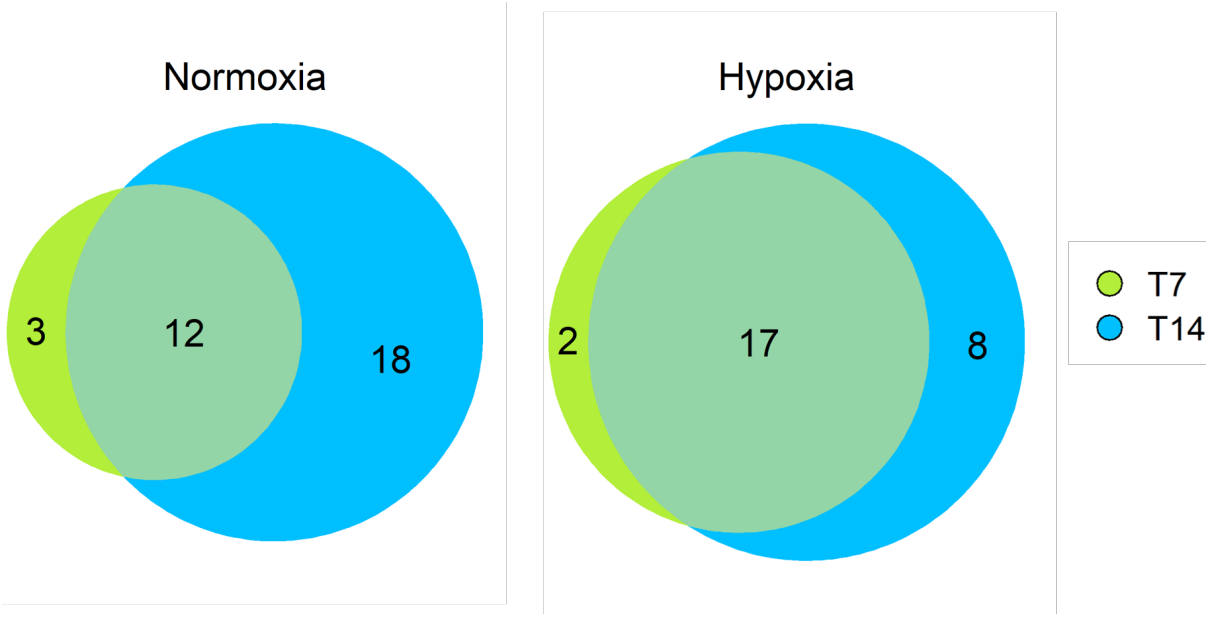

Figure 100: Overlap of miRNA hits found after 7 and 14 days of selection. Euler plots showing the overlap of miRNAs which were significant hits after 7 days (T7) and 14 days (T14) of selection. Under both normoxic and hypoxic conditions, there are more hits identified at T14. The majority of hits which are detected at T7 are also found at T14.

6.3.9 Essential miRNAs in MDA-MB-231 cells

Comparing the abundance of guides between T0 and T7/T14 can be used to identify miRNAs that are essential for growth or survival in this cell line. Figure 101 and Table 12 show miRNAs which are significant hits in each condition. miRNAs which are hits in both the normoxia and hypoxia arms of the experiment are more likely to be true hits and are more likely to be relevant across different environmental conditions. Those which are significant at T7 as well as T14 have greater potency since the knockout population is depleted within a shorter time frame. The miRNAs whose guides are significantly depleted and have a $\log_2$ fold change $\leqslant 1$ in all four conditions are: hsa-mir-6724, hsa-mir-3687, hsa-mir-3648, hsa-mir-663a, and hsa-mir-483.

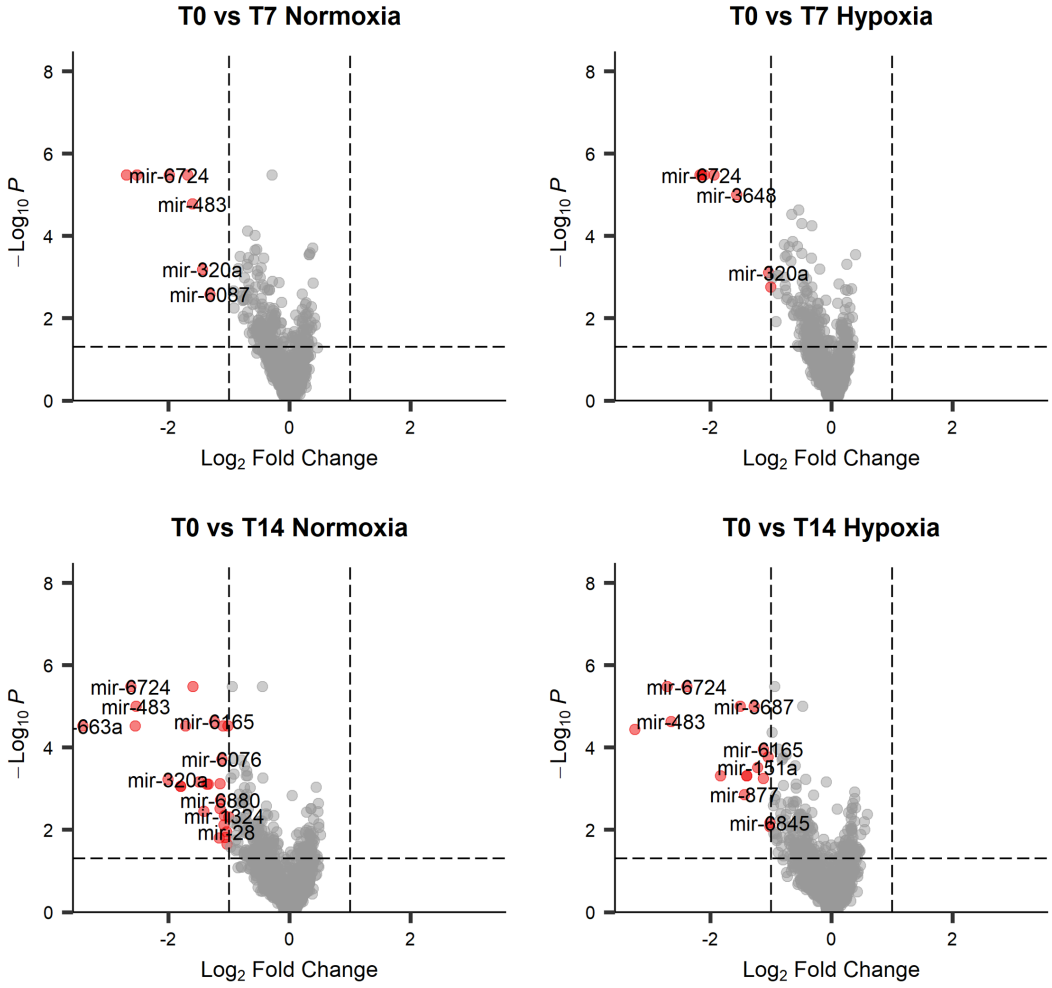

Figure 101: Essential miRNAs. Volcano plots for each time comparison showing miRNAs whose targeting guides were selected against over time.

Table 12: List of essential miRNAs. miRNAs whose knockout causes a significant (adjusted *p* value $\leqslant$ 0.05) growth/survival disadvantage in any comparison with T0 are displayed. ** means this miRNA was significantly depleted and had a $\log_2$ fold change $\leqslant$ -1. * means this miRNA was significantly depleted but had a $\log_2$ fold change between -1 and 0.

miRNA	T0 vs T7 Normoxia	T0 vs T7 Hypoxia	T0 vs T14 Normoxia	T0 vs T14 Hypoxia
hsa-mir-6724	**	**	**	**
hsa-mir-3687	**	**	**	**
hsa-mir-3648	**	**	**	**
hsa-mir-663a	**	**	**	**
hsa-mir-3180	*	*	*	*
hsa-mir-483	**	**	**	**
hsa-mir-1226	*	*	**	**
hsa-mir-6165	*	*	**	**
hsa-mir-1184	*	*	*	*
hsa-mir-658	*	*	-	*
hsa-mir-7111	*	-	-	-
hsa-mir-3652	*	*	*	*
hsa-mir-769	*	*	-	*
hsa-mir-4315	*	*	**	*
hsa-mir-4435	*	*	*	**
hsa-mir-1282	-	*	*	*
hsa-mir-449c	-	*	-	*
hsa-mir-6760	-	*	**	*
hsa-mir-6871	-	*	-	-
hsa-mir-3179	-	*	*	-
hsa-mir-6076	-	-	**	-
hsa-mir-199a	-	-	*	-
hsa-mir-4467	-	-	*	*
hsa-mir-4769	-	-	*	-
hsa-mir-3940	-	-	*	-
hsa-mir-6820	-	-	*	-
hsa-mir-320a	-	-	**	**
hsa-mir-1306	-	-	**	-
hsa-mir-6784	-	-	**	-
hsa-mir-877	-	-	**	-
hsa-mir-4477b	-	-	**	**
hsa-mir-4477a	-	-	**	**
hsa-mir-1299	-	-	**	-
hsa-mir-130b	-	-	*	-
hsa-mir-6798	-	-	*	-
hsa-mir-6777	-	-	-	*
hsa-mir-151a	-	-	-	**
hsa-mir-4783	-	-	-	**
hsa-mir-937	-	-	-	*

An important factor when concluding that a miRNA is essential for growth or survival is establishing that, under normal circumstances, the cell would express that particular miRNA. This is important since miRNA genes are often contained within other genes, so if a miRNA is a hit in a CRISPR screen, but is typically not expressed in the cell, it is more likely that the effect on growth/survival is caused by the guide having an effect on the encompassing gene. The expression level of the miRNAs listed in Table 12 was assessed using the MDA-MB-231 small RNA sequencing data from Section 4, a published dataset of small RNA sequencing also on MDA-MB-231 cells [230], and in breast cancer samples found in TCGA. The top five hits in the screen (hsa-mir-6724, hsa-mir-3687, hsa-mir-3648, hsa-mir-663a, and hsa-mir-483) were either not expressed, or expressed at very low levels in my MDA-MB-231 small RNA-seq dataset (Figure 102). The miRNAs were not detected at all in the publicly available dataset (Figure 103). Of this list, hsa-mir-483 was the only miRNA to be detected in breast cancer samples of the basal subtype from TCGA (Figure 104).

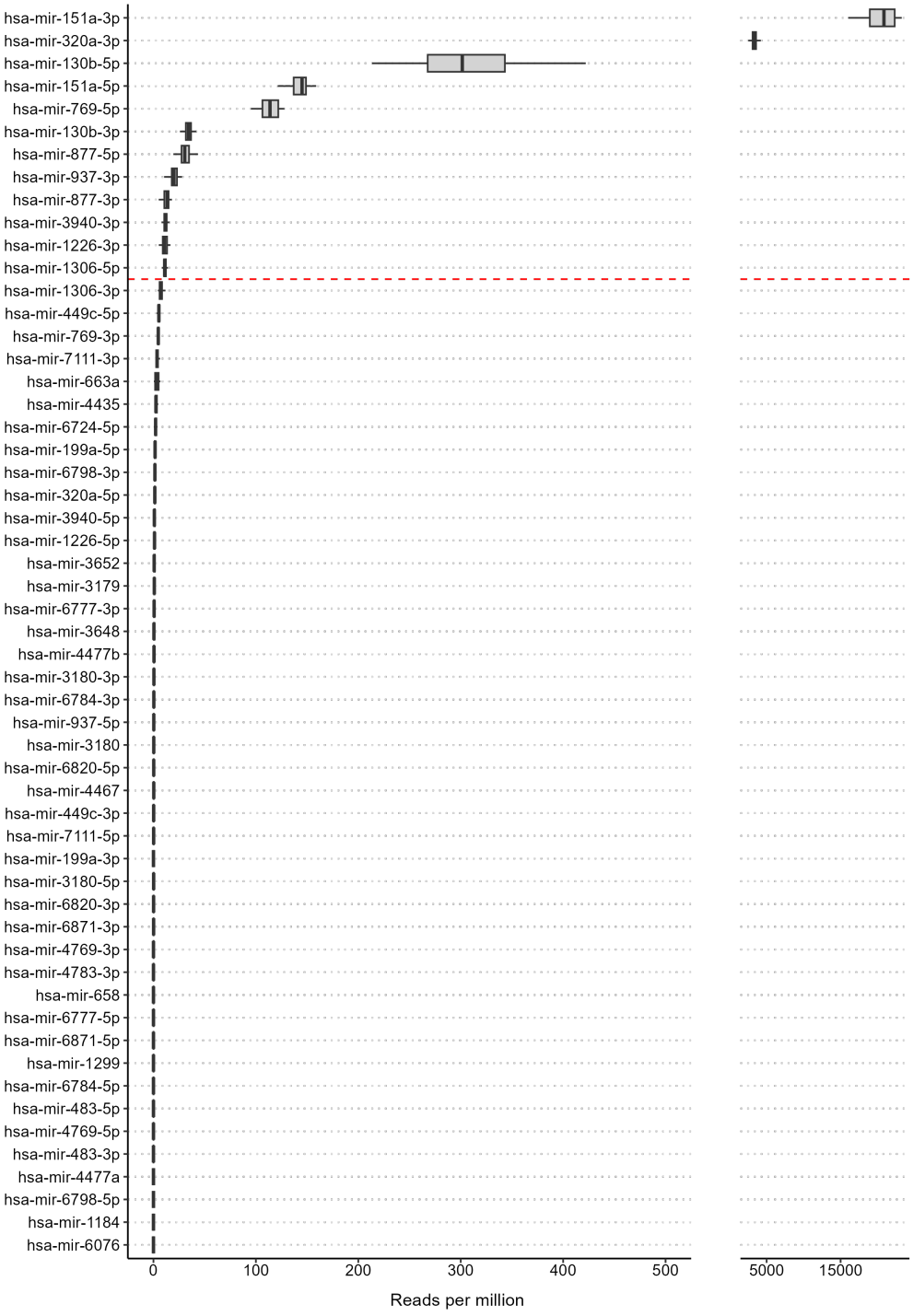

Figure 102: Expression of essential miRNAs in my MDA-MB-231 dataset. Expression of the essential miRNAs in the small RNA-seq dataset from Section 4. miRNAs above the red dashed line have a median expression level of at least 10 reads per million. Each boxplot displays the median and the first and third quartiles. The whiskers extend to the highest/lowest point that lies no more than 1.5 times the interquartile range from the first/third quartiles. hsa-mir-6724, hsa-mir-3687, hsa-mir-3648, hsa-mir-663a, and hsa-mir-483 are expressed at very low levels in this dataset. A split x axis has been used.

Figure 103: Expression of essential miRNAs in a published MDA-MB-231 dataset. Expression of the essential miRNAs in a published dataset of miRNA sequencing on MDA-MB-231 cells. miRNAs above the red dashed line have at least 10 reads per million. hsa-mir-6724, hsa-mir-3687, hsa-mir-3648, hsa-mir-663a, and hsa-mir-483 are not expressed in this dataset.

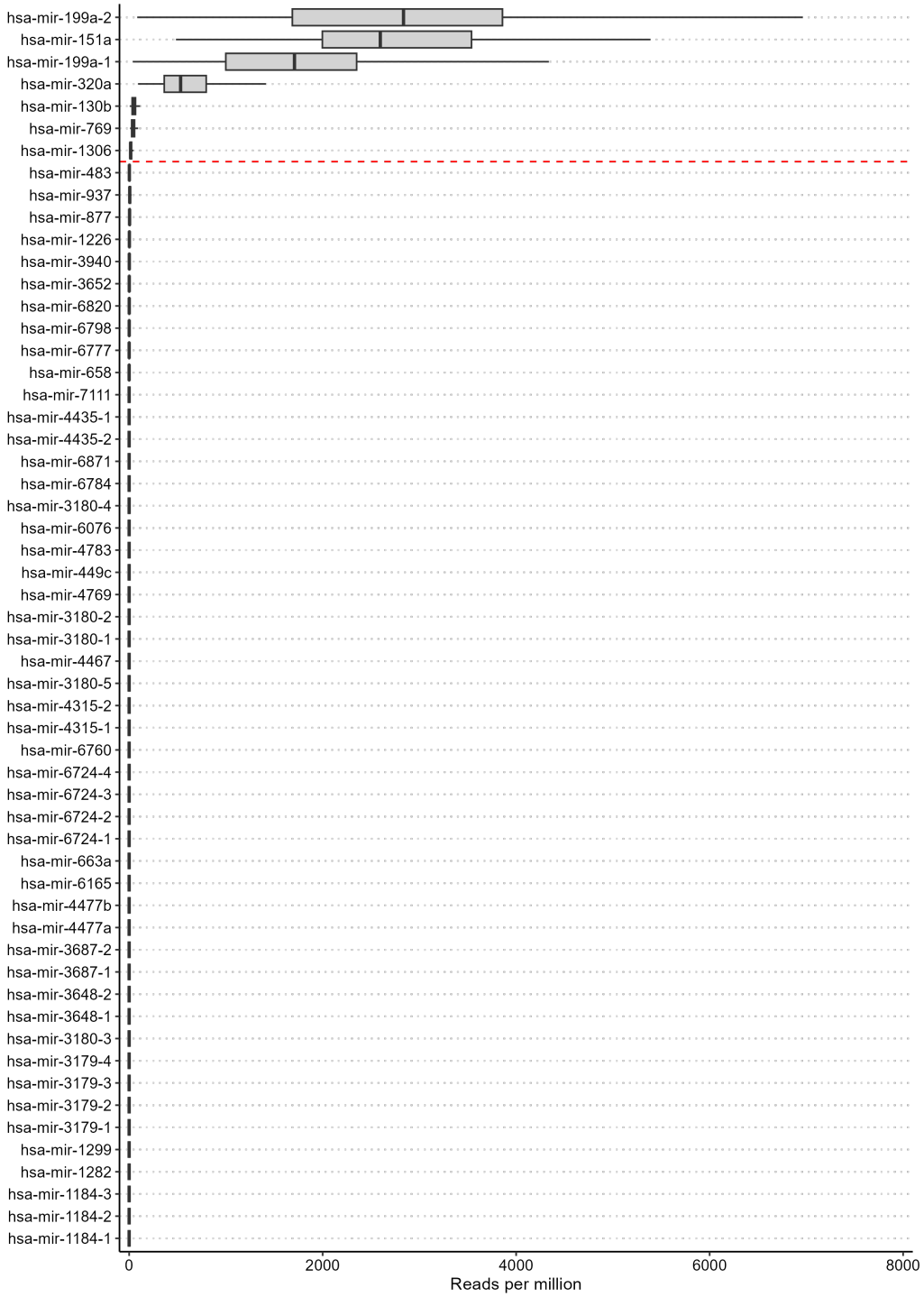

Figure 104: Expression of essential miRNAs in TCGA. Expression of the essential miRNAs in breast cancer samples of the basal subtype from TCGA. miRNAs above the red dashed line have at least 10 reads per million. Each boxplot displays the median and the first and third quartiles. The whiskers extend to the highest/lowest point that lies no more than 1.5 times the inter-quartile range from the first/third quartiles. Outlier points are not displayed. hsa-mir-6724, hsa-mir-3687, hsa-mir-3648, hsa-mir-663a, and hsa-mir-483 are not detected at significant levels in these samples.

Given that miRNA molecules work in a 1:1 ratio by binding to complementary mRNA sequences, it is unlikely that these lowly expressed miRNAs are having a significant functional effect, despite the guides targeting these regions causing a significant growth/survival disadvantage in the CRISPR screen. Further investigation into the genomic loci of these miRNAs offers some explanation as to how these guides have such a drastic effect on the cell. One of the miRNAs, hsa-mir-6724, is found at four different genomic loci, with three of these in close proximity to ribosomal RNA (rRNA) genes. Since the expression of miR-6724 is low in MDA-MB-231 cells, it is likely that the negative effects conferred by guides targeting this miRNA are caused by either the large number of double strand breaks which would occur given that the guides target all four genomic copies of miR-6724, or by off target effects relating to the proximal 5.8S rRNA gene. hsa-mir-3687 has been withdrawn from miRbase since the creation of the pLX-miR library, but its location was annotated to the same genomic region as three of the four hsa-mir-6724 genomic copies, again placing it in proximity to a 5.8S rRNA gene. hsa-mir-3648 has two genomic copies, and hsa-mir-3648-1 is contained within this same 5.8S rRNA gene. The close genomic proximity of these miRNAs indicates that their significance in this study is caused by off target effects rather than knock out of the miRNAs themselves. The two remaining miRNAs are both located within other genes which could explain their significance in the screen. hsa-mir-663a is located within a lncRNA (*MIR663AHG*) which does not have a known function. hsa-mir-483 is located within *IGF2*, meaning this protein could be affected by the sgRNAs. However, knockout of insulin-like growth factor 2 (IGF2) did not affect MDA-MB-231 cell growth in the genome-wide CRISPR screen (T0 vs T14 normoxia: $\log_2$ fold change 0.05, adjusted p value 1. T0 vs T14 hypoxia: $\log_2$ fold change = 0.07, adjusted p value = 1). Given that these miRNAs are not expressed in the relevant samples, and most have plausible alternative explanations for why the sgRNAs may have had an effect, it is very likely that these hits are false positives.

Most of the miRNAs in Table 12 are expressed at low levels. However, a few of the miRNAs are highly expressed in both MDA-MB-231 cells and breast tumour samples. These miRNAs are hsa-mir-151a, hsa-mir-320a, hsa-mir-130b, hsa-mir-769, and hsa-mir-1306. As these miRNAs are well expressed in the relevant samples, these are more likely to be true positive hits. However, upon investigation, many of these miRNAs again lie in close proximity to known essential genes. hsa-mir-151a is located in an intron of *PTK2*. This gene is a hit in the genome-

wide CRISPR screen from Section 4 (T0 vs T14 normoxia: $\log_2$ fold change = -1.54, adjusted *p* value = 0.039. T0 vs T14 hypoxia: $\log_2$ fold change = -2.21, adjusted *p* value = 0.056) and is reported as "strongly selective" in DepMap [231]. hsa-mir-320a lies slightly upstream of the *POLR3D* gene, which is reported as "common essential" by DepMap and trends towards significance in my genome-wide CRISPR screen (T0 vs T14 normoxia: $\log_2$ fold change = -1.55, adjusted *p* value = 0.113. T0 vs T14 hypoxia: $\log_2$ fold change = -1.31, adjusted *p* value = 0.155). hsa-mir-130b does not lie in close proximity to any protein-coding genes so is less likely to have secondary effects via sgRNAs affecting protein expression. hsa-mir-769 is located downstream of *CCDC61*. This gene is not reported as essential in DepMap, nor was it a hit in my genome-wide CRISPR screen (T0 vs T14 normoxia: $\log_2$ fold change = -0.27, adjusted *p* value = 1. T0 vs T14 hypoxia: $\log_2$ fold change = 0.01, adjusted *p* value = 1). Finally, hsa-mir-1306 lies within an exon of *DGCR8*, a component of the microprocessor complex. DepMap identifies this gene as "common essential", and this is supported by my genome-wide CRISPR screen (T0 vs T14 normoxia: $\log_2$ fold change = -1.35, adjusted *p* value = 0.021. T0 vs T14 hypoxia: $\log_2$ fold change = -2.05, adjusted *p* value = 0.013). In the absence of laboratory validation, the most likely true positive hits from this screen are hsa-mir-130b and hsa-mir-769. They are both expressed in relevant samples both *in vitro* and *in vivo*, and there are no obvious secondary effects which could explain the reduction in growth/survival caused by the sgRNAs. Their guide profiles over time are shown in Figure 105.

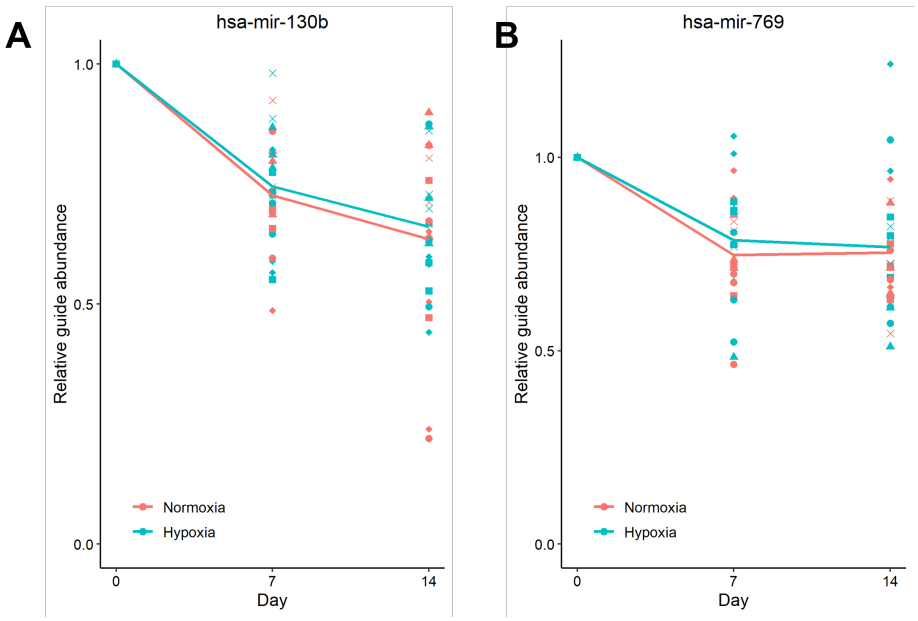

Figure 105: miR-130b and miR-769. The abundance of each miRNA targeting guide relative to the T0 condition. The shapes of the symbols represent each unique guide. This allows guides which have little effect, and therefore may have low guide efficiency, to be identified. Points are connected from the mean at each timepoint. **(A)** Guides targeting hsa-mir-130b show a gradual depletion over time. **(B)** Guides targeting hsa-mir-769 are reduced by day 7 but do not appear to show selection following this.

6.3.10 Potential downstream effects of miR-130b knockout

As the most likely true positive hits, the predicted targets of hsa-mir-130b and hsa-mir-769 were identified in an attempt to explain why these miRNAs could be essential to MDA-MB-231 cells. Their predicted targets were identified using miRNAtap which combines results of five different target prediction tools. This is important because miRNAs do not perfectly base-pair with their mRNA target, meaning it is difficult to accurately predict targets and different algorithms can lead to very different predictions [232]. The targets were filtered to retain only targets which are expressed in MDA-MB-231 cells using the RNA-seq results from Section 4. This is because miRNAs function by binding to mRNA molecules, so they are unable to have a biological effect on genes which have no mRNA molecules present in the cell of interest. The top 30 predicted targets of miR-130b after filtering are shown in Table 13.

There are some interesting targets of miR-130b, some of which have converging functions. For example, two of the top 30 targets are inhibitors of p53 activity. *FBXO11* encodes F-box only protein 11, which neddylates p53, suppressing its activity [233, 234], while *UBE2D2* encodes

Table 13: Predicted targets of miR-130b. The top 30 predicted targets of miR-130b according to miRNAtap. Genes were predicted as targets by a minimum of three independent programs. Each program assigns a score to each target and orders these to assign a rank (displayed in the relevant column). For each target, the geometric mean of its ranking in each program is calculated. The list of genes is then sorted based on this geometric mean and the "Final Rank" is assigned. Genes which did not have an average of $\geqslant 1$ count per sample in the MDA-MB-231 RNA-Seq data were removed, but the "Final Rank" column was not updated to reflect this, leaving some missing values. For example, the predicted target with the second highest ranking, *UNC5D0*, is removed from this table as it is not expressed in MDA-MB-231 cells.

Gene	Ensembl ID	Final Rank	DIANA	miRanda	miRDB	PicTar	TargetScan
PIK3CB	ENSG00000051382	1	10	5	9	-	-
SAMD4A	ENSG00000020577	3	4.5	-	27	7	-
FBXO11	ENSG00000138081	5	28	-	2	34	22
PLXNA4	ENSG00000221866	7	53	-	17	2	-
DIO2	ENSG00000211448	8	13	-	62	3	-
MAP3K13	ENSG00000073803	9.5	103.5	2	-	12	-
SPTY2D1	ENSG00000179119	9.5	3	-	36	23	-
UBE2D2	ENSG00000131508	11	15	-	20	-	9
SMARCA1	ENSG00000102038	13	42	-	24	33	5
POU3F2	ENSG00000184486	14	28	-	16	6	62
TAF4B	ENSG00000141384	15	6	8	-	-	76
DDX6	ENSG00000110367	17	14	-	73	4	-
HAND1	ENSG00000113196	19	21	-	15	53	15
PHACTR2	ENSG00000112419	21	22	-	-	9	27
STIMATE	ENSG00000213533	22	62	3	-	29	60
UQCRB	ENSG00000156467	23	82	-	70	-	1
TCAF1	ENSG00000198420	25	2	-	43	-	69
IKZF3	ENSG00000161405	26	25	-	37	5	73.5
CHRDL1	ENSG00000101938	27	23	-	13	21	55
LENG8	ENSG00000167615	28	99	6	-	32	19.5
GATA6	ENSG00000141448	31	26	-	5	62	50.5
CNKSR2	ENSG00000149970	33	16	-	31	27	38.5
SH3BGRL2	ENSG00000198478	34	50	-	18	17	44
CPEB4	ENSG00000113742	35	4.5	-	58	39	-
MEF2C	ENSG00000081189	36	46.5	-	10	24	-
PDCD6IP	ENSG00000170248	37	57.5	-	-	16	13
TSPAN2	ENSG00000134198	40	39	-	6	54	-
DIABLO	ENSG00000184047	41	-	14	66	-	14
SLC4A11	ENSG00000088836	42	8	-	49	-	33.5
IGFBP2	ENSG00000115457	45	30	-	23	59	

an E2 ligase known to ubiquitinate p53 thus marking it for degradation [235]. If these are true targets, knockout of miR-130b would increase the levels of these proteins, which would in turn reduce the amount of p53. Though this seems contradictory to reduced survival, MDA-MB-231 cells have been shown to be dependent on their mutant p53 for survival [236, 237], so this could explain the decrease in cell growth/survival caused by miR-130b knockout. *SH3BGRL2* encodes a known tumour suppressor gene [238, 239], so knockout of miR-130b would increase its abundance within the cell. *DIABLO* encodes a pro-apoptotic mitochondrial protein. It binds to inhibitor of apoptosis proteins (IAPs) which are thought to sequester caspases and inhibit their activity. When DIABLO binds to the IAPs, caspases are freed and apoptosis can occur [240]. The knockout of miR-130b could increase the amount of DIABLO present in the cell and thus trigger apoptosis.

6.3.11 Potential downstream effects of miR-769 knockout

The top 30 predicted targets of miR-769 are shown in Table 14. There are fewer targets of miR-769 that have converging functions than those of miR-130b. Despite this, there are some predicted targets which have plausible mechanisms for their increased expression causing a growth defect in the CRISPR screen. For example, reduced expression of S100 calcium-binding protein A2 (S100A2) is implicated in many cancer types [241–243], and exogenous expression of S100A2 caused growth repression in breast cancer cell lines [244]. Knockout of miR-769 would have the same effect by increasing the amount of S100A2 protein present. SMAD2 is a mediator of TGF-β signalling and is implicated in cancer. It has been reported to have tumour suppressive functions [245], though a study in MDA-MB-231 cells found that although SMAD2 knockdown did create a more aggressive phenotype with increased metastases, it did not increase cell growth [246]. This type of change would not be detected in the CRISPR screen. *SSNA1* is an interesting predicted target as it encodes a microtubule stabilising protein and has been compared to taxanes in its activity. Knockout of miR-769 would cause an increase in SSNA1, stabilising microtubules and preventing mitosis [247].

Table 14: Predicted targets of miR-769. The top 30 predicted targets of miR-769 according to miRNAtap. Genes were predicted as targets by a minimum of three independent programs. Each program assigns a score to each target and orders these to assign a rank (displayed in the relevant column). For each target, the geometric mean of its ranking in each program is calculated. The list of genes is then sorted based on this geometric mean and the "Final Rank" is assigned. Genes which did not have an average of $\geqslant 1$ count per sample in the MDA-MB-231 RNA-Seq data were removed, but the "Final Rank" column was not updated to reflect this, leaving some missing values. For example, the predicted target with the third highest ranking, *WFDC11*, is removed from this table as it is not expressed in MDA-MB-231 cells.

Gene	Ensembl ID	Final Rank	DIANA	miRanda	miRDB	PicTar	TargetScan
TFAM	ENSG00000108064	1	1.5	3	16	-	8.5
TRIM66	ENSG00000166436	2	3	38	-	1	-
TSEN34	ENSG00000170892	4	30	1	-	-	5
S100A2	ENSG00000196754	5	4	-	30	-	2
IMPDH1	ENSG00000106348	7	34	31	3	7	14
PRKAA1	ENSG00000132356	9	24.5	-	1	-	38
AKT3	ENSG00000117020	12	10	-	19	5	46
SMAD2	ENSG00000175387	13	49	-	13	3	25
SET	ENSG00000119335	14	5	50	4	-	82
STK35	ENSG00000125834	15	42	-	20	2	61.5
ZBTB44	ENSG00000196323	16	50.5	-	5	11	-
CCDC24	ENSG00000159214	17	70	2	-	-	20.5
CEP57	ENSG00000166037	18	37	-	10	19	19
ZNF415	ENSG00000170954	21	41	26	17	-	10
FAM117B	ENSG00000138439	22	14	-	34	13	35
TRAPPC3	ENSG00000054116	24	59	22	26	21	22
QRSL1	ENSG00000130348	25	9	42	24	-	31.5
FAXDC2	ENSG00000170271	26	11	-	21	-	26
CERCAM	ENSG00000167123	29	35	11	-	-	18
SSNA1	ENSG00000176101	30	-	16	59	-	8.5
SPAST	ENSG00000021574	31	19	-	11	-	41
AGPAT1	ENSG00000204310	32	-	9	18	-	54
WDR37	ENSG00000047056	33	76	34	37	8	67
UPF3B	ENSG00000125351	35	7	-	44	-	34
LRP12	ENSG00000147650	36	29	-	7	-	52
SYTL2	ENSG00000137501	38	24.5	-	12	-	40
C1QTNF1	ENSG00000173918	39	20	44	45	-	23.5
TDRD6	ENSG00000180113	40	21	37	40	-	30
GGCT	ENSG00000006625	41	69	47	-	-	4
MRPS16	ENSG00000182180	42	33	-	27	-	15

6.3.12 miRNAs with a hypoxia specific effect on growth/survival

Despite there being a known miRNA transcriptional response to hypoxia, no miRNAs were significant hits when comparing the T14 normoxia and T14 hypoxia conditions. The $\log_2$ fold changes of guides in these comparisons were relatively small, indicating samples in normoxia and hypoxia are very similar (Figure 106).

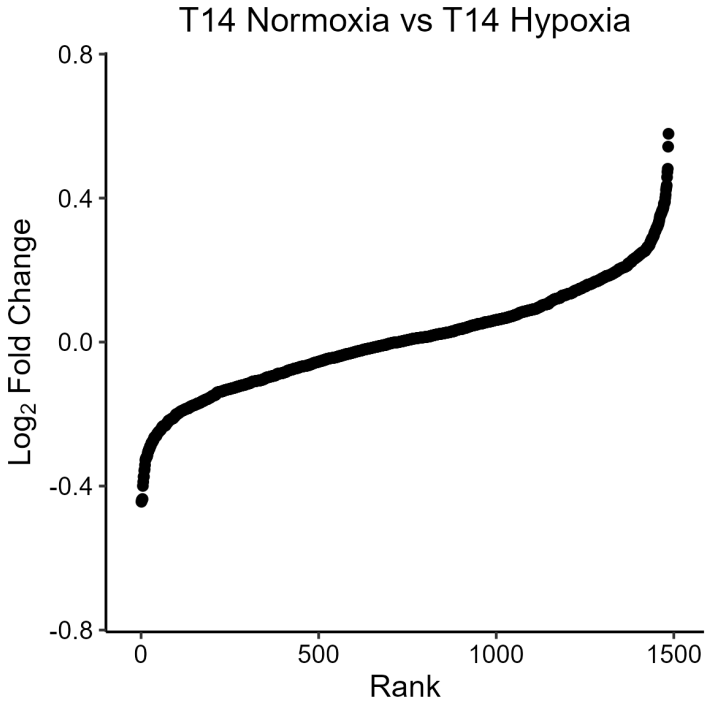

Figure 106: Normoxia vs hypoxia samples after 14 days of selection. The $\log_2$ fold change of the guides when comparing T14 normoxia and T14 hypoxia directly. Most $\log_2$ fold changes lie between -0.4 and 0.4, indicating very few differences in selection of guides between the two samples.

This can also be observed in the PCA whereby the normoxia samples are in close proximity to their hypoxic counterparts (Figure 99). The small changes in guide abundance result in no significant results after correcting for multiple testing. Relaxing the filtering parameters (i.e. using the *p* value instead of the adjusted *p* value and requiring only a +/-0.3 $\log_2$ fold change) shows the miRNAs which are most likely to have an effect on survival under hypoxia (Figure 107).

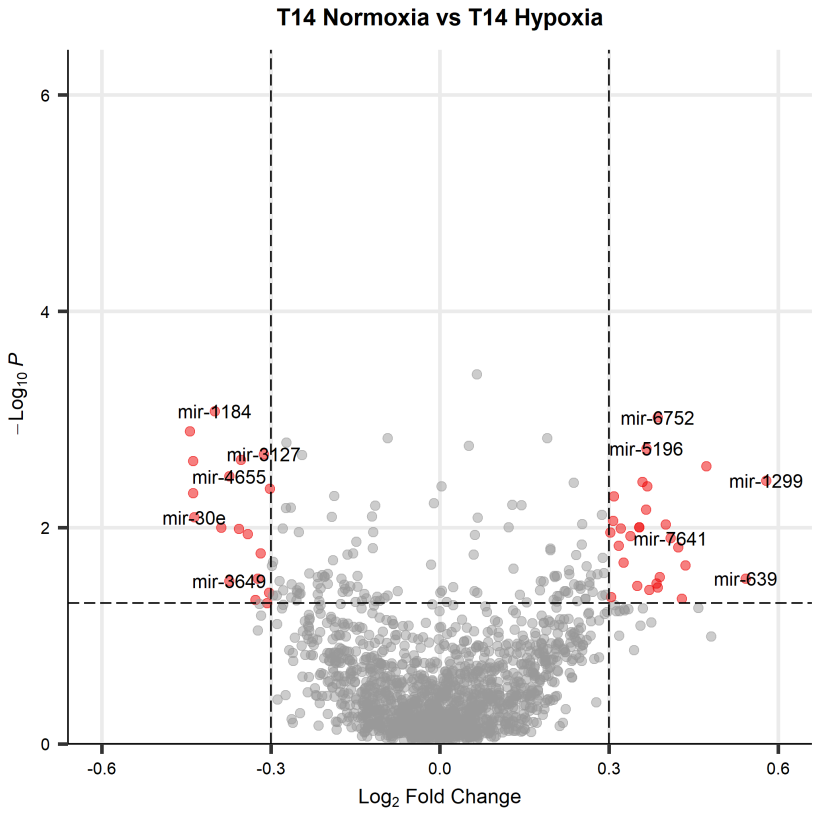

Figure 107: Hypoxia specific miRNA hits. miRNAs which are identified as hits in the screen when using a p value $\leqslant 0.05$ cutoff rather than adjusted p value $\leqslant 0.05$.

Although the direct comparison between normoxia and hypoxia produced no significant results, some insight can nonetheless be gained from trends in the results. One example is the opposing effect of hsa-mir-1299 and hsa-mir-3127 knockout on normoxic and hypoxic cells (Figure 107). These miRNAs are inhibitors and activators of signal transducer and activator of transcription 3 (STAT3) signalling respectively. miR-3127 has been shown to indirectly increase STAT3 phosphorylation and therefore its activity [248]. miR-1299 reduces protein expression of STAT3 through the canonical miRNA pathway [249]. Given that STAT3 is known to act alongside HIF1 and can induce Twist and Snail expression, there are possible modes of action for these miRNAs [250, 251]. Additionally, these two miRNAs have been shown to have opposing effects on autophagy, with miR-3127 being an inhibitor and miR-1299 an activator [248, 252]. Autophagy is typically increased under hypoxia, so mediation of this process could also explain the differing selection effects of miR-1299 and miR-3127 knockout.

miR-210 is known to be strongly upregulated under hypoxia and is prognostic in many studies [44, 253–255]. However, there was no evidence that cells in either normoxia or hypoxia relied upon the expression of miR-210 for their survival (Figure 108).

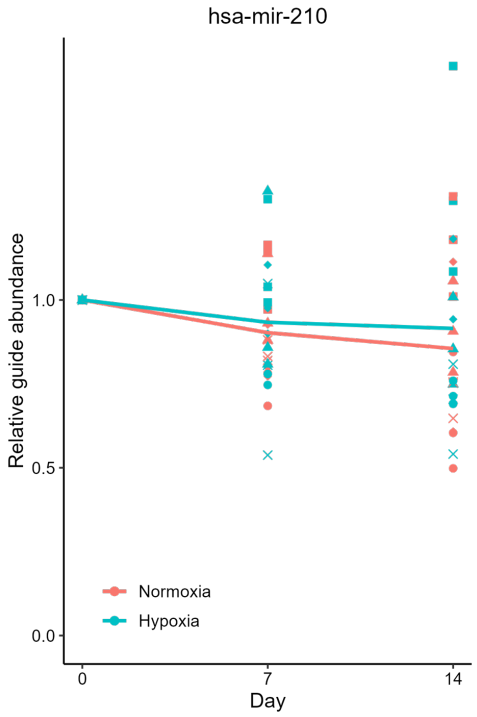

Figure 108: miR-210. The abundance of each guide targeting hsa-mir-210 relative to the T0 condition. The shapes of the symbols represent each unique guide. Lines are connected at the mean. The guides targeting hsa-mir-210 do not decrease in abundance over time.

6.4 Discussion

The miRNA CRISPR screen identified putative essential miRNAs in MDA-MB-231 cells. These are miR-130b and miR-769. In itself, this is an important finding. Given that the hypoxic cells have gone through fewer cell doublings (an average of 7.8 vs 9.4), an equal $\log_2$ fold change in normoxia and hypoxia samples might actually indicate increased sensitivity in hypoxic cells. Since hypoxia often increases treatment resistance [15, 28], finding drug targets that reduce cancer growth equally in normoxia and hypoxia is key. Additionally, there were trends towards hypoxia specific results. This may indicate that not enough time passed during the screen to pick out subtle changes. To identify miRNAs which may affect survival under hypoxia, longer selection times should be used.

miRNAs which were expected to be identified as hits in this screen based on previous experiments in breast cancer did not present. For example, miR-23b and miR-27b, which have been shown to be oncogenic, did not alter cell growth in this screen [200]. miR-21, which has been shown to be essential in MDA-MB-231 cells, was not a significant hit [201]. However, it still ranked within the top 100 hits in both the normoxia and hypoxia arms of the experiment. Kurata and Lin [97] also expected miR-21 to be essential in their screen of HeLa cells using this library, but it was not identified. This could indicate low guide efficiency of the sgRNAs targeting miR-21.

Generally, there is a reasonable level of overlap between Kurata and Lin's [97] results in HeLa cells and my MDA-MB-231 screen, but none with the NCI-N87 cells they used in their second screen. Using an adjusted *p* value cut off of 0.1, the hits from my screen which also appear in their HeLa cell screen are: hsa-mir-151a (hypoxia only), hsa-mir-320a, hsa-mir-663a, hsa-mir-1226, hsa-mir-3648 and hsa-mir-3687. Out of these, hsa-mir-3648 and hsa-mir-3687 could have had the same off target effects in their screen caused by impairment of the 5.8S rRNA. hsa-mir-663a was dismissed in my screen due to a lack of expression in MDA-MB-231 cells, but this miRNA was expressed in their HeLa cells and does not overlap a known essential gene so could be a true hit. Despite being expressed in my MDA-MB-231 dataset, hsa-mir-1226 did not have significant expression in samples from TCGA so was not investigated further. However, this miRNA is an intron of *DHX30*, a gene which is not marked as essential DepMap [231], so again could be a true hit. hsa-mir-151a and hsa-mir-320a lie

in close proximity to known essential genes, which would cause the same off-target results in both of our screens. However, Kurata and Lin have since validated miR-151a in their cell line (bioRxiv preprint [256]). This means that miR-151a could also be relevant in my system, especially given its high expression levels in both MDA-MB-231 cells and samples from TCGA (Figures 102 to 104). Kurata and Lin (bioRxiv preprint [256]) found that p53 is a direct target of miR-151a-3p, while miR-151a-5p was able to reduce p53 levels via an unknown indirect mechanism. Knockout of hsa-mir-151a increased p53 levels and triggered G1 arrest in HeLa cells.

miR-130b is known to be upregulated in malignant tissues [257–259]. A validated target of miR-130b-5p is cyclin-G2, a cell cycle regulator. Low levels of cyclin-G2 have been associated with poor prognosis in TNBC [260]. This means that when hsa-mir-130b is knocked out in the CRISPR screen, cyclin-G2 levels rise and initiate cell cycle arrest. This could explain the results seen in this study. Another target of miR-130b-3p is phosphatase and tensin homolog (PTEN) [259]. PTEN is an inhibitor of the PI3K/Akt signalling pathway and a known tumour suppressor. De-repression of PTEN via hsa-mir-130b knockout would decrease PI3K/Akt signalling, thus decreasing pro-survival/proliferation signalling.

There are conflicting reports in the literature regarding the effect of miR-769 on cancer [261–264], though I was unable to find studies which used breast cancer as a model. It is possible that the effect of miR-769 is dependent on the cellular context, so further research is required to elucidate the role of miR-769 in TNBC.

Due to the pLX-miR library being the only one available, it was selected for use despite some shortcomings. Firstly, all control sgRNAs are non-targeting. This means they cannot be used to create a null distribution as the lack of double-strand breaks in these cells gives them a survival advantage compared to any cell with a targeting guide. This is demonstrated in their own study, whereby the non-targeting guides display a shift towards a positive $\log_2$ fold change in the polyploid HeLa cells, but not the diploid NCI-N87 cells [97]. To mitigate this, control guides should be used that are still able to cause double strand breaks, but in a location of the DNA that does not have a known function. Additionally, since there are no universal, well-validated essential miRNAs, the lack of positive controls in the library is a limitation. Typically, in protein-coding CRISPR screens, some sgRNAs will be included that target known essential

genes to provide confidence that the screen has worked as expected. To enhance this library, a small selection of sgRNAs should be added that target known essential protein-coding genes that can act as a positive control.

The nature of targeting non-coding RNAs using CRISPR adds difficulty due to the lack of a frame shift effect. In protein-coding genes, small 1-2 bp insertion/deletion events can cause dramatic changes in the final protein because of the frame shift that ensues. When RNA is the final functional product, frame shift cannot occur. Although research has shown that small insertion/deletion events can still cause knockout of miRNAs [207], it is still likely that the general guide efficiency is lower in non-coding CRISPR screens than in their coding counterparts. To mitigate this issue, I chose to use a high coverage in my screen (750x). This is above the 200x coverage used by Kurata and Lin, and the commonly recommended coverage of 500x [97, 213]. The extra power gained from this should help to counter some of the effects of reduced guide efficiency, and since the library is relatively small, the number of cells required for coverage was still manageable. I also used the MAGeCK RRA algorithm to identify CRISPR hits as this was found to function best with low guide efficiencies [222].

Identifying miRNAs of interest using CRISPR screening is also confounded by the existence of miRNA families. There is redundancy in function within a miRNA family, meaning that if one family member is knocked out, others may be able to compensate for the lack of function [265]. Another limitation of targeting miRNAs by CRISPR on a large scale is that approximately 76% of miRNAs lie within other genes or regulatory elements [97]. This means that when the miRNA gene is mutated, it can also affect the function of the surrounding gene(s). This is unavoidable, but in small studies which focus on one miRNA, the relevant assays can be conducted to ensure the surrounding genes are not impacted (e.g. qPCR of the miRNA of interest and the neighbouring gene to ensure only the miRNA levels are reduced). For large scale screens where this is not possible, investigation of the relevant genomic loci can indicate the most likely true/false positives, identifying the most suitable hits for validation.

Future work would involve attempting to validate miR-130b and miR-769 as essential miRNAs in MDA-MB-231 cells. Creating individual knockouts of these miRNAs using the sgRNAs that had the greatest effect then measuring the effect on growth would be the simplest validation. Additionally, to prove that the effects are not caused by secondary effects at the double-strand

break site, traditional miRNA inhibition methods should be used to complement this. Steric blocking of miRNAs using anti-miR oligonucleotides or miRNA sponging could be used to inhibit their function without targeting the genomic DNA [41, 266, 267]. An important point to note is that that these methods will only target one arm of the miRNA (i.e. either 3p or 5p form). This is in contrast to CRISPR knockout which prevents production of both arms. Therefore, prior to validation experiments, the arm which is most likely to be the effector in this context should be identified.

Should the miRNA validate as an essential miRNA, follow up experiments to identify the gene targets and mechanism of action could ensue. Direct targets may be validated by engineering the 3' UTR of the predicted target on to a luciferase reporter. If the gene is a direct target of the miRNA, inhibition of the miRNA will increase luciferase reporter activity and over-expression of the miRNA will decrease it.

miRNAs are possible drug targets and are easy to target with a high degree of specificity using competing oligonucleotides. This also provides a way to increase protein expression, something which is generally difficult to do therapeutically (i.e. most approved drugs are small molecules which are predominantly inhibitory). This means we can aim to increase expression of tumour suppressor genes through targeting miRNAs, rather than being limited to inhibiting oncogenes. The greatest problem faced in targeting miRNAs therapeutically is delivering the competing oligonucleotides to the cell *in vivo* due to the instability of oligonucleotides in biological fluids. Advances have been made in this area, with chemical modifications such as locked nucleic acids or molecular scaffolds used to increase stability [150, 209] and attached aptamers used to improve targeting to cancer cells [209]. Lipid based nanoparticles can also be used for packaging oligonucleotides which increase their half-life [150, 268]. As well as oligonucleotides, small molecules have been designed with high specificity towards oncogenic miRNAs. These can inhibit their processing [269], or target nucleases to trigger their degradation [270]. Given the new plausibility of targeting miRNAs in cancer therapeutics, identifying such miRNA targets in relevant disease models is of increasing importance.

7 Final discussion

The first aim of this project was to characterise the transcriptional response to chronic hypoxia in TNBC. The classical hypoxia response, which is largely driven by HIF, was initiated by the 24 hour timepoint and remained active over the full 14 day time course. Genes which were differentially regulated under chronic hypoxia across multiple cell lines were predominantly involved in adhesion and/or EMT. Validation experiments implied that an atypical form of EMT was occurring under chronic hypoxia in some TNBC cell lines, which could represent a hybrid E/M state.

None of the transcriptional responses to acute or chronic hypoxia were found to be essential for cell survival under hypoxia in this study. In general, there was very little overlap in genes that were selectively essential under hypoxia between cell lines. Only two genes were identified as hits (adjusted p value $\leqslant 0.05$ and $\log_2$ fold change $\leqslant$ -1) in both TNBC cell lines: GCN2 and GPI. GCN2 is an activator of the ISR and is triggered by amino acid starvation [57]. Here, I showed that GCN2 was activated under chronic hypoxia, though whether this was indirectly caused by amino acid starvation was not tested. Nutrient deprivation in my model is unlikely due to the excess of media and frequent media changes employed, but nevertheless, amino acid depletion cannot be ruled out. Validation experiments showed that chemical inhibition of GCN2 reduced cell growth, and this appeared to occur to a greater extent in cells under chronic hypoxia.

Overall, GCN2 constitutes a promising therapeutic target. It is overexpressed in many cancers [55, 184, 186], and analysis of DepMap data has shown GCN2 dependency in 13% of cancer cell lines, which is greater than the dependency on any of the other eIF2α kinases [189]. Inhibition of the ISR downstream of GCN2 has been shown to reduce cancer stem cell phenotypes and increase sensitivity to paclitaxel *in vivo* [271]. Crucially, GCN2 knockout mice are developmentally normal unless fed a diet restricted in amino acids [55, 187]. This means a GCN2 targeting drug is likely to have a good toxicity profile and therapeutic window. This is compounded by GCN2 being active only under stress conditions, as this means GCN2 inhibitors would potentially have a greater impact in the tumour microenvironment, where stressors exist, than in healthy tissue. The implication of this is that GCN2 inhibitors may

actually function better *in vivo* than expected, provided they can gain access to the tumour. This is a known challenge in targeting hypoxic tumours as they, by definition, have poor access to vasculature and therefore drugs delivered via the bloodstream. The efficacy of targeting GCN2 in cancer has been shown in mice [272] and a GCN2 inhibitor is currently in phase I clinical trials for the treatment of advanced solid tumours [273].

No miRNAs were found to be significant for hypoxia specific survival in this study. However, this could be due to a lack of power in the experiment. The time spent under hypoxia during the screen may not have been sufficient for essential miRNAs in hypoxia to be identified. Longer incubation periods would allow any differences to be observed. Some miRNAs which appeared to be essential in MDA-MB-231 cells under both normoxia and hypoxia were identified, one of which was also detected by Kurata and Lin [97].

In general, there was little overlap between the cell lines tested in both the hypoxia driven transcriptional response and gene essentiality. This is in agreement with other researchers who found great discrepancies between cell lines [112, 178, 192]. This is representative of both the noise in large scale assays and innate differences between cell lines. This emphasises the importance of not relying on a single cell line as a model. Multiple cell lines must be tested and, where possible, data from large repositories (e.g. TCGA) should be utilised.

The three main chapters of this thesis have produced results which do not greatly intersect. However, one key cellular process that is implicated in both EMT and the ISR is autophagy. Autophagy, though sometimes playing opposing roles in tumourigenesis [190], is known to be protective under stress conditions in cancer and can drive the development of cancer stem cell properties [71, 190, 274]. The activation of autophagy is a key function of the ISR which enables stress tolerance by the cell. Additionally, autophagy was shown to increase when EMT was stimulated in MCF7 cells [275], while autophagy was required for maintenance of the mesenchymal phenotype in MDA-MB-231 cells [276]. This links the upregulation of autophagy as a possible linking factor between the results of the transcriptomic analysis and the genome-wide CRISPR screen. Of interest, the tentative hypoxia-specific hits in the miRNA CRISPR screen (miR-1299 and miR-3127) are also implicated in autophagy [248, 252]. Together, the data in these chapters suggest that modulation of autophagy through different cell mechanisms is key in driving adaptation to chronic hypoxia exposure.

7.1 Limitations

One limitation of this research is that all experiments were conducted using cell lines cultured in a monolayer. This is far removed from the microenvironment of the tumour, which is comprised of a multitude of cells and structural/extracellular proteins. There is also a lack of perfusion in cell culture which means nutrients are depleted over time while waste products accumulate. Again, this is not representative of the tumour microenvironment. An attempt was made to mitigate this by using excess media volumes and changing media frequently. However, this also meant that my cells were unlikely to be in a state of nutrition deprivation as cells in a hypoxic tumour might be. Although 3D cell culture can allow researchers to mimic the hypoxic core of a tumour, this approach was not suitable for this project as it would not allow clear separation between the normoxic and hypoxic conditions. The nature of genome-wide CRISPR screening requires many cells to be cultured and this largely prohibits the use of complex culture methods such as these. A panel of cell lines was utilised in this project and results selected which were common among them. This enables results which more likely to be relevant across systems, and *in vivo*, to be identified.

Another limitation of this research is the use of atmospheric oxygen levels as a control. As shown in Figure 1, the level of oxygen found in tissues in the body ($<$ 10% O_2) is not the same as that found in the atmosphere (21% O_2). This means that the controls used across these experiments were hyperoxic compared to even well vascularised, healthy breast tissue. This has the potential to skew experimental results, but is unfortunately a widespread issue across hypoxia research. For this study, the use of physiological oxygen concentrations for control samples was prohibited by the availability of hypoxia chambers.

7.2 Future work

Future work should focus on elucidating the mechanism by which GCN2 is activated under hypoxia and how it confers its protective activity. To identify if amino acid depletion was in fact causing GCN2 activation in my model, spent culture media could be collected and the amino acid levels determined by mass spectrometry. Alternatively, supplementing extra amino acids to see whether this rescues the growth perturbation could indirectly shed light upon the activation mechanism of GCN2. RNA-seq experiments on GCN2 knockout cells

could identify pathways that are activated by GCN2 specifically under hypoxia to elucidate the downstream players in GCN2 mediated selection. The importance of GCN2 *in vivo* should also be investigated. This could involve monitoring p-GCN2 levels in xenografts alongside markers of tumour hypoxia to show whether GCN2 is active in tumours.

Finally, a greater body of research on the chronic hypoxia response is needed. This should ideally encompass many cell types and models so that any common adaptive mechanisms can be identified.

8 References

[1] Sung, H, Ferlay, J, Siegel, RL, Laversanne, M, Soerjomataram, I, Jemal, A, and Bray, F. *Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.* CA: A Cancer Journal for Clinicians, (2021). pp. 209–249. https://doi.org/10.3322/caac.21660.

[2] UK, CR. *Breast cancer statistics.* (2017). https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer#heading-Zero.

[3] SEER. *Female Breast Cancer Subtypes - Cancer Stat Facts.* (2020). https://seer.cancer.gov/statfacts/html/breast-subtypes.html.

[4] Yin, L, Duan, JJ, Bian, XW, and Yu, SC. *Triple-negative breast cancer molecular subtyping and treatment progress.* Breast Cancer Research, (2020). https://doi.org/10.1186/s13058-020-01296-5.

[5] Hsu, JY, Chang, CJ, and Cheng, JS. *Survival, treatment regimens and medical costs of women newly diagnosed with metastatic triple-negative breast cancer.* Scientific Reports, (2022). pp. 729. https://doi.org/10.1038/s41598-021-04316-2.

[6] Lehmann, BD, Pietenpol, JA, and Tan, AR. *Triple-negative breast cancer: molecular subtypes and new targets for therapy.* American Society of Clinical Oncology Educational Book, (2015). pp. e31–e39. https://doi.org/10.14694/EdBook_AM.2015.35.e31.

[7] Vtorushin, S, Dulesova, A, and Krakhmal, N. *Luminal androgen receptor (LAR) subtype of triple-negative breast cancer: molecular, morphological, and clinical features.* Journal of Zhejiang University, (2022). pp. 617–624. https://doi.org/10.1631/jzus.B2200113.

[8] Muz, B, Puente, P de la, Azab, F, and Kareem Azab, A. *The role of hypoxia in cancer progression, angiogenesis, metastasis, and resistance to therapy.* Hypoxia, (2015). pp. 83–92. https://doi.org/10.2147/HP.S93413.

[9] Liu, Q, Palmgren, VA, Danen, EH, and Le Dévédec, SE. *Acute vs. chronic vs. intermittent hypoxia in breast Cancer: a review on its application in in vitro research.*

Molecular Biology Reports, (2022). pp. 10961–10973. https://doi.org/10.1007/s11033-022-07802-6.

[10] Krock, BL, Skuli, N, and Simon, MC. *Hypoxia-induced angiogenesis: good and evil.* Genes & Cancer, (2011). pp. 1117–1133. https://doi.org/10.1177/1947601911423654.

[11] Tasoulis, MK and Douzinas, EE. *Hypoxemic reperfusion of ischemic states: an alternative approach for the attenuation of oxidative stress mediated reperfusion injury.* Journal of Biomedical Science, (2016). pp. 1–8. https://doi.org/10.1186/s12929-016-0220-0.

[12] Dengler, VL, Galbraith, MD, and Espinosa, JM. *Transcriptional regulation by hypoxia inducible factors.* Critical Reviews in Biochemistry and Molecular Biology, (2014). pp. 1–15. https://doi.org/10.3109/10409238.2013.838205.

[13] Bayer, C and Vaupel, P. *Acute versus chronic hypoxia in tumors. Controversial data concerning time frames and biological consequences.* Strahlentherapie Und Onkologie, (2012). pp. 616–627. https://doi.org/10.1007/s00066-012-0085-4.

[14] Emami Nejad, A, Najafgholian, S, Rostami, A, Sistani, A, Shojaeifar, S, Esparvarinha, M, Nedaeinia, R, Haghjooy Javanmard, S, Taherian, M, Ahmadlou, M, et al. *The role of hypoxia in the tumor microenvironment and development of cancer stem cell: a novel approach to developing treatment.* Cancer Cell International, (2021). pp. 1–26. https://doi.org/10.1186/s12935-020-01719-5.

[15] Al Tameemi, W, Dale, TP, Al-Jumaily, RMK, and Forsyth, NR. *Hypoxia-modified cancer cell metabolism.* Frontiers in Cell and Developmental Biology, (2019). pp. 4. https://doi.org/10.3389/fcell.2019.00004.

[16] Sullivan, M, Galea, P, and Latif, S. *What is the appropriate oxygen tension for in vitro culture?* Molecular Human Reproduction, (2006). pp. 653–653. https://doi.org/10.1093/molehr/gal081.

[17] Nabavizadeh, SA, Akbari, H, Ware, JB, Nasrallah, M, Guiry, S, Bagley, SJ, Desai, A, Levy, S, Sarchiapone, W, Prior, T, et al. *Arterial spin labeling and dynamic susceptibility contrast-enhanced MR imaging for evaluation of arteriovenous shunting and tumor hypoxia in glioblastoma.* Scientific reports, (2019). pp. 8747. https://doi.org/10.1038/s41598-019-45312-x.

[18] Vaupel, P and Mayer, A. *Hypoxia in tumors: pathogenesis-related classification, characterization of hypoxia subtypes, and associated biological and clinical implications.* Oxygen transport to tissue XXXVI, (2014). pp. 19–24. https://doi.org/10.1007/978-1-4939-0620-8_3.

[19] Shrieve, DC and Harris, JW. *The in vitro sensitivity of chronically hypoxic EMT6/SF cells to X-radiation and hypoxic cell radiosensitizers.* International Journal of Radiation Biology and Related Studies in Physics, Chemistry and Medicine, (1985). pp. 127–138. https://doi.org/10.1080/09553008514551131.

[20] Sweet, R, Paul, A, and Zastre, J. *Hypoxia induced upregulation and function of the thiamine transporter, SLC19A3 in a breast cancer cell line.* Cancer Biology & Therapy, (2010). pp. 1101–1111. https://doi.org/10.4161/cbt.10.11.13444.

[21] Pires, I, Bencokova, Z, Milani, M, Folkes, L, Li, J, Stratford, M, Harris, A, and Hammond, E. *Effects of acute versus chronic hypoxia on DNA damage responses and genomic instability.* Cancer Research, (2010). https://doi.org/10.1158/0008-5472.CAN-09-2715.

[22] Alqawi, O, Wang, HP, Espiritu, M, and Singh, G. *Chronic hypoxia promotes an aggressive phenotype in rat prostate cancer cells.* Free Radical Research, (2007). pp. 788–797. https://doi.org/10.1080/10715760701361531.

[23] Kindrick, JD and Mole, DR. *Hypoxic regulation of gene transcription and chromatin: cause and effect.* International Journal of Molecular Sciences, (2020). pp. 8320. https://doi.org/10.3390/ijms21218320.

[24] Bartoszewski, R, Moszynska, A, Serocki, M, Cabaj, A, Polten, A, Ochocka, R, Dell'Italia, L, Bartoszewska, S, Kroliczewski, J, Dabrowski, MD, and Collawn, JF. *Primary endothelial cell-specific regulation of hypoxia-inducible factor (HIF)-1 and HIF-2 and their target gene expression profiles during hypoxia.* The FASEB Journal, (2019). pp. 7929–7941. https://doi.org/10.1096/fj.201802650RR.

[25] Koh, MY and Powis, G. *Passing the baton: the HIF switch.* Trends in Biochemical Sciences, (2012). pp. 364–372. https://doi.org/10.1016/j.tibs.2012.06.004.

[26] Michiels, C. *Physiological and pathological responses to hypoxia.* The American Journal of Pathology, (2004). pp. 1875–1882. https://doi.org/10.1016/S0002-9440(10)63747-9.

[27] Diao, X, Ye, F, Zhang, M, Ren, X, Tian, X, Lu, J, Sun, X, Hou, Z, Chen, X, Li, F, et al. *Identification of oleoylethanolamide as an endogenous ligand for HIF-3α.* Nature Communications, (2022). pp. 2529. https://doi.org/10.1038/s41467-022-30338-z.

[28] Keith, B, Johnson, RS, and Simon, MC. *HIF1α and HIF2α: sibling rivalry in hypoxic tumour growth and progression.* Nature Reviews Cancer, (2012). pp. 9–22. https://doi.org/10.1038/nrc3183.

[29] Abd-Aziz, N, Stanbridge, EJ, and Shafee, N. *Bortezomib attenuates HIF-1-but not HIF-2-mediated transcriptional activation.* Oncology Letters, (2015). pp. 2192–2196. https://doi.org/10.3892/ol.2015.3545.

[30] Sowter, HM, Raval, R, Moore, J, Ratcliffe, PJ, and Harris, AL. *Predominant role of hypoxia-inducible transcription factor (HIF)-1α versus Hif-2α in regulation of the transcriptional response to hypoxia.* Cancer Research, (2003). pp. 6130–6134.

[31] Janaszak-Jasiecka, A, Bartoszewska, S, Kochan, K, Piotrowski, A, Kalinowski, L, Kamysz, W, Ochocka, RJ, Bartoszewski, R, and Collawn, JF. *miR-429 regulates the transition between Hypoxia-Inducible Factor HIF1A and HIF3A expression in human endothelial cells.* Scientific Reports, (2016). pp. 22775–22775. https://doi.org/10.1038/srep22775.

[32] Serocki, M, Bartoszewska, S, Janaszak-Jasiecka, A, Ochocka, RJ, Collawn, JF, and Bartoszewski, R. *miRNAs regulate the HIF switch during hypoxia: a novel therapeutic target.* Angiogenesis, (2018). pp. 183–202. https://doi.org/10.1007/s10456-018-9600-2.

[33] Ravenna, L, Principessa, L, Verdina, A, Salvatori, L, Russo, MA, and Petrangeli, E. *Distinct phenotypes of human prostate cancer cells associate with different adaptation to hypoxia and pro-inflammatory gene expression.* PloS One, (2014). pp. e96250–e96250. https://doi.org/10.1371/journal.pone.0096250.

[34] Batie, M and Rocha, S. *JmjC histone demethylases act as chromatin oxygen sensors.* Molecular & Cellular Oncology, (2019). pp. 1608501. https://doi.org/10.1080/23723556.2019.1608501.

[35] Batie, M, Del Peso, L, and Rocha, S. *Hypoxia and chromatin: a focus on transcriptional repression mechanisms.* Biomedicines, (2018). pp. 47. https://doi.org/10.3390/biomedicines6020047.

[36] Chi, JT, Wang, Z, Nuyten, DSA, Rodriguez, EH, Schaner, ME, Salim, A, Wang, Y, Kristensen, GB, Helland, A, Børresen-Dale, AL, et al. *Gene expression programs in response to hypoxia: cell type specificity and prognostic significance in human cancers.* PLoS Medicine, (2006). pp. e47. https://doi.org/10.1371/journal.pmed.0030047.

[37] Abou Khouzam, R, Sharda, M, Rao, SP, Kyerewah-Kersi, SM, Zeinelabdin, NA, Mahmood, AS, Nawafleh, H, Khan, MS, Venkatesh, GH, and Chouaib, S. *Chronic hypoxia is associated with transcriptomic reprogramming and increased genomic instability in cancer cells.* Frontiers in Cell and Developmental Biology, (2023). pp. 1095419. https://doi.org/10.3389/fcell.2023.1095419.

[38] Farina, AR, Cappabianca, L, Sebastiano, M, Zelli, V, Guadagni, S, and Mackay, AR. *Hypoxia-induced alternative splicing: the 11th Hallmark of Cancer.* Journal of Experimental & Clinical Cancer Research, (2020). pp. 1–30. https://doi.org/0.1186/s13046-020-01616-9.

[39] Han, J, Li, J, Ho, JC, Chia, GS, Kato, H, Jha, S, Yang, H, Poellinger, L, and Lee, KL. *Hypoxia is a key driver of alternative splicing in human breast cancer cells.* Scientific Reports, (2017). pp. 4108. https://doi.org/10.1038/s41598-017-04333-0.

[40] Shen, G, Li, X, Jia, Yf, Piazza, GA, and Xi, Y. *Hypoxia-regulated microRNAs in human cancer.* Acta Pharmacologica Sinica, (2013). pp. 336–341. https://doi.org/10.1038/aps.2012.195.

[41] Lima, JF, Cerqueira, L, Figueiredo, C, Oliveira, C, and Azevedo, NF. *Anti-miRNA oligonucleotides: A comprehensive guide for design.* RNA Biology, (2018). pp. 338–352. https://doi.org/10.1080/15476286.2018.1445959.

[42] O'Brien, J, Hayder, H, Zayed, Y, and Peng, C. *Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation.* Frontiers in Endocrinology, (2018). https://doi.org/10.3389/fendo.2018.00402.

[43] Choudhry, H and Mole, DR. *Hypoxic regulation of the noncoding genome and NEAT1.* Briefings in Functional Genomics, (2016). pp. 174–185. https://doi.org/10.1093/bfgp/elv050.

[44] Huang, X and Zuo, JH. *Emerging roles of miR-210 and other non-coding RNAs in the hypoxic response.* Acta Biochimica Et Biophysica Sinica, (2014). pp. 220–232. https://doi.org/10.1093/abbs/gmt141.

[45] Jin, FF, Wang, YB, Zhu, YN, Li, S, Liu, Y, Chen, C, Wang, XH, Zen, K, and Li, LM. *The miR-125a/HK2 axis regulates cancer cell energy metabolism reprogramming in hepatocellular carcinoma.* Scientific Reports, (2017). https://doi.org/10.1038/s41598-017-03407-3.

[46] Ma, JL, Guo, WL, and Chen, XM. *Overexpressing microRNA-150 attenuates hypoxia-induced human cardiomyocyte cell apoptosis by targeting glucose-regulated protein-94.* Molecular Medicine Reports, (2018). pp. 4181–4186. https://doi.org/10.3892/mmr.2018.8375.

[47] Wouters, BG and Koritzinsky, M. *Hypoxia signalling through mTOR and the unfolded protein response in cancer.* Nature Reviews Cancer, (2008). pp. 851–864. https://doi.org/10.1038/nrc2501.

[48] Chipurupalli, S, Kannan, E, Tergaonkar, V, D'Andrea, R, and Robinson, N. *Hypoxia induced ER stress response as an adaptive mechanism in cancer.* International Journal of Molecular Sciences, (2019). pp. 749. https://doi.org/10.3390/ijms20030749.

[49] Pakos-Zebrucka, K, Koryga, I, Mnich, K, Ljujic, M, Samali, A, and Gorman, AM. *The integrated stress response.* EMBO Reports, (2016). pp. pp1374–1395. https://doi.org/10.15252/embr.201642195.

[50] Somers, J, Pöyry, T, and Willis, AE. *A perspective on mammalian upstream open reading frame function.* The International Journal of Biochemistry & Cell Biology, (2013). pp. 1690–1700. https://doi.org/10.1016/j.biocel.2013.04.020.

[51] Liu, Y, László, C, Liu, Y, Liu, W, Chen, X, Evans, S, and Wu, S. *Regulation of G1 arrest and apoptosis in hypoxia by PERK and GCN2-mediated eIF2α phosphorylation.* Neoplasia, (2010). pp. 61–68. https://doi.org/10.1593/neo.91354.

[52] Neill, G and Masson, GR. *A stay of execution: ATF4 regulation and potential outcomes for the integrated stress response.* Frontiers in Molecular Neuroscience, (2023). pp. 1112253–1112253. https://doi.org/10.3389/fnmol.2023.1112253.

[53] Bartoszewska, S and Collawn, JF. *Unfolded protein response (UPR) integrated signaling networks determine cell fate during hypoxia.* Cellular & Molecular Biology Letters, (2020). pp. 18–18. https://doi.org/10.1186/s11658-020-00212-1.

[54] Wang, P, Li, J, Tao, J, and Sha, B. *The luminal domain of the ER stress sensor protein PERK binds misfolded proteins and thereby triggers PERK oligomerization.* The Journal of Biological Chemistry, (2018). pp. 4110–4121. https://doi.org/10.1074/jbc.RA117.001294.

[55] Ye, J, Kumanova, M, Hart, LS, Sloane, K, Zhang, H, De Panis, DN, Bobrovnikova-Marjon, E, Diehl, JA, Ron, D, and Koumenis, C. *The GCN2-ATF4 pathway is critical for tumour cell survival and proliferation in response to nutrient deprivation.* EMBO Journal, (2010). pp. 2082–2096. https://doi.org/10.1038/emboj.2010.81.

[56] Hu, XQ and Zhang, L. *Hypoxia and the integrated stress response promote pulmonary hypertension and preeclampsia: Implications in drug development.* Drug Discovery Today, (2021). pp. 2754–2773. https://doi.org/10.1016/j.drudis.2021.07.011.

[57] Gold, LT and Masson, GR. *GCN2: roles in tumour development and progression.* Biochemical Society Transactions, (2022). pp. 737–745. https://doi.org/10.1042/BST20211252.

[58] JIANG, HY and WEK, RC. *GCN2 phosphorylation of eIF2α activates NF-κB in response to UV irradiation.* Biochemical Journal, (2005). pp. 371–380. https://doi.org/10.1042/BJ20041164.

[59] Deng, J, Harding, HP, Raught, B, Gingras, AC, Berlanga, JJ, Scheuner, D, Kaufman, RJ, Ron, D, and Sonenberg, N. *Activation of GCN2 in UV-Irradiated Cells Inhibits Translation.* Current Biology, (2002). pp. 1279–1286. https://doi.org/10.1016/s0960-9822(02)01037-0.

[60] Inglis, AJ, Masson, GR, Shao, S, Perisic, O, McLaughlin, SH, Hegde, RS, and Williams, RL. *Activation of GCN2 by the ribosomal P-stalk.* Proceedings of the National Academy of Sciences, (2019). pp. 4946–4954. https://doi.org/10.1073/pnas.1813352116.

[61] Ishimura, R, Nagy, G, Dotu, I, Chuang, JH, and Ackerman, SL. *Activation of GCN2 kinase by ribosome stalling links translation elongation with translation initiation.* eLife, (2016). https://doi.org/10.7554/eLife.14295.

[62] Wang, SF, Chen, MS, Chou, YC, Ueng, YF, Yin, PH, Yeh, TS, and Lee, HC. *Mitochondrial dysfunction enhances cisplatin resistance in human gastric cancer cells via the ROS-activated GCN2-eIF2α-ATF4-xCT pathway.* Oncotarget, (2016). pp. 74132–74151. https://doi.org/10.18632/oncotarget.12356.

[63] Bi, M, Naczki, C, Koritzinsky, M, Fels, D, Blais, J, Hu, N, Harding, H, Novoa, I, Varia, M, Raleigh, J, Scheuner, D, Kaufman, RJ, Bell, J, Ron, D, Wouters, BG, and Koumenis, C. *ER stress-regulated translation increases tolerance to extreme hypoxia and promotes tumor growth.* EMBO Journal, (2005). pp. 3470–3481. https://doi.org/10.1038/sj.emboj.7600777.

[64] Boulefour, W, Rowinski, E, Louati, S, Sotton, S, Wozny, AS, Moreno-Acosta, P, Mery, B, Rodriguez-Lafrasse, C, and Magne, N. *A review of the role of hypoxia in radioresistance in cancer therapy.* Medical Science Monitor: International Medical Journal of Experimental and Clinical Research, (2021). pp. e934116–1. https://doi.org/10.12659/MSM.934116.

[65] Graham, K and Unger, E. *Overcoming tumor hypoxia as a barrier to radiotherapy, chemotherapy and immunotherapy in cancer treatment.* International Journal of Nanomedicine, (2018). pp. 6049–6058. https://doi.org/10.2147/IJN.S140462.

[66] Li, D, Zhou, L, Huang, J, and Xiao, X. *Effect of multidrug resistance 1/P-glycoprotein on the hypoxia-induced multidrug resistance of human laryngeal cancer cells.* Oncology Letters, (2016). pp. 1569–1574. https://doi.org/10.3892/ol.2016.4749.

[67] Lv, Y, Zhao, S, Han, J, Zheng, L, Yang, Z, and Zhao, L. *Hypoxia-inducible factor-1α induces multidrug resistance protein in colon cancer.* OncoTargets and Therapy, (2015). pp. 1941–1948. https://doi.org/10.2147/OTT.S82835.

[68] Liu, XW, Su, Y, Zhu, H, Cao, J, Ding, WJ, Zhao, YC, He, QJ, and Yang, B. *HIF-1α-dependent autophagy protects HeLa cells from fenretinide (4-HPR)-induced apoptosis in hypoxia.* Pharmacological Research, (2010). pp. 416–425. https://doi.org/10.1016/j.phrs.2010.07.002.

[69] Sun, Y, Xing, X, Liu, Q, Wang, Z, Xin, Y, Zhang, P, Hu, C, and Liu, Y. *Hypoxia-induced autophagy reduces radiosensitivity by the HIF-1α/miR-210/Bcl-2 pathway in colon cancer cells.* International Journal of Oncology, (2015). pp. 750–756. https://doi.org/10.3892/ijo.2014.2745.

[70] Hanahan, D and Weinberg, RA. *The hallmarks of cancer.* Cell, (2000). pp. 57–70. https://doi.org/10.1016/S0092-8674(00)81683-9.

[71] Hanahan, D and Weinberg, RA. *Hallmarks of cancer: the next generation.* Cell, (2011). pp. 646–674. https://doi.org/10.1016/j.cell.2011.02.013.

[72] Hanahan, D. *Hallmarks of cancer: new dimensions.* Cancer discovery, (2022). pp. 31–46. https://doi.org/10.1158/2159-8290.CD-21-1059.

[73] Sermeus, A, Genin, M, Maincent, A, Fransolet, M, Notte, A, Leclere, L, Riquier, H, Arnould, T, and Michiels, C. *Hypoxia-induced modulation of apoptosis and BCL-2 family proteins in different cancer cell types.* PloS One, (2012). pp. e47519. https://doi.org/10.1371/journal.pone.0047519.

[74] Zhang, C, Liu, J, Wang, J, Zhang, T, Xu, D, Hu, W, and Feng, Z. *The interplay between tumor suppressor p53 and hypoxia signaling pathways in cancer.* Frontiers in Cell and Developmental Biology, (2021). pp. 648808. https://doi.org/10.3389/fcell.2021.648808.

[75] Dong, Z, Venkatachalam, MA, Wang, J, Patel, Y, Saikumar, P, Semenza, GL, Force, T, and Nishiyama, J. *Up-regulation of apoptosis inhibitory protein IAP-2 by hypoxia: HIF-1-independent mechanisms.* Journal of Biological Chemistry, (2001). pp. 18702–18709. https://doi.org/10.1074/jbc.M011774200.

[76] Bhandari, V et al. *Molecular landmarks of tumor hypoxia across cancer types.* Nature Genetics, (2019). pp. 308–+. https://doi.org/10.1038/s41588-018-0318-2.

[77] Barratt, SL, Blythe, T, Ourradi, K, Jarrett, C, Welsh, GI, Bates, DO, and Millar, AB. *Effects of hypoxia and hyperoxia on the differential expression of VEGF-A isoforms and receptors in Idiopathic Pulmonary Fibrosis (IPF).* Respiratory Research, (2018). pp. 1–5. https://doi.org/10.1186/s12931-017-0711-x.

[78] Chen, Z, Han, F, Du, Y, Shi, H, and Zhou, W. *Hypoxic microenvironment in cancer: molecular mechanisms and therapeutic interventions.* Signal Transduction and Targeted Therapy, (2023). pp. 70. https://doi.org/10.1038/s41392-023-01332-8.

[79] Choi, JY, Jang, YS, Min, SY, and Song, JY. *Overexpression of MMP-9 and HIF-1α in breast cancer cells under hypoxic conditions.* Journal of Breast Cancer, (2011). pp. 88–95. https://doi.org/10.4048/jbc.2011.14.2.88.

[80] Yang, J et al. *Guidelines and definitions for research on epithelial-mesenchymal transition.* Nature Reviews Molecular Cell Biology, (2020). pp. 341–352. https://doi.org/10.1038/s41580-020-0237-9.

[81] Debnath, P, Huirem, RS, Dutta, P, and Palchaudhuri, S. *Epithelial-mesenchymal transition and its transcription factors.* Bioscience Reports, (2022). https://doi.org/10.1042/BSR20211754.

[82] Hapke, RY and Haake, SM. *Hypoxia-induced epithelial to mesenchymal transition in cancer.* Cancer Letters, (2020). pp. 10–20. https://doi.org/10.1016/j.canlet.2020.05.012.

[83] Zhang, L, Huang, G, Li, X, Zhang, Y, Jiang, Y, Shen, J, Liu, J, Wang, Q, Zhu, J, Feng, X, et al. *Hypoxia induces epithelial-mesenchymal transition via activation of SNAI1 by hypoxia-inducible factor-1α in hepatocellular carcinoma.* BMC Cancer, (2013). pp. 1–9. https://doi.org/10.1186/1471-2407-13-108.

[84] Tang, T, Yang, Z, Zhu, Q, Wu, Y, Sun, K, Alahdal, M, Zhang, Y, Xing, Y, Shen, Y, Xia, T, Xi, T, Pan, Y, and Jin, L. *Up-regulation of miR-210 induced by a hypoxic microenvironment promotes breast cancer stem cell metastasis, proliferation, and self-renewal by targeting E-cadherin.* The FASEB Journal, (2018). pp. 6965–6981. https://doi.org/10.1096/fj.201801013R.

[85] Aiello, NM and Kang, Y. *Context-dependent EMT programs in cancer metastasis.* Journal of Experimental Medicine, (2019). pp. 1016–1026. https://doi.org/10.1084/jem.20181827.

[86] Peyre, L, Meyer, M, Hofman, P, and Roux, J. *TRAIL receptor-induced features of epithelial-to-mesenchymal transition increase tumour phenotypic heterogeneity: potential cell survival mechanisms.* British Journal of Cancer, (2021). pp. 91–101. https://doi.org/10.1038/s41416-020-01177-w.

[87] Song, KA, Niederst, MJ, Lochmann, TL, Hata, AN, Kitai, H, Ham, J, Floros, KV, Hicks, MA, Hu, H, Mulvey, HE, et al. *Epithelial-to-mesenchymal transition antagonizes response to targeted therapies in lung cancer by suppressing BIM.* Clinical Cancer Research, (2018). pp. 197–208. https://doi.org/10.1158/1078-0432.CCR-17-1577.

[88] Kwok, WK, Ling, MT, Lee, TW, Lau, TC, Zhou, C, Zhang, X, Chua, CW, Chan, KW, Chan, FL, Glackin, C, et al. *Up-regulation of TWIST in prostate cancer and its implication as a therapeutic target.* Cancer Research, (2005). pp. 5153–5162. https://doi.org/10.1158/0008-5472.CAN-04-3785.

[89] Staalduinen, J van, Baker, D, Ten Dijke, P, and Dam, H van. *Epithelial–mesenchymal-transition-inducing transcription factors: new targets for tackling chemoresistance in cancer?* Oncogene, (2018). pp. 6195–6211. https://doi.org/10.1038/s41388-018-0378-x.

[90] Hassan Venkatesh, G, Bravo, P, Shaaban Moustafa Elsayed, W, Amirtharaj, F, Wojtas, B, Abou Khouzam, R, Hussein Nawafleh, H, Mallya, S, Satyamoorthy, K, Dessen, P, et al. *Hypoxia increases mutational load of breast cancer cells through frameshift mutations.* Oncoimmunology, (2020). pp. 1750750. https://doi.org/10.1080/2162402X.2020.1750750.

[91] Bhandari, V, Li, CH, Bristow, RG, and Boutros, PC. *Divergent mutational processes distinguish hypoxic and normoxic tumours.* Nature Communications, (2020). pp. 737. https://doi.org/10.1038/s41467-019-14052-x.

[92] Crosby, ME, Kulshreshtha, R, Ivan, M, and Glazer, PM. *MicroRNA regulation of DNA repair gene expression in hypoxic stress.* Cancer Research, (2009). pp. 1221–1229. https://doi.org/10.1158/0008-5472.CAN-08-2516.

[93] Vilaplana-Lopera, N, Besh, M, and Moon, EJ. *Targeting hypoxia: Revival of old remedies.* Biomolecules, (2021). pp. 1604. https://doi.org/10.3390/biom11111604.

[94] Spiegelberg, L, Houben, R, Niemans, R, Ruysscher, D de, Yaromina, A, Theys, J, Guise, CP, Smaill, JB, Patterson, AV, Lambin, P, et al. *Hypoxia-activated prodrugs and (lack of) clinical progress: The need for hypoxia-based biomarker patient selection in phase III clinical trials.* Clinical and Translational Radiation Oncology, (2019). pp. 62–69. https://doi.org/10.1016/j.ctro.2019.01.005.

[95] Li, Y, Zhao, L, and Li, XF. *Targeting hypoxia: hypoxia-activated prodrugs in cancer therapy.* Frontiers in Oncology, (2021). pp. 700407. https://doi.org/10.3389/fonc.2021.700407.

[96] Liu, B, Shyr, Y, Cai, J, and Liu, Q. *Interplay between miRNAs and host genes and their role in cancer.* Briefings in Functional Genomics, (2019). pp. 255–266. https://doi.org/10.1093/bfgp/elz002.

[97] Kurata, JS and Lin, RJ. *MicroRNA-focused CRISPR-Cas9 library screen reveals fitness-associated miRNAs.* RNA, (2018). pp. 966–981. https://doi.org/10.1261/rna.066282.118.

[98] Rueda, A, Barturen, G, Lebrón, R, Gómez-Martín, C, Alganza, Á, Oliver, JL, and Hackenberg, M. *sRNAtoolbox: an integrated collection of small RNA research tools.* Nucleic Acids Research, (2015). pp. W467–W473. https://doi.org/10.1093/nar/gkv555.

[99] Barturen, G, Rueda, A, Hamberg, M, Alganza, A, Lebron, R, Kotsyfakis, M, Shi, BJ, Koppers-Lalic, D, and Hackenberg, M. *sRNAbench: profiling of small RNAs and its sequence variants in single or multi-species high-throughput experiments.* Methods in Next Generation Sequencing, (2014). pp. 21–31.

[100] Li, J, Kho, AT, Chase, RP, Pantano, L, Farnam, L, Amr, SS, and Tantisira, KG. *COMPSRA: a COMprehensive platform for small RNA-Seq data analysis.* Scientific Reports, (2020). pp. 4552. https://doi.org/10.1038/s41598-020-61495-0.

[101] Ziemann, M, Kaspi, A, and El-Osta, A. *Evaluation of microRNA alignment techniques.* RNA, (2016). pp. 1120–1138. https://doi.org/10.1261/rna.055509.115.

[102] Potla, P, Ali, SA, and Kapoor, M. *A bioinformatics approach to microRNA-sequencing analysis.* Osteoarthritis and Cartilage Open, (2021). pp. 100131. https://doi.org/10.1016/j.ocarto.2020.100131.

[103] Wu, X, Kim, TK, Baxter, D, Scherler, K, Gordon, A, Fong, O, Etheridge, A, Galas, DJ, and Wang, K. *sRNAnalyzer—a flexible and customizable small RNA sequencing data analysis pipeline.* Nucleic Acids Research, (2017). pp. 12140–12151. https://doi.org/10.1093/nar/gkx999.

[104] Deschamps-Francoeur, G, Simoneau, J, and Scott, MS. *Handling multi-mapped reads in RNA-seq.* Computational and Structural Biotechnology Journal, (2020). pp. 1569–1576. https://doi.org/10.1016/j.csbj.2020.06.014.

[105] Handzlik, JE, Tastsoglou, S, Vlachos, IS, and Hatzigeorgiou, AG. *Manatee: detection and quantification of small non-coding RNAs from next-generation sequencing data.* Scientific Reports, (2020). pp. 705. https://doi.org/10.1038/s41598-020-57495-9.

[106] Camps, C, Saini, HK, Mole, DR, Choudhry, H, Reczko, M, Guerra-Assuncao, JA, Tian, YM, Buffa, FM, Harris, AL, Hatzigeorgiou, AG, Enright, AJ, and Ragoussis, J. *Integrated analysis of microRNA and mRNA expression and association with HIF binding reveals the complexity of microRNA expression regulation under hypoxia.* Molecular Cancer, (2014). https://doi.org/10.1186/1476-4598-13-28.

[107] Lee, P, Chandel, NS, and Simon, MC. *Cellular adaptation to hypoxia through hypoxia inducible factors and beyond.* Nature Reviews Molecular Cell Biology, (2020). pp. 268–283. https://doi.org/10.1038/s41580-020-0227-y.

[108] Choudhry, H and Harris, AL. *Advances in hypoxia-inducible factor biology.* Cell Metabolism, (2018). pp. 281–298. https://doi.org/10.1016/j.cmet.2017.10.005.

[109] Carcereri de Prati, A, Butturini, E, Rigo, A, Oppici, E, Rossin, M, Boriero, D, and Mariotto, S. *Metastatic Breast Cancer Cells Enter Into Dormant State and Express Cancer Stem Cells Phenotype Under Chronic Hypoxia.* Journal of Cellular Biochemistry, (2017). pp. 3237–3248. https://doi.org/10.1002/jcb.25972.

[110] Chen, SC, Chen, FW, Hsu, YL, and Kuo, PL. *Systematic Analysis of Transcriptomic Profile of Renal Cell Carcinoma under Long-Term Hypoxia Using Next-Generation Sequencing and Bioinformatics.* International Journal of Molecular Sciences, (2017). pp. 2657. https://doi.org/10.3390/ijms18122657.

[111] Jarman, EJ, Ward, C, Turnbull, AK, Martinez-Perez, C, Meehan, J, Xintaropoulou, C, Sims, AH, and Langdon, SP. *HER2 regulates HIF-2α and drives an increased hypoxic response in breast cancer.* Breast Cancer Research, (2019). pp. 10–10. https://doi.org/10.1186/s13058-019-1097-0.

[112] Danielsson, F, Fasterius, E, Sullivan, D, Hases, L, Sanli, K, Zhang, C, Mardinoglu, A, Al-Khalili, C, Huss, M, Uhlén, M, Williams, C, and Lundberg, E. *Transcriptome profiling of the interconnection of pathways involved in malignant transformation and response to hypoxia.* Oncotarget, (2018). pp. 19730–19744. https://doi.org/10.18632/oncotarget.24808.

[113] Tseng, LM, Chiu, JH, Liu, CY, Tsai, YF, Wang, YL, Yang, CW, and Shyr, YM. *A comparison of the molecular subtypes of triple-negative breast cancer among non-Asian and Taiwanese women.* Breast cancer research and treatment, (2017). pp. 241–254. https://doi.org/10.1007/s10549-017-4195-7.

[114] Barton, VN, D'Amato, NC, Gordon, MA, Lind, HT, Spoelstra, NS, Babbs, BL, Heinz, RE, Elias, A, Jedlicka, P, Jacobsen, BM, et al. *Multiple molecular subtypes of triple-negative breast cancer critically rely on androgen receptor and respond to enzalutamide in vivo.* Molecular cancer therapeutics, (2015). pp. 769–778. https://doi.org/10.1158/1535-7163.MCT-14-0926.

[115] Lamble, S, Batty, E, Attar, M, Buck, D, Bowden, R, Lunter, G, Crook, D, El-Fahmawi, B, and Piazza, P. *Improved workflows for high throughput library preparation using the transposome-based nextera system.* BMC Biotechnology, (2013). https://doi.org/10.1186/1472-6750-13-104.

[116] Buffa, FM, Harris, AL, West, CM, and Miller, CJ. *Large meta-analysis of multiple cancers reveals a common, compact and highly prognostic hypoxia metagene.* British Journal of Cancer, (2010). pp. 428–435. https://doi.org/10.1038/sj.bjc.6605450.

[117] Dhawan, A, Barberis, A, Cheng, WC, Domingo, E, West, C, Maughan, T, Scott, JG, Harris, AL, and Buffa, FM. *Guidelines for using sigQC for systematic evaluation of gene signatures.* Nature Protocols, (2019). pp. 1377–1400. https://doi.org/10.1038/s41596-019-0136-8.

[118] Jain, IH, Calvo, SE, Markhard, AL, Skinner, OS, To, TL, Ast, T, and Mootha, VK. *Genetic Screen for Cell Fitness in High or Low Oxygen Highlights Mitochondrial and Lipid Metabolism.* Cell, (2020). pp. 716–+. https://doi.org/10.1016/j.cell.2020.03.029.

[119] Oki, S and Ohta, T. *ChIP-Atlas.* (2015). https://chip-atlas.org/.

[120] Garcia-Alonso, L, Holland, CH, Ibrahim, MM, Turei, D, and Saez-Rodriguez, J. *Benchmark and integration of resources for the estimation of human transcription factor activities.* Genome Research, (2019). pp. 1363–1375. https://doi.org/10.1101/gr.240663.118.

[121] Lake, DF and Faigel, DO. *The emerging role of QSOX1 in cancer.* Antioxidants & Redox Signaling, (2014). pp. 485–496. https://doi.org/10.1089/ars.2013.5572.

[122] Pernodet, N, Hermetet, F, Adami, P, Vejux, A, Descotes, F, Borg, C, Adams, M, Pallandre, JR, Viennet, G, Esnard, F, Jouvenot, M, and Despouy, G. *High expression of QSOX1 reduces tumorogenesis, and is associated with a better outcome for breast cancer patients.* Breast Cancer Research, (2012). pp. R136–R136. https://doi.org/10.1186/bcr3341.

[123] Katchman, BA, Ocal, IT, Cunliffe, HE, Chang, YH, Hostetter, G, Watanabe, A, LoBello, J, and Lake, DF. *Expression of quiescin sulfhydryl oxidase 1 is associated with a highly invasive phenotype and correlates with a poor prognosis in Luminal B breast cancer.* Breast Cancer Research, (2013). pp. R28–R28. https://doi.org/10.1186/bcr3407.

[124] Ilani, T, Alon, A, Grossman, I, Horowitz, B, Kartvetishvily, E, Cohen, SR, and Fass, D. *A Secreted Disulfide Catalyst Controls Extracellular Matrix Composition and Function.* Science, (2013). pp. 74–76. https://doi.org/10.1126/science.1238279.

[125] Rousselle, P and Scoazec, JY. *Laminin 332 in cancer: When the extracellular matrix turns signals from cell anchorage to cell movement.* Seminars in Cancer Biology, (2020). pp. 149–165. https://doi.org/10.1016/j.semcancer.2019.09.026.

[126] Miyazaki, K. *Laminin-5 (laminin-332): Unique biological activity and role in tumor growth and invasion.* Cancer Science, (2006). pp. 91–98. https://doi.org/10.1111/j.1349-7006.2006.00150.x.

[127] Zeisberg, M and Neilson, EG. *Biomarkers for epithelial-mesenchymal transitions.* The Journal of Clinical Investigation, (2009). pp. 1429–1437. https://doi.org/10.1172/JCI36183.

[128] Carpenter, PM, Dao, AV, Arain, ZS, Chang, MK, Nguyen, HP, Arain, S, Wang-Rodriguez, J, Kwon, SY, and Wilczynski, SP. *Motility Induction in Breast Carcinoma by Mammary Epithelial Laminin 332 (Laminin 5).* Molecular Cancer Research, (2009). pp. 462–475. https://doi.org/10.1158/1541-7786.MCR-08-0148.

[129] Garg, M, Kanojia, D, Okamoto, R, Jain, S, Madan, V, Chien, W, Sampath, A, Ding, LW, Xuan, M, Said, JW, Doan, NB, Liu, LZ, Yang, H, Gery, S, Braunstein, GD, and Koeffler, HP. *Laminin-5γ-2 (LAMC2) Is Highly Expressed in Anaplastic Thyroid Carcinoma and Is Associated With Tumor Progression, Migration, and Invasion by Modulating Signaling of EGFR.* The Journal of Clinical Endocrinology and Metabolism, (2014). pp. E62–E72. https://doi.org/10.1210/jc.2013-2994.

[130] Qi, Y and Xu, R. *Roles of PLODs in Collagen Synthesis and Cancer Progression.* Frontiers in Cell and Developmental Biology, (2018). pp. 66–66. https://doi.org/10.3389/fcell.2018.00066.

[131] Gilkes, DM, Bajpai, S, Wong, CC, Chaturvedi, P, Hubbi, ME, Wirtz, D, and Semenza, GL. *Procollagen lysyl hydroxylase 2 is essential for hypoxia-induced breast cancer metastasis.* Molecular Cancer Research, (2013). pp. 456–466. https://doi.org/10.1158/1541-7786.MCR-12-0629.

[132] Xu, F, Zhang, J, Hu, G, Liu, L, and Liang, W. *Hypoxia and TGF-β1 induced PLOD2 expression improve the migration and invasion of cervical cancer cells by promoting epithelial-to-mesenchymal transition (EMT) and focal adhesion formation.* Cancer Cell International, (2017). pp. 54–54. https://doi.org/10.1186/s12935-017-0420-z.

[133] Chen, Z, Fang, Z, and Ma, J. *Regulatory mechanisms and clinical significance of vimentin in breast cancer.* Biomedicine & Pharmacotherapy, (2021). pp. 111068–111068. https://doi.org/10.1016/j.biopha.2020.111068.

[134] Berr, AL, Wiese, K, Dos Santos, G, Koch, CM, Anekalla, KR, Kidd, M, Davis, JM, Cheng, Y, Hu, YS, and Ridge, KM. *Vimentin is required for tumor progression and metastasis in a mouse model of non-small cell lung cancer.* Oncogene, (2023). pp. 2074–2087. https://doi.org/10.1038/s41388-023-02703-9.

[135] McInroy, L and Määttä, A. *Down-regulation of vimentin expression inhibits carcinoma cell migration and adhesion.* Biochemical and Biophysical Research Communications, (2007). pp. 109–114. https://doi.org/10.1016/j.bbrc.2007.06.036.

[136] Qiu, Z, Wang, Y, Zhang, Z, Qin, R, Peng, Y, Tang, W, Xi, Y, Tian, G, and Zhang, Y. *Roles of intercellular cell adhesion molecule-1 (ICAM-1) in colorectal cancer: expression, functions, prognosis, tumorigenesis, polymorphisms and therapeutic implications.* Frontiers in Oncology, (2022). pp. 1052672–1052672. https://doi.org/10.3389/fonc.2022.1052672.

[137] Guo, P, Huang, J, Wang, L, Jia, D, Yang, J, Dillon, DA, Zurakowski, D, Mao, H, Moses, MA, and Auguste, DT. *ICAM-1 as a molecular target for triple negative breast cancer.* Proceedings of the National Academy of Sciences, (2014). pp. 14710–14715. https://doi.org/10.1073/pnas.1408556111.

[138] Chen, M, Wu, C, Fu, Z, and Liu, S. *ICAM1 promotes bone metastasis via integrin-mediated TGF-β/EMT signaling in triple-negative breast cancer.* Cancer Science, (2022). pp. 3751–3765. https://doi.org/10.1111/cas.15532.

[139] Rosette, C, Roth, RB, Oeth, P, Braun, A, Kammerer, S, Ekblom, J, and Denissenko, MF. *Role of ICAM1 in invasion of human breast cancer cells.* Carcinogenesis, (2005). pp. 943–950. https://doi.org/10.1093/carcin/bgi070.

[140] Hu, X, Liu, Y, Bing, Z, Ye, Q, and Li, C. *High Moesin Expression Is a Predictor of Poor Prognosis of Breast Cancer: Evidence From a Systematic Review With Meta-Analysis.* Frontiers in Oncology, (2021). pp. 650488–650488. https://doi.org/10.3389/fonc.2021.650488.

[141] Alam, F, Mezhal, F, EL Hasasna, H, Nair, VA, Aravind, S, Saber Ayad, M, El-Serafi, A, and Abdel-Rahman, WM. *The role of p53-microRNA 200-Moesin axis in invasion and drug resistance of breast cancer cells.* Tumor Biology, (2017). pp. 1010428317714634–1010428317714634. https://doi.org/10.1177/1010428317714634.

[142] Zhao, K, Yi, Y, Ma, Z, and Zhang, W. *INHBA is a Prognostic Biomarker and Correlated With Immune Cell Infiltration in Cervical Cancer.* Frontiers in Genetics, (2022). pp. 705512–705512. https://doi.org/10.3389/fgene.2021.705512.

[143] Yu, Z, Cheng, L, Liu, X, Zhang, L, and Cao, H. *Increased Expression of INHBA Is Correlated With Poor Prognosis and High Immune Infiltrating Level in Breast Cancer.* Frontiers in Bioinformatics, (2022). pp. 729902–729902. https://doi.org/10.3389/fbinf.2022.729902.

[144] Yu, Y, Wang, W, Lu, W, Chen, W, and Shang, A. *Inhibin β-A (INHBA) induces epithelial-mesenchymal transition and accelerates the motility of breast cancer cells by activating the TGF-β signaling pathway.* Bioengineered, (2021). pp. 4681–4696. https://doi.org/10.1080/21655979.2021.1957754.

[145] Yang, MH, Wu, MZ, Chiou, SH, Chen, PM, Chang, SY, Liu, CJ, Teng, SC, and Wu, KJ. *Direct regulation of TWIST by HIF-1α promotes metastasis.* Nature Cell Biology, (2008). pp. 295–305. https://doi.org/10.1038/ncb1691.

[146] Bakiri, L, Macho-Maschler, S, Custic, I, Niemiec, J, Guío-Carrión, A, Hasenfuss, SC, Eger, A, Müller, M, Beug, H, and Wagner, EF. *Fra-1/AP-1 induces EMT in*

*mammary epithelial cells by modulating Zeb1/2 and TGF*β *expression.* Cell Death and Differentiation, (2015). pp. 336–350. https://doi.org/10.1038/cdd.2014.157.

[147] Chen, D, Sun, Y, Yuan, Y, Han, Z, Zhang, P, Zhang, J, You, MJ, Teruya-Feldstein, J, Wang, M, Gupta, S, Hung, MC, Liang, H, and Ma, L. *miR-100 induces epithelial-mesenchymal transition but suppresses tumorigenesis, migration and invasion.* PLoS Genetics, (2014). pp. e1004177. https://doi.org/10.1371/journal.pgen.1004177.

[148] Liang, YK, Lin, HY, Dou, XW, Chen, M, Wei, XL, Zhang, YQ, Wu, Y, Chen, CF, Bai, JW, Xiao, YS, et al. *MiR-221/222 promote epithelial-mesenchymal transition by targeting Notch3 in breast cancer cell lines.* NPJ breast cancer, (2018). pp. 20. https://doi.org/10.1038/s41523-018-0073-7.

[149] Dwivedi, SKD, Mustafi, SB, Mangala, LS, Jiang, D, Pradeep, S, Rodriguez-Aguayo, C, Ling, H, Ivan, C, Mukherjee, P, Calin, GA, Lopez-Berestein, G, Sood, AK, and Bhattacharya, R. *Therapeutic evaluation of microRNA-15a and microRNA-16 in ovarian cancer.* Oncotarget, (2016). pp. 15093–15104. https://doi.org/10.18632/oncotarget.7618.

[150] El Sayed, SR, Cristante, J, Guyon, L, Denis, J, Chabre, O, and Cherradi, N. *MicroRNA Therapeutics in Cancer: Current Advances and Challenges.* Cancers, (2021). https://doi.org/10.3390/cancers13112680.

[151] Zhang, HY, Li, CH, Wang, XC, Luo, YQ, Cao, XD, and Chen, JJ. *MiR-199 inhibits EMT and invasion of hepatoma cells through inhibition of Snail expression.* European Review for Medical & Pharmacological Sciences, (2019). https://doi.org/10.26355/eurrev_201909_18998.

[152] Ning An, B and Bo Zheng, B. *MiR-203a-3p Inhibits Pancreatic Cancer Cell Proliferation, EMT, and Apoptosis by Regulating SLUG.* Technology in Cancer Research & Treatment, (2020). pp. 1533033819898729–1533033819898729. https://doi.org/10.1177/1533033819898729.

[153] Wellner, U, Schubert, J, Burk, UC, Schmalhofer, O, Zhu, F, Sonntag, A, Waldvogel, B, Vannier, C, Darling, D, Hausen, A zur, Brunton, VG, Morton, J, Sansom, O, Schüler, J, Stemmler, MP, Herzberger, C, Hopt, U, Keck, T, Brabletz, S, and Brabletz, T. *The EMT-activator ZEB1 promotes tumorigenicity by repressing stemness-inhibiting microRNAs.* Nature Cell Biology, (2009). pp. 1487–1495. https://doi.org/10.1038/ncb1998.

[154] Fan, X, Chen, W, Fu, Z, Zeng, L, Yin, Y, and Yuan, H. *MicroRNAs, a subpopulation of regulators, are involved in breast cancer progression through regulating breast cancer stem cells.* Oncology Letters, (2017). pp. 5069–5076. https://doi.org/10.3892/ol.2017.6867.

[155] Pastushenko, I et al. *Identification of the tumour transition states occurring during EMT.* Nature, (2018). pp. 463–468. https://doi.org/10.1038/s41586-018-0040-3.

[156] Brown, MS, Abdollahi, B, Wilkins, OM, Lu, H, Chakraborty, P, Ognjenovic, NB, Muller, KE, Jolly, MK, Christensen, BC, Hassanpour, S, and Pattabiraman, DR. *Phenotypic heterogeneity driven by plasticity of the intermediate EMT state governs disease progression and metastasis in breast cancer.* Science Advances, (2022). pp. eabj8002–eabj8002. https://doi.org/10.1126/sciadv.abj8002.

[157] Furuta, C, Miyamoto, T, Takagi, T, Noguchi, Y, Kaneko, J, Itoh, S, Watanabe, T, and Itoh, F. *Transforming growth factor-β signaling enhancement by long-term exposure to hypoxia in a tumor microenvironment composed of Lewis lung carcinoma cells.* Cancer Science, (2015). pp. 1524–1533. https://doi.org/10.1111/cas.12773.

[158] McMahon, S, Charbonneau, M, Grandmont, S, Richard, DE, and Dubois, CM. *Transforming Growth Factor β1 Induces Hypoxia-inducible Factor-1 Stabilization through Selective Inhibition of PHD2 Expression.* The Journal of Biological Chemistry, (2006). pp. 24171–24181. https://doi.org/10.1074/jbc.M604507200.

[159] Han, WQ, Zhu, Q, Hu, J, Li, PL, Zhang, F, and Li, N. *Hypoxia-inducible factor prolyl-hydroxylase-2 mediates transforming growth factor beta 1-induced epithelial–mesenchymal transition in renal tubular cells.* Biochimica et Biophysica Acta. Molecular Cell Research, (2013). pp. 1454–1462. https://doi.org/10.1016/j.bbamcr.2013.02.029.

[160] Malagoli Tagliazucchi, G, Wiecek, AJ, Withnell, E, and Secrier, M. *Genomic and microenvironmental heterogeneity shaping epithelial-to-mesenchymal trajectories in cancer.* Nature Communications, (2023). pp. 789–789. https://doi.org/10.1038/s41467-023-36439-7.

[161] Chang, CJ, Chao, CH, Xia, W, Yang, JY, Xiong, Y, Li, CW, Yu, WH, Rehman, SK, Hsu, JL, Lee, HH, et al. *p53 regulates epithelial-mesenchymal transition and stem*

cell properties through modulating miRNAs. Nature Cell Biology, (2011). pp. 317–323. https://doi.org/10.1038/ncb2173.

[162] De Craene, B and Berx, G. *Regulatory networks defining EMT during cancer initiation and progression.* Nature Reviews Cancer, (2013). pp. 97–110. https://doi.org/10.1038/nrc3447.

[163] Jolly, MK, Boareto, M, Huang, B, Jia, D, Lu, M, Ben-Jacob, E, Onuchic, JN, and Levine, H. *Implications of the Hybrid Epithelial/Mesenchymal Phenotype in Metastasis.* Frontiers in Oncology, (2015). pp. 155–155. https://doi.org/10.3389/fonc.2015.00155.

[164] Ribeiro, AS and Paredes, J. *P-Cadherin Linking Breast Cancer Stem Cells and Invasion: A Promising Marker to Identify an "Intermediate/Metastable" EMT State.* Frontiers in Oncology, (2015). pp. 371–371. https://doi.org/10.3389/fonc.2014.00371.

[165] Plygawko, AT, Kan, S, and Campbell, K. *Epithelial-mesenchymal plasticity: emerging parallels between tissue morphogenesis and cancer metastasis.* Philosophical Transactions of the Royal Society of London. Series B. Biological Sciences, (2020). pp. 20200087–20200087. https://doi.org/10.1098/rstb.2020.0087.

[166] Hecht, I, Bar-El, Y, Balmer, F, Natan, S, Tsarfaty, I, Schweitzer, F, and Ben-Jacob, E. *Tumor invasion optimization by mesenchymal-amoeboid heterogeneity.* Scientific Reports, (2015). pp. 10622–10622. https://doi.org/10.1038/srep10622.

[167] Feldker, N, Ferrazzi, F, Schuhwerk, H, Widholz, SA, Guenther, K, Frisch, I, Jakob, K, Kleemann, J, Riegel, D, Bönisch, U, Lukassen, S, Eccles, RL, Schmidl, C, Stemmler, MP, Brabletz, T, and Brabletz, S. *Genome-wide cooperation of EMT transcription factor ZEB1 with YAP and AP-1 in breast cancer.* EMBO Journal, (2020). pp. e103209–n/a. https://doi.org/10.15252/embj.2019103209.

[168] Fischer, KR, Durrans, A, Lee, S, Sheng, J, Li, F, Wong, ST, Choi, H, El Rayes, T, Ryu, S, Troeger, J, et al. *Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance.* Nature, (2015). pp. 472–476.

[169] Zheng, X, Carstens, JL, Kim, J, Scheible, M, Kaye, J, Sugimoto, H, Wu, CC, LeBleu, VS, and Kalluri, R. *Epithelial-to-mesenchymal transition is dispensable for metastasis but*

induces chemoresistance in pancreatic cancer. Nature, (2015). pp. 525–530. https://doi.org/10.1038/nature16064.

[170] Wang, JY and Doudna, JA. *CRISPR technology: A decade of genome editing is only the beginning.* Science, (2023). pp. eadd8643–eadd8643. https://doi.org/10.1126/science.add8643.

[171] Joung, J, Konermann, S, Gootenberg, JS, Abudayyeh, OO, Platt, RJ, Brigham, MD, Sanjana, NE, and Zhang, F. *Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening.* Nature Protocols, (2017). pp. 828–863. https://doi.org/10.1038/nprot.2017.016.

[172] Doench, JG. *Am I ready for CRISPR? A user's guide to genetic screens.* Nature Reviews Genetics, (2018). pp. 67–80. https://doi.org/10.1038/nrg.2017.97.

[173] Hart, T et al. *Evaluation and Design of Genome-Wide CRISPR/SpCas9 Knockout Screens.* G3-GENES GENOMES GENETICS, (2017). pp. 2719–2727. https://doi.org/10.1534/g3.117.041277.

[174] Nakamura, A, Nambu, T, Ebara, S, Hasegawa, Y, Toyoshima, K, Tsuchiya, Y, Tomita, D, Fujimoto, J, Kurasawa, O, Takahara, C, Ando, A, Nishigaki, R, Satomi, Y, Hata, A, and Hara, T. *Inhibition of GCN2 sensitizes ASNS-low cancer cells to asparaginase by disrupting the amino acid response.* Proceedings of the National Academy of Sciences of The United States of America, (2018). pp. E7776–E7785. https://doi.org/10.1073/pnas.1805523115.

[175] Li, W, Koster, J, Xu, H, Chen, CH, Xiao, TF, Liu, JS, Brown, M, and Liu, XS. *Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR.* Genome Biology, (2015). https://doi.org/10.1186/s13059-015-0843-6.

[176] Li, W. *MAGeCK Wiki.* (2019). https://sourceforge.net/p/mageck/wiki/Home/.

[177] Cross, BCS, Lawo, S, Archer, CR, Hunt, JR, Yarker, JL, Riccombeni, A, Little, AS, McCarthy, NJ, and Moore, JD. *Increasing the performance of pooled CRISPR-Cas9 drop-out screening.* Scientific Reports, (2016). https://doi.org/10.1038/srep31782.

[178] Thomas, LW, Esposito, C, Morgan, RE, Price, S, Young, J, Williams, SP, Maddalena, LA, McDermott, U, and Ashcroft, M. *Genome-wide CRISPR/Cas9 deletion screen defines mitochondrial gene essentiality and identifies routes for tumour cell viability in*

hypoxia. Communications Biology, (2021). pp. 615–615. https://doi.org/10.1038/s42003-021-02098-x.

[179] Thorén, MM, Vaapil, M, Staaf, J, Planck, M, Johansson, ME, Mohlin, S, and Påhlman, S. *Myc-induced glutaminolysis bypasses HIF-driven glycolysis in hypoxic small cell lung carcinoma cells.* Oncotarget, (2017). pp. 48983. https://doi.org/10.18632/oncotarget.16904.

[180] Blais, JD, Filipenko, V, Bi, M, Harding, HP, Ron, D, Koumenis, C, Wouters, BG, and Bell, JC. *Activating Transcription Factor 4 Is Translationally Regulated by Hypoxic Stress.* Molecular and Cellular Biology, (2004). pp. 7469–7482. https://doi.org/10.1128/MCB.24.17.7469-7482.2004.

[181] Koumenis, C, Naczki, C, Koritzinsky, M, Rastani, S, Diehl, A, Sonenberg, N, Koromilas, A, and Wouters, BG. *Regulation of Protein Synthesis by Hypoxia via Activation of the Endoplasmic Reticulum Kinase PERK and Phosphorylation of the Translation Initiation Factor eIF2α.* Molecular and Cellular Biology, (2002). pp. 7405–7416. https://doi.org/10.1128/MCB.22.21.7405-7416.2002.

[182] Rouschop, KM, Dubois, LJ, Keulers, TG, Beucken, T van den, Lambin, P, Bussink, J, Kogel, AJ van der, Koritzinsky, M, and Wouters, BG. *PERK/eIF2α signaling protects therapy resistant hypoxic cells through induction of glutathione synthesis and protection against ROS.* Proceedings of the National Academy of Sciences, (2013). pp. 4622–4627. https://doi.org/10.1073/pnas.1210633110.

[183] Furnish, M, Boulton, DP, Genther, V, Grofova, D, Ellinwood, ML, Romero, L, Lucia, MS, Cramer, SD, and Caino, MC. *MIRO2 Regulates Prostate Cancer Cell Growth via GCN1-Dependent Stress Signaling.* Molecular Cancer Research, (2022). pp. 607–621. https://doi.org/10.1158/1541-7786.MCR-21-0374.

[184] Ge, L, Chen, W, Cao, W, Liu, G, Zhang, Q, Zhuang, J, Zhang, M, Yang, J, Guo, S, Zhao, X, and Guo, H. *GCN2 is a potential prognostic biomarker for human papillary renal cell carcinoma.* Cancer Biomarkers, (2018). pp. 395–403. https://doi.org/10.3233/CBM-170922.

[185] Gold, LT, Bray, SE, Kernohan, NM, Henderson, N, Nowicki, M, and Masson, GR. *The amino-acid stress sensing eIF2α kinase GCN2 is a survival biomarker for malignant*

mesothelioma. BJC Reports, (2023). pp. 4. https://doi.org/10.1038/s44276-023-00004-y.

[186] Wang, Y, Ning, Y, Alam, GN, Jankowski, BM, Dong, Z, Nör, JE, and Polverini, PJ. *Amino acid deprivation promotes tumor angiogenesis through the GCN2/ATF4 pathway.* Neoplasia, (2013). pp. 989–997. https://doi.org/10.1593/neo.13262.

[187] Zhang, P, McGrath, BC, Reinert, J, Olsen, DS, Lei, L, Gill, S, Wek, SA, Vattem, KM, Wek, RC, Kimball, SR, et al. *The GCN2 eIF2α kinase is required for adaptation to amino acid deprivation in mice.* Molecular and Cellular Biology, (2002). pp. 6681–6688. https://doi.org/10.1128/MCB.22.19.6681-6688.2002.

[188] Masuoka, HC and Townes, TM. *Targeted disruption of the activating transcription factor 4 gene results in severe fetal anemia in mice.* Blood, (2002). pp. 736–745. https://doi.org/10.1182/blood.v99.3.736.

[189] Saavedra-Garcia, P et al. *Systems level profiling of chemotherapy-induced stress resolution in cancer cells reveals druggable trade-offs.* Proceedings of the National Academy of Sciences of The United States of America, (2021). https://doi.org/10.1073/pnas.2018229118.

[190] Zaarour, RF, Azakir, B, Hajam, EY, Nawafleh, H, Zeinelabdin, NA, Engelsen, AST, Thiery, J, Jamora, C, and Chouaib, S. *Role of Hypoxia-Mediated Autophagy in Tumor Cell Death and Survival.* Cancers, (2021). pp. 533. https://doi.org/10.3390/cancers13030533.

[191] D'Aiuto, N, Hochmann, J, Millán, M, Di Paolo, A, Bologna-Molina, R, Sotelo Silveira, J, and Arocena, M. *Hypoxia, acidification and oxidative stress in cells cultured at large distances from an oxygen source.* Scientific Reports, (2022). pp. 21699–21699. https://doi.org/10.1038/s41598-022-26205-y.

[192] Bao, MHR, Yang, C, Tse, APW, Wei, L, Lee, D, Zhang, MS, Goh, CC, Chiu, DKC, Yuen, VWH, Law, CT, Chin, WC, Chui, NNQ, Wong, BPY, Chan, CYK, Ng, IOL, Chung, CYS, Wong, CM, and Wong, CCL. *Genome-wide CRISPR-Cas9 knockout library screening identified PTPMT1 in cardiolipin synthesis is crucial to survival in hypoxia in liver cancer.* Cell Reports, (2021). pp. 108676–108676. https://doi.org/10.1016/j.celrep.2020.108676.

[193] Bartel, DP. *Metazoan MicroRNAs.* Cell, (2018). pp. 20–51. https://doi.org/10.1016/j.cell.2018.03.006.

[194] Daugaard, I, Sanders, K, Idica, A, Vittayarukskul, K, Hamdorf, M, Krog, J, Chow, R, Jury, D, Hansen, L, Hager, H, et al. *miR-151a induces partial EMT by regulating E-cadherin in NSCLC cells.* Oncogenesis, (2017). pp. e366–e366. https://doi.org/10.1038/oncsis.2017.66.

[195] Zaravinos, A. *The Regulatory Role of MicroRNAs in EMT and Cancer.* Journal of Oncology, (2015). https://doi.org/10.1155/2015/865816.

[196] Peng, Y and Croce, CM. *The role of MicroRNAs in human cancer.* Signal Transduction and Targeted Therapy, (2016). https://doi.org/10.1038/sigtrans.2015.4.

[197] Qattan, A. *Novel miRNA Targets and Therapies in the Triple-Negative Breast Cancer Microenvironment: An Emerging Hope for a Challenging Disease.* International Journal of Molecular Sciences, (2020). https://doi.org/10.3390/ijms21238905.

[198] Zografos, E, Zagouri, F, Kalapanida, D, Zakopoulou, R, Kyriazoglou, A, Apostolidou, K, Gazouli, M, and Dimopoulos, MA. *Prognostic role of microRNAs in breast cancer: A systematic review.* Oncotarget, (2019). pp. 7156. https://doi.org/10.18632/oncotarget.27327.

[199] Si, ML, Zhu, S, Wu, H, Lu, Z, Wu, F, and Mo, YY. *miR-21-mediated tumor growth.* Oncogene, (2007). pp. 2799–2803. https://doi.org/10.1038/sj.onc.1210083.

[200] Hannafon, BN, Cai, A, Calloway, CL, Xu, YF, Zhang, R, Fung, KM, and Ding, WQ. *miR-23b and miR-27b are oncogenic microRNAs in breast cancer: evidence from a CRISPR/Cas9 deletion study.* BMC Cancer, (2019). https://doi.org/10.1186/s12885-019-5839-2.

[201] Yan, LX, Wu, QN, Zhang, Y, Li, YY, Liao, DZ, Hou, JH, Fu, J, Zeng, MS, Yun, JP, Wu, QL, Zeng, YX, and Shao, JY. *Knockdown of miR-21 in human breast cancer cell lines inhibits proliferation, in vitro migration and in vivo tumor growth.* Breast Cancer Research, (2011). https://doi.org/10.1186/bcr2803.

[202] Costales, MG, Haga, CL, Velagapudi, SP, Childs-Disney, JL, Phinney, DG, and Disney, MD. *Small Molecule Inhibition of microRNA-210 Reprograms an Oncogenic Hypoxic Circuit.* Journal of the American Chemical Society, (2017). pp. 3446–3455. https://doi.org/10.1021/jacs.6b11273.

[203] Favaro, E, Ramachandran, A, McCormick, R, Gee, H, Blancher, C, Crosby, M, Devlin, C, Blick, C, Buffa, F, Li, JL, et al. *MicroRNA-210 regulates mitochondrial free radical response to hypoxia and krebs cycle in cancer cells by targeting iron sulfur cluster protein ISCU.* PloS One, (2010). pp. e10345. https://doi.org/10.1371/journal.pone.0010345.

[204] Yang, W, Wei, J, Sun, T, and Liu, F. *Effects of knockdown of miR-210 in combination with ionizing radiation on human hepatoma xenograft in nude mice.* Radiation Oncology, (2013). pp. 1–10. https://doi.org/10.1186/1748-717X-8-102.

[205] Babar, IA, Czochor, J, Steinmetz, A, Weidhaas, JB, Glazer, PM, and Slack, FJ. *Inhibition of hypoxia-induced miR-155 radiosensitizes hypoxic lung cancer cells.* Cancer Biology & Therapy, (2011). pp. 908–14. https://doi.org/10.4161/cbt.12.10.17681.

[206] Santos, JC, Lima, NdS, Sarian, LO, Matheu, A, Ribeiro, ML, and Derchain, SFM. *Exosome-mediated breast cancer chemoresistance via miR-155 transfer.* Scientific Reports, (2018). pp. 829. https://doi.org/10.1038/s41598-018-19339-5.

[207] Chang, H, Yi, B, Ma, RX, Zhang, XG, Zhao, HY, and Xi, YG. *CRISPR/cas9, a novel genomic tool to knock down microRNA in vitro and in vivo.* Scientific Reports, (2016). https://doi.org/10.1038/srep22312.

[208] Sanjana, NE, Shalem, O, and Zhang, F. *Improved vectors and genome-wide libraries for CRISPR screening.* Nature Methods, (2014). pp. 783–784. https://doi.org/10.1038/nmeth.3047.

[209] Yin, HR, Xiong, GF, Guo, SJ, Xu, CC, Xu, R, Guo, PX, and Shu, D. *Delivery of Anti-miRNA for Triple-Negative Breast Cancer Therapy Using RNA Nanoparticles Targeting Stem Cell Marker CD133.* Molecular Therapy, (2019). pp. 1252–1261. https://doi.org/10.1016/j.ymthe.2019.04.018.

[210] Chen, YC, Gao, DY, and Huang, L. *In vivo delivery of miRNAs for cancer therapy: Challenges and strategies.* Advanced Drug Delivery Reviews, (2015). pp. 128–141. https://doi.org/10.1016/j.addr.2014.05.009.

[211] Chin, AR, Fong, MY, Somlo, G, Wu, J, Swiderski, P, Wu, XW, and Wang, SE. *Cross-kingdom inhibition of breast cancer growth by plant miR159.* Cell Research, (2016). pp. 217–228. https://doi.org/10.1038/cr.2016.13.

[212] Kurata, JS and Lin, RJ. *Addgene Lin Human miRNA CRISPR Knockout Library.* (2018). https://www.addgene.org/pooled-library/lin-human-crispr-mirna/.

[213] Joung, J, Konermann, S, Gootenberg, JS, Abudayyeh, OO, Platt, RJ, Brigham, MD, Sanjana, NE, and Zhang, F. *Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening.* Nature Protocols, (2017). pp. 828–863. https://doi.org/10.1038/nprot.2017.016.

[214] Hines, WC, Yaswen, P, and Bissell, MJ. *Modelling breast cancer requires identification and correction of a critical cell lineage-dependent transduction bias.* Nature Communications, (2015). https://doi.org/10.1038/ncomms7927.

[215] Davis, E. *Lentivirus: That's My MOI, and I'm Sticking To It.* (2014). https://www.genecopoeia.com/resource/lentivirus-thats-my-moi-and-im-sticking-to-it-lentivirus-moi-cell-lines/.

[216] Bio, T. *Everything you need to know about CRISPR library screening.* (2022). https://www.takarabio.com/learning-centers/gene-function/gene-editing/genome-wide-screening/crispr-library-screening.

[217] Qiagen. *Qiagen Genomic DNA Handbook.* Generic. (2015).

[218] Shalem, O, Sanjana, NE, Hartenian, E, Shi, X, Scott, DA, Mikkelsen, TS, Heckl, D, Ebert, BL, Root, DE, Doench, JG, and Zhang, F. *Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells.* Science, (2014). pp. 84–87. https://doi.org/10.1126/science.1247005.

[219] Li, W, Xu, H, Xiao, TF, Cong, L, Love, MI, Zhang, F, Irizarry, RA, Liu, JS, Brown, M, and Liu, XS. *MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens.* Genome Biology, (2014). https://doi.org/10.1186/s13059-014-0554-4.

[220] Chen, CH, Xiao, T, Xu, H, Jiang, P, Meyer, CA, Li, W, Brown, M, and Liu, XS. *Improved design and analysis of CRISPR knockout screens.* Bioinformatics, (2018). pp. 4095–4101. https://doi.org/10.1093/bioinformatics/bty450.

[221] Benjamini, Y and Hochberg, Y. *Controlling the false discovery rate - a practical and powerful approach to multiple testing.* Journal of the Royal Statistical Society Series B-Statistical Methodology, (1995). pp. 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

[222] Bodapati, S, Daley, TP, Lin, XQ, Zou, J, and Qi, LS. *A benchmark of algorithms for the analysis of pooled CRISPR screens.* Genome Biology, (2020). https://doi.org/10.1186/s13059-020-01972-x.

[223] Pajak, M and Simpson, I. *miRNAtap: microRNA Targets - Aggregated Predictions.* (2022). https://bioconductor.org/packages/devel/bioc/manuals/miRNAtap/man/miRNAtap.pdf.

[224] Maragkakis, M, Vergoulis, T, Alexiou, P, Reczko, M, Plomaritou, K, Gousis, M, Kourtis, K, Koziris, N, Dalamagas, T, and Hatzigeorgiou, AG. *DIANA-microT Web server upgrade supports Fly and Worm miRNA target prediction and bibliographic miRNA to disease association.* Nucleic Acids Research, (2011). pp. W145–W148. https://doi.org/10.1093/nar/gkr294.

[225] Enright, AJ, John, B, Gaul, U, Tuschl, T, Sander, C, and Marks, DS. *MicroRNA targets in Drosophila.* Genome Biology, (2004). https://doi.org/10.1186/gb-2003-5-1-r1.

[226] Lall, S, Grun, D, Krek, A, Chen, K, Wang, YL, Dewey, CN, Sood, P, Colombo, T, Bray, N, MacMenamin, P, Kao, HL, Gunsalus, KC, Pachter, L, Piano, F, and Rajewsky, N. *A genome-wide map of conserved microRNA targets in C. elegans.* Current Biology, (2006). pp. 460–471. https://doi.org/10.1016/j.cub.2006.01.050.

[227] Friedman, RC, Farh, KKH, Burge, CB, and Bartel, DP. *Most mammalian mRNAs are conserved targets of microRNAs.* Genome Research, (2009). pp. 92–105. https://doi.org/10.1101/gr.082701.108.

[228] Wong, N and Wang, XW. *miRDB: an online resource for microRNA target prediction and functional annotations.* Nucleic Acids Research, (2015). pp. D146–D152. https://doi.org/10.1093/nar/gku1104.

[229] Pelt-Verkuil, E van, Belkum, A van, and Hays, JP. *Principles and Technical Aspects of PCR Amplification.* Springer Science & Business Media., (2008).

[230] Boo, L, Ho, WY, Alit, NM, Yeap, SK, Ky, H, Chan, KG, Yin, WF, Satharasinghe, DA, Liew, WC, Tan, SW, Cheong, SK, and Ong, HK. *Phenotypic and microRNA transcriptomic profiling of the MDA-MB-231 spheroid-enriched CSCs with comparison of MCF-7 microRNA profiling dataset.* PeerJ, (2017). https://doi.org/10.7717/peerj.3551.

[231] Institute, B. *DepMap: The Cancer Dependency Map Project at Broad Institute.* (2021). https://depmap.org/portal/.

[232] Zheng, X, Chen, L, Li, X, Zhang, Y, Xu, S, and Huang, X. *Prediction of miRNA targets by learning from interaction sequences.* Plos One, (2020). pp. e0232578. https://doi.org/10.1371/journal.pone.0232578.

[233] Abida, WM, Nikolaev, A, Zhao, WH, Zhang, WZ, and Gu, W. *FBXO11 promotes the neddylation of p53 and inhibits its transcriptional activity.* Journal of Biological Chemistry, (2007). pp. 1797–1804. https://doi.org/10.1074/jbc.M609001200.

[234] Tateossian, H, Morse, S, Simon, MM, Dean, CH, and Brown, SDM. *Interactions between the otitis media gene, Fbxo11, and p53 in the mouse embryonic lung.* Disease Models & Mechanisms, (2015). pp. 1531–1542. https://doi.org/10.1242/dmm.022426.

[235] Lee, JY, Tokumoto, M, Fujiwara, Y, Hasegawa, T, Seko, Y, Shimada, A, and Satoh, M. *Accumulation of p53 via down-regulation of UBE2D family genes is a critical pathway for cadmium-induced renal toxicity.* Scientific Reports, (2016). https://doi.org/10.1038/srep21968.

[236] Hui, L, Zheng, Y, Yan, Y, Bargonetti, J, and Foster, DA. *Mutant p53 in MDA-MB-231 breast cancer cells is stabilized by elevated phospholipase D activity and contributes to survival signals generated by phospholipase D.* Oncogene, (2006). pp. 7305–7310. https://doi.org/10.1038/sj.onc.1209735.

[237] Lisek, K, Campaner, E, Ciani, Y, Walerych, D, and Del Sal, G. *Mutant p53 tunes the NRF2-dependent antioxidant response to support survival of cancer cells.* Oncotarget, (2018). pp. 20508–20523. https://doi.org/10.18632/oncotarget.24974.

[238] Dufies, M. *SH3BGRL2, a new downregulated tumor suppressor in clear cell renal cell carcinomas.* EBioMedicine, (2020). https://doi.org/10.1016/j.ebiom.2020.102641.

[239] Li, DD, Deng, L, Hu, SY, Zhang, FL, and Li, DQ. *SH3BGRL2 exerts a dual function in breast cancer growth and metastasis and is regulated by TGF-beta 1.* American Journal of Cancer Research, (2020). pp. 1238–+.

[240] Duckett, CS. *IAP proteins: sticking it to Smac.* Biochemical Journal, (2005). pp. e1–2. https://doi.org/10.1042/bj20041800.

[241] Maelandsmo, GM, Florenes, VA, Mellingsaeter, T, Hovig, E, Kerbel, RS, and Fodstad, O. *Differential expression patterns of S100A2, S100A4 and S100A6 during progression of human malignant melanoma.* International Journal of Cancer, (1997). pp. 464–469. https://doi.org/10.1002/(sici)1097-0215(19970822)74:4<464::aid-ijc19>3.3.co;2-g.

[242] Gupta, S, Hussain, T, MacLennan, GT, Fu, PF, Patel, J, and Mukhtar, H. *Differential expression of S100A2 and S100A4 during progression of human prostate adenocareinoma.* Journal of Clinical Oncology, (2003). pp. 106–112. https://doi.org/10.1200/jco.2003.03.024.

[243] Liu, D, Rudland, PS, Sibson, DR, Platt-Higgins, A, and Barraclough, R. *Expression of calcium binding protein S100A2 in breast lesions.* British Journal of Cancer, (2000). pp. 1473–1479. https://doi.org/10.1054/bjoc.2000.1488.

[244] Buckley, NE, D'Costa, Z, Kaminska, M, and Mullan, PB. *S100A2 is a BRCA1/p63 coregulated tumour suppressor gene with roles in the regulation of mutant p53 stability.* Cell Death & Disease, (2014). https://doi.org/10.1038/cddis.2014.31.

[245] Yang, JY, Wahdan-Alaswad, R, and Danielpour, D. *Critical Role of Smad2 in Tumor Suppression and Transforming Growth Factor-beta-Induced Apoptosis of Prostate Epithelial Cells.* Cancer Research, (2009). pp. 2185–2190. https://doi.org/10.1158/0008-5472.can-08-3961.

[246] Petersen, M, Pardali, E, Horst, G van der, Cheung, H, Hoogen, C van den, Pluijm, G van der, and Dijke, P ten. *Smad2 and Smad3 have opposing roles in breast cancer bone metastasis by differentially affecting tumor angiogenesis.* Oncogene, (2010). pp. 1351–1361. https://doi.org/10.1038/onc.2009.426.

[247] Lawrence, EJ, Arpag, G, Arnaiz, C, and Zanic, M. *SSNA1 stabilizes dynamic microtubules and detects microtubule damage.* eLife, (2021). https://doi.org/10.7554/eLife.67282;10.7554/eLife.67282.sa0;10.7554/eLife.67282.sa1;10.7554/eLife.67282.sa2.

[248] Tang, DF, Zhao, DD, Wu, Y, Yao, RY, Zhou, L, Lu, LM, Gao, W, and Sun, YF. *The miR-3127-5p/p-STAT3 axis up-regulates PD-L1 inducing chemoresistance in non-small-cell lung cancer.* Journal of Cellular and Molecular Medicine, (2018). pp. 3847–3856. https://doi.org/10.1111/jcmm.13657.

[249] Wang, Y, Lu, Z, Wang, NN, Zhang, M, Zeng, XD, and Zhao, W. *MicroRNA-1299 is a negative regulator of STAT3 in colon cancer.* Oncology Reports, (2017). pp. 3227–3234. https://doi.org/10.3892/or.2017.5605.

[250] Pawlus, M, Wang, L, and Hu, C. *STAT3 and HIF1α cooperatively activate HIF1 target genes in MDA-MB-231 and RCC4 cells.* Oncogene, (2014). pp. 1670–1679. https://doi.org/10.1038/onc.2013.115.

[251] Gugnoni, M, Sancisi, V, Manzotti, G, Gandolfi, G, and Ciarrocchi, A. *Autophagy and epithelial–mesenchymal transition: an intricate interplay in cancer.* Cell Death & Disease, (2016). pp. e2520–e2520. https://doi.org/10.1038/cddis.2016.415.

[252] Deng, KY, Huang, LJ, Sun, XY, and Zang, YH. *The role and underlying mechanism of miR-1299 in cancer.* Future Science OA, (2021). https://doi.org/10.2144/fsoa-2021-0014.

[253] Wang, J, Zhao, J, Shi, M, Ding, Y, Sun, H, Yuan, F, and Zou, Z. *Elevated expression of miR-210 predicts poor survival of cancer patients: a systematic review and meta-analysis.* PloS One, (2014). pp. e89223. https://doi.org/10.1371/journal.pone.0089223.

[254] Wang, W, Qu, A, Liu, W, Liu, Y, Zheng, G, Du, L, Zhang, X, Yang, Y, Wang, C, and Chen, X. *Circulating miR-210 as a diagnostic and prognostic biomarker for colorectal cancer.* European Journal of Cancer Care, (2017). pp. e12448. https://doi.org/10.1111/ecc.12448.

[255] Tang, Y, Zhou, X, Ji, J, Chen, L, Cao, J, Luo, J, and Zhang, S. *High expression levels of miR-21 and miR-210 predict unfavorable survival in breast cancer: a systemic review and meta-analysis.* The International Journal of Biological Markers, (2015). pp. 347–358. https://doi.org/10.5301/jbm.5000160.

[256] Kurata, JS and Lin, RJ. *Pan-Cancer Biomarker miR-151a Regulates p21 Partially Through p53.* bioRxiv, (2023).

[257] Fort, RS, Matho, C, Oliveira-Rizzo, C, Garat, B, Sotelo-Silveira, JR, and Duhagon, MA. *An integrated view of the role of miR-130b/301b miRNA cluster in prostate cancer.* Experimental Hematology & Oncology, (2018). https://doi.org/10.1186/s40164-018-0102-0.

[258] Chang, YY, Kuo, WH, Hung, JH, Lee, CY, Lee, YH, Chang, YC, Lin, WC, Shen, CY, Huang, CS, Hsieh, FJ, Lai, LC, Tsai, MH, Chang, KJ, and Chuang, EY. *Deregulated microRNAs in triple-negative breast cancer revealed by deep sequencing.* Molecular Cancer, (2015). https://doi.org/10.1186/s12943-015-0301-9.

[259] Miao, Y, Zheng, W, Li, NN, Su, Z, Zhao, LF, Zhou, HM, and Jia, L. *MicroRNA-130b targets PTEN to mediate drug resistance and proliferation of breast cancer cells via the PI3K/Akt signaling pathway.* Scientific Reports, (2017). https://doi.org/10.1038/srep41942.

[260] Montagner, M, Enzo, E, Forcato, M, Zanconato, F, Parenti, A, Rampazzo, E, Basso, G, Leo, G, Rosato, A, Bicciato, S, Cordenonsi, M, and Piccolo, S. *SHARP1 suppresses breast cancer metastasis by promoting degradation of hypoxia-inducible factors.* Nature, (2012). pp. 380–384. https://doi.org/10.1038/nature11207.

[261] Lee, D. *miR-769-5p is associated with prostate cancer recurrence and modulates proliferation and apoptosis of cancer cells.* Experimental and Therapeutic Medicine, (2021). https://doi.org/10.3892/etm.2021.9766.

[262] Liu, WS, Wang, BY, Duan, A, Shen, K, Zhang, Q, Tang, XL, Wei, YZ, Tang, J, and Zhang, S. *Exosomal transfer of miR-769-5p promotes osteosarcoma proliferation and metastasis by targeting DUSP16.* Cancer Cell International, (2021). https://doi.org/10.1186/s12935-021-02257-4.

[263] Han, C, Song, YM, and Lian, CH. *MiR-769 Inhibits Colorectal Cancer Cell Proliferation and Invasion by Targeting HEY1.* Medical Science Monitor, (2018). pp. 9232–9239. https://doi.org/10.12659/msm.911663.

[264] Yang, Z, He, J, Gao, P, Niu, Y, Zhang, J, Wang, L, Liu, MY, Wei, XM, Liu, CL, Zhang, C, Wang, W, Du, JY, Li, HM, Hu, WN, and Sun, GG. *miR-769-5p suppressed cell proliferation, migration and invasion by targeting TGFBR1 in non-small cell lung carcinoma.* Oncotarget, (2017). pp. 113558–113570. https://doi.org/10.18632/oncotarget.23060.

[265] Hammond, SM. *An overview of microRNAs.* Advanced Drug Delivery Reviews, (2015). pp. 3–14. https://doi.org/10.1016/j.addr.2015.05.001.

[266] Bak, RO and Mikkelsen, JG. *miRNA sponges: soaking up miRNAs for regulation of gene expression.* Wiley Interdisciplinary Reviews - RNA, (2014). pp. 317–333. https://doi.org/10.1002/wrna.1213.

[267] Oh, SY, Ju, Y, and Park, H. *A highly effective and long-lasting inhibition of miRNAs with PNA-based antisense oligonucleotides.* Molecules and Cells, (2009). pp. 341–345. https://doi.org/10.1007/s10059-009-0134-8.

[268] Schmidt, MF. *Drug target miRNAs: chances and challenges.* Trends in Biotechnology, (2014). pp. 578–585. https://doi.org/10.1016/j.tibtech.2014.09.002.

[269] Suresh, BM, Li, WC, Zhang, PY, Wang, KW, Yildirim, I, Parker, CG, and Disney, MD. *A general fragment-based approach to identify and optimize bioactive ligands targeting RNA.* Proceedings of the National Academy of Sciences of the United States of America, (2020). pp. 33197–33203. https://doi.org/10.1073/pnas.2012217117.

[270] Costales, MG, Aikawa, H, Li, Y, Childs-Disney, JL, Abegg, D, Hoch, DG, Velagapudi, SP, Nakai, Y, Khan, T, Wang, KW, Yildirim, I, Adibekian, A, Wang, ET, and Disney, MD. *Small-molecule targeted recruitment of a nuclease to cleave an oncogenic RNA in a mouse model of metastatic cancer.* Proceedings of the National Academy of Sciences of the United States of America, (2020). pp. 2406–2411. https://doi.org/10.1073/pnas.1914286117.

[271] Jewer, M, Lee, L, Leibovitch, M, Zhang, G, Liu, J, Findlay, SD, Vincent, KM, Tandoc, K, Dieters-Castator, D, Quail, DF, et al. *Translational control of breast cancer plasticity.* Nature Communications, (2020). pp. 2498. https://doi.org/10.1038/s41467-020-16352-z.

[272] Jackson, JJ, Shibuya, GM, Ravishankar, B, Adusumilli, L, Bradford, D, Brockstedt, DG, Bucher, C, Bui, M, Cho, C, Colas, C, et al. *Potent GCN2 Inhibitor Capable of Reversing MDSC-Driven T Cell Suppression Demonstrates In Vivo Efficacy as a Single Agent and in Combination with Anti-Angiogenesis Therapy.* Journal of Medicinal Chemistry, (2022). pp. 12895–12924. https://doi.org/10.1021/acs.jmedchem.2c00736.

[273] Pelster, M, Torre, M, Kannan, GS, Gargano, MA, Mattson, P, Rathmann, B, Bose, N, and Iglesias, JL. *A multicenter, open-label, phase 1a/b study of HC-7366, a modulator of integrated stress response (ISR) kinase GCN2 in subjects with advanced solid tumors.* (2022). https://doi.org/10.1200/JCO.2022.40.16_suppl.TPS3179.

[274] Nazio, F, Bordi, M, Cianfanelli, V, Locatelli, F, and Cecconi, F. *Autophagy and cancer stem cells: molecular mechanisms and therapeutic applications.* Cell Death & Differentiation, (2019). pp. 690–702. https://doi.org/10.1038/s41418-019-0292-y.

[275] Akalay, I, Janji, B, Hasmim, M, Noman, MZ, André, F, De Cremoux, P, Bertheau, P, Badoual, C, Vielh, P, Larsen, AK, et al. *Epithelial-to-mesenchymal transition and autophagy induction in breast carcinoma promote escape from T-cell–mediated lysis.* Cancer Research, (2013). pp. 2418–2427. https://doi.org/10.1158/0008-5472.CAN-12-2432.

[276] Cufí, S, Vazquez-Martin, A, Oliveras-Ferraros, C, Martin-Castillo, B, Vellon, L, and Menendez, JA. *Autophagy positively regulates the CD44+ CD24-/low breast cancer stem-like phenotype.* Cell Cycle, (2011). pp. 3871–3885. https://doi.org/10.4161/cc.10.22.17976.

[277] Pagès, H, Carlson, M, Falcon, S, and Li, N. *AnnotationDbi: Manipulation of SQLite-based annotations in Bioconductor.* (2023). https://doi.org/10.18129/B9.bioc.AnnotationDbi.

[278] Langmead, B, Trapnell, C, Pop, M, and Salzberg, SL. *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.* Genome Biology, (2009). pp. 1–10. https://doi.org/10.1186/gb-2009-10-3-r25.

[279] Langmead, B and Salzberg, SL. *Fast gapped-read alignment with Bowtie 2.* Nature Methods, (2012). pp. 357–359. https://doi.org/10.1038/nmeth.1923.

[280] Wu, T, Hu, E, Xu, S, Chen, M, Guo, P, Dai, Z, Feng, T, Zhou, L, Tang, W, Zhan, L, et al. *clusterProfiler 4.0: A universal enrichment tool for interpreting omics data.* The Innovation, (2021). https://doi.org/10.1016/j.xinn.2021.100141.

[281] Gu, Z, Eils, R, and Schlesner, M. *Complex heatmaps reveal patterns and correlations in multidimensional genomic data.* Bioinformatics, (2016). https://doi.org/10.1093/bioinformatics/btw313.

[282] Martin, M. *Cutadapt removes adapter sequences from high-throughput sequencing reads.* EMBnet.journal, (2011). pp. 10–12. https://doi.org/10.14806/ej.17.1.200.

[283] Badia-i-Mompel, P, Santiago, JV, Braunger, J, Geiss, C, Dimitrov, D, Müller-Dott, S, Taus, P, Dugourd, A, Holland, CH, Flores, ROR, and Saez-Rodriguez, J. *decoupleR:*

ensemble of computational methods to infer biological activities from omics data. Bioinformatics Advances, (2022). https://doi.org/10.1093/bioadv/vbac016.

[284] Love, MI, Huber, W, and Anders, S. *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biology, (2014). pp. 1–21. https://doi.org/10.1186/s13059-014-0550-8.

[285] Blighe, K, Rana, S, and Lewis, M. *EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling.* (2023). https://bioconductor.org/packages/devel/bioc/vignettes/EnhancedVolcano/inst/doc/EnhancedVolcano.html.

[286] Andrews, S. *A quality control tool for high throughput sequence data.* (2010). https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

[287] Wickham, H. *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York, (2016).

[288] Cunningham, F, Allen, JE, Allen, J, Alvarez-Jarreta, J, Amode, MR, Armean, IM, Austine-Orimoloye, O, Azov, AG, Barnes, I, Bennett, R, et al. *Ensembl 2022.* Nucleic Acids Research, (2022). pp. D988–D995. https://doi.org/10.1093/nar/gkab1049.

[289] Kozomara, A, Birgaoanu, M, and Griffiths-Jones, S. *miRBase: from microRNA sequences to function.* Nucleic Acids Research, (2019). pp. D155–D162. https://doi.org/10.1093/nar/gky1141.

[290] Carlson, M. *org.Hs.eg.db: Genome wide annotation for Human.* (2019).

[291] Danecek, P, Bonfield, JK, Liddle, J, Marshall, J, Ohan, V, Pollard, MO, Whitwham, A, Keane, T, McCarthy, SA, Davies, RM, et al. *Twelve years of SAMtools and BCFtools.* Gigascience, (2021). https://doi.org/10.1093/gigascience/giab008.

[292] Dobin, A and Gingeras, TR. *Mapping RNA-seq Reads with STAR.* Current Protocols in Bioinformatics, (2015). pp. 11.14.1–11.14.19. https://doi.org/10.1002/0471250953.bi1114s51.

[293] Liao, Y, Smyth, GK, and Shi, W. *featureCounts: an efficient general purpose program for assigning sequence reads to genomic features.* Bioinformatics, (2014). pp. 923–930. https://doi.org/10.1093/bioinformatics/btt656.

[294] Colaprico, A, Silva, TC, Olsen, C, Garofano, L, Cava, C, Garolini, D, Sabedot, TS, Malta, TM, Pagnotta, SM, Castiglioni, I, et al. *TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data.* Nucleic Acids Research, (2016). https://doi.org/10.1093/nar/gkv1507.

[295] Lexogen. *3‘ mRNA-Seq Library Prep Kit User Guide.* (2023).

9 Appendix

9.1 Appendix: Methods

Table S1: Consumables

Product name	Supplier	Catalogue no.
Agarose	Sigma	A7431
Amersham ECL Prime Western Blotting Detection Reagent	Cytivia	RPN2236
β-mercaptoethanol	Enzo	28625
Blasticidin S hydrochloride	Cambridge BioScience	14499
Bovine Serum Albumin (BSA)	Fisher Bioreagents	BP9704-100
Cas9 Lenti Particles	Merck	LVCAS9NEO-1EA
Cell Counting Kit-8	Stratech	B34304-SEL
cOmplete Proteinase Inhibitor Cocktail	Roche	11873580001
CyQUANT Cell Proliferation Assay	Thermo Fisher Scientific	C7026
DC Protein Assay	Bio-Rad	5000116
DEPC treated water	Invitrogen	AM9906
Dimethyl sulfoxide (DMSO)	Insight Biotechnology	sc-358801
DNA Gel Loading Dye (6X)	Thermo Fisher Scientific	R0611
Dulbecco's Modified Eagle Medium (DMEM), low glucose	Thermo Fisher Scientific	31885
Ethanol	Merck	32221-M
Fetal Bovine Serum (FBS)	Sigma	F9665
Full skirted 96 well PCR plate	Appleton Woods	TA841
G418	Sigma	G8168
GCN2i (GCN2iB)	Selleck Chemicals	S8929
Immobilon Blotting Filter Paper	Merck Millipore	IBFP0813C
Immobilon-P PVDF Membrane	Merck Millipore	IPVH00010
Incucyte ImageLock 96-well plates	Sartorius	4379
Isopropanol	WoldoClean	-
Laemmli buffer (4X)	Bio-Rad	1610747
Methanol	Sigma	32213-M
miRNeasy Mini Kit	QIAGEN	217004
MISSION lentiviral transduction particles	Merck	SHCLNV
MycoAlert Mycoplasma Detection Assay	Lonza	LT07-318
NEBNext High-Fidelity 2X PCR Master Mix	New England BioLabs	M0541L
NEBNext Small RNA Library kit	New England BioLabs	E7330L
NEBNext Ultra II Q5 Master Mix	New England BioLabs	M0544
Nonfat dried milk powder	AppliChem	A0830
NuPAGE 3 -8% Tris-Acetate Gel	Invitrogen	EA03752BOX
NuPAGE 4-12% Bis-Tris Gel	Invitrogen	NP0335
NuPAGE LDS Sample Buffer	Invitrogen	NP0007
NuPAGE MES SDS Running Buffer	Invitrogen	NP0002
NuPAGE MOPS SDS Running Buffer	Invitrogen	NP0001
NuPAGE Sample Reducing Agent	Invitrogen	NP0004

Table S1 continued: Consumables

Product name	Supplier	Catalogue no.
NuPAGE Transfer Buffer	Invitrogen	NP00061
PEIpro	VWR	1010000
Penicillin streptomycin (P/S)	Thermo Fisher Scientific	15140122
Phosphate buffered saline (PBS) (pH 7.4)	Thermo Fisher Scientific	10010
Pierce BCA Protein Assay Kit	Thermo Fisher Scientific	23227
Pierce ECL Plus Western Blotting Substrate	Thermo Fisher Scientific	32134
Pierce Phosphatase Inhibitor Mini Tablets	Thermo Fisher Scientific	A32957
pLX-miR (Lin Human miRNA CRISPR Knockout Library)	Addgene	112200
pMD2.G	Addgene	12259
Polybrene	Insight Biotechnology	sc-134220
Precision Plus Protein Dual Color Standards	Bio-Rad	1610374
PrepX reagents	Takara	64010
psPAX2	Addgene	12260
PureLink RNase A	Thermo Fisher Scientific	12091021
QIAGEN Blood & Cell Culture DNA Midi Kit	QIAGEN	13343
QIAGEN maxi column	QIAGEN	12163
QIAquick Gel Extraction Kit	QIAGEN	28704
QIAquick PCR Purification Kit	QIAGEN	28104
Quant-it RiboGreen RNA Assay Kit	Invitrogen	R11490
QuantSeq 3'mRNA-Seq Kit	Lexogen	191-195
Qubit dsDNA High Sensitivity Assay	Invitrogen	Q33231
Qubit RNA Broad Range Assay	QIAGEN	Q10211
Quick Start Bovine Serum Albumin Standard Set	Bio-Rad	5000207
Quick-Load Purple 100 bp DNA Ladder	New England BioLabs	N0551S
QuickTiter Lentivirus Titer Kit	Cell BioLabs	VPK-107
Rainbow Molecular Weight Markers (full-range)	VWR	RPN800E
RIPA buffer	Sigma	R0278
RNA ScreenTape assay	Agilent	5067
Sodium acetete	Invitrogen	AM9740
SYBR Safe DNA Gel Stain	Invitrogen	S33102
TKOv3 (Toronto KnockOut CRISPR Library - Version 3)	Addgene	90294
Tri Reagent	Sigma	T9424
Tris buffered saline (TBS)	Santa Cruz Biotechnology	sc-362186
Tris-acetate-EDTA (TAE) buffer	Thermo Fisher Scientific	B49
Tris-EDTA (TE) buffer	VWR	A8569
Trypan blue solution	Thermo Fisher Scientific	15250061
Trypsin	Lonza	BE02-007E
TURBO DNA-free Kit	Invitrogen	AM1907
Tween 20	National Diagnostics	EC-607
Western Blotting Filter Paper	Thermo Fisher Scientific	84783
Wizard Genomic DNA Purification Kit	Promega	A1620

Table S2: Antibodies

Target	Antibody	Supplier	Catalogue no.
β-actin	Mouse monoclonal antibody	Abcam	ab8226
β-actin	Mouse monoclonal antibody HRP	Sigma	A3854
Cas9	Mouse monoclonal antibody	Cell Signalling Technology	14697
E-cadherin	Rabbit monoclonal antibody	Cell Signalling Technology	3195
GCN2	Rabbit polyclonal antibody	Cell Signalling Technology	3302S
p-GCN2	Rabbit monoclonal recombinant antibody	Abcam	ab75836
HIF-1α	Mouse monoclonal antibody	BD Biosciences	610959
HIF-2α	Rabbit polyclonal antibody	Novus Bio	NB100-122
Vimentin	Rabbit monoclonal antibody	Cell Signalling Technology	5741
	Secondary antibodies		
Anti-mouse	Goat anti-mouse IgG HRP	Dako	P0447
Anti-rabbit	Goat anti-rabbit IgG HRP	Invitrogen	65-6120

Table S3: Machines

Machine	Supplier
Apollo next-generation sequencing library preparation system	WaferGen
Bio-Rad C1000 Touch Thermal Cycler	Bio-Rad
BluePippin	SageScience
FLUOstar OPTIMA microplate reader	BMG Labtech
ImageQuant LAS 4000 Mini	GE Healthcare
Incucyte S3	Sartorius
Incucyte Zoom	Sartorius
Invitrogen iBright 1500	Invitrogen
Invitrogen XCell SureLock Protein Gel Electrophoresis tank	Invitrogen
Invivo$_2$ 500 workstation	Baker Ruskinn
LUNA II automated cell counter	Labtech International
Mini Trans-Blot Cell	Bio-Rad
NanoDrop One Spectrophotometer	Thermo Fisher Scientific
NextSeq 500	Illumina
NovaSeq 6000	Illumina
Novex Semi-Dry Blotter	Invitrogen
POLARstar OMEGA microplate reader	BMG Labtech
Qubit 4 Fluorometer	Invitrogen
TapeStation	Agilent
Tetrad PCR Thermal Cycler	Bio-Rad
Trans-Blot Turbo Transfer System	Bio-Rad

Table S4: Programs and databases

Program	Citation
AnnotationDbi	Pagès et al. [277]
Bowtie	Langmead et al. [278]
Bowtie2	Langmead and Salzberg [279]
clusterProfiler	Wu et al. [280]
ComplexHeatmap	Gu et al. [281]
Cutadapt	Martin [282]
decoupleR	Badia-i-Mompel et al. [283]
DESeq2	Love et al. [284]
EnhancedVolcano	Blighe et al. [285]
FastQC	Andrews [286]
ggplot2	Wickham [287]
Hsa.GRCh38	Cunningham et al. [288]
MAGeCK	Li et al. [219]
Manatee	Handzlik et al. [105]
miRBase	Kozomara et al. [289]
miRNAtap	Pajak and Simpson [223]
org.Hs.eg.db	Carlson [290]
SAMtools	Danecek et al. [291]
SigQC	Dhawan et al. [117]
STAR	Dobin and Gingeras [292]
Subread	Liao et al. [293]
TCGAbiolinks	Colaprico et al. [294]

9.2 Appendix: Optimisation of a bioinformatics pipeline to quantify microRNAs in small RNA sequencing data

The custom python script used to add multi-mapping statistics to the Bowtie output.

```
import os
cwd = os.getcwd()
with open (cwd+"/samplelist.txt", "rt") as samplelist: # open the list of samples
    for line in samplelist:
        filepath = str(cwd+"/"+str(line)) # extract the sample name from the list of
            samples and add the current working directory to create the filepath
        filepath = str(filepath).rstrip("\n")
        sample = str(line).rstrip("\n") # create a new variable containing the sample
            name
        with open(filepath, "rt") as sam, open(cwd+"/"+sample+"_NH.sam", "wt",
            encoding='utf-8') as output: # open the SAM file, and create a new file
            for the output to be stored in
            header = [] # create empty list
            reads = [] # create empty list
            for ln in sam:
                if ln.startswith("@"):
                    header.append(ln[0:]) # append the line to the header list
                else: # if the line doesn't begin with an @ (i.e. the rest of the SAM
                    file)
                    a = ln.split("\t")[-1] # split the line by tabs and take the last
                        value (e.g. XM:i:2)
                    ab = int(a.split(":")[-1]) -1 # split this by colons, and take the
                        last value (e.g. 2). The is the number of times the read has
                        aligned plus 1. Convert this to an integer and minus 1 from it,
                        giving the number of times the read mapped
                    abc = "NH:i:"+str(ab) # add the value generated above to the string
                        NH:i: to create the NH tag relevant for this line
                    reads.append(ln.rstrip("\n") +"\t"+abc+"\n") # to the empty "reads"
                        list, add the full text line with the NH tag appended
            output.write(''.join(header)) # add the header to the output file
            output.write(''.join(reads)) # add the lines containing the reads with
                their new NH tag to the output file
```

9.3 Appendix: Transcriptional adaptation to chronic hypoxia

9.3.1 Optimisation of 3' mRNA-seq alignment

Passing 3' QuantSeq data through the traditional RNA alignment pipeline used by the Buffa lab achieved low mapping efficiency (Figure S1). The original pipeline had been designed and optimised for use on polyA selected RNA-seq data. However, many of the parameters chosen were not suitable for 3' QuantSeq data as demonstrated by the low percentage of uniquely mapped reads. A mapping rate as low as 47% indicates a problem with library preparation or processing [292]. The first step of optimisation involved modifying the pipeline to treat the dataset as single-ended (i.e. read 2 was excluded from the analysis). This is because, during the sequencing step, read 2 is primed before the polyA tail and the low complexity results in poor sequencing quality (Figure S2). The second step of optimisation included adapter and polyA trimming, achieving a final mapping rate of 85% (Figure S1). This was considered sufficient to continue with downstream analysis [292].

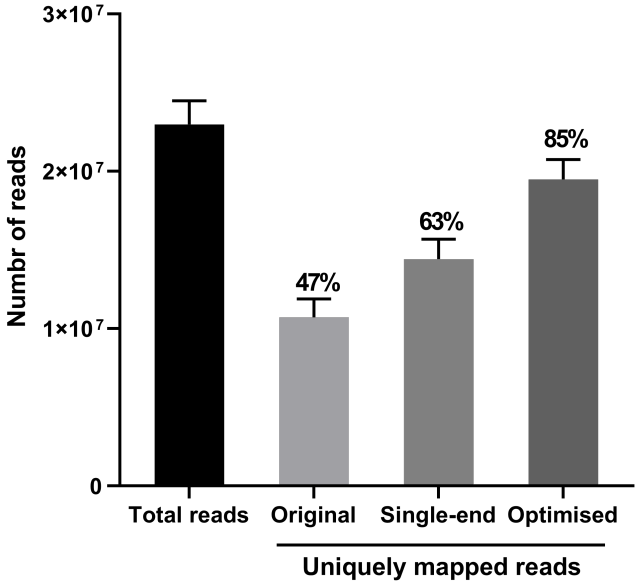

Figure S1: Optimisation of 3' mRNA-seq alignment. Reads were processed using one of three pipelines and the number of uniquely mapped reads was calculated. The original pipeline (designed for processing a polyA selected RNA-seq library) mapped 47% of the total reads. The mapping rate was improved by inputting only read 1 into the pipeline (single-end) and by adding steps to trim adapters and polyA sequences (optimised).

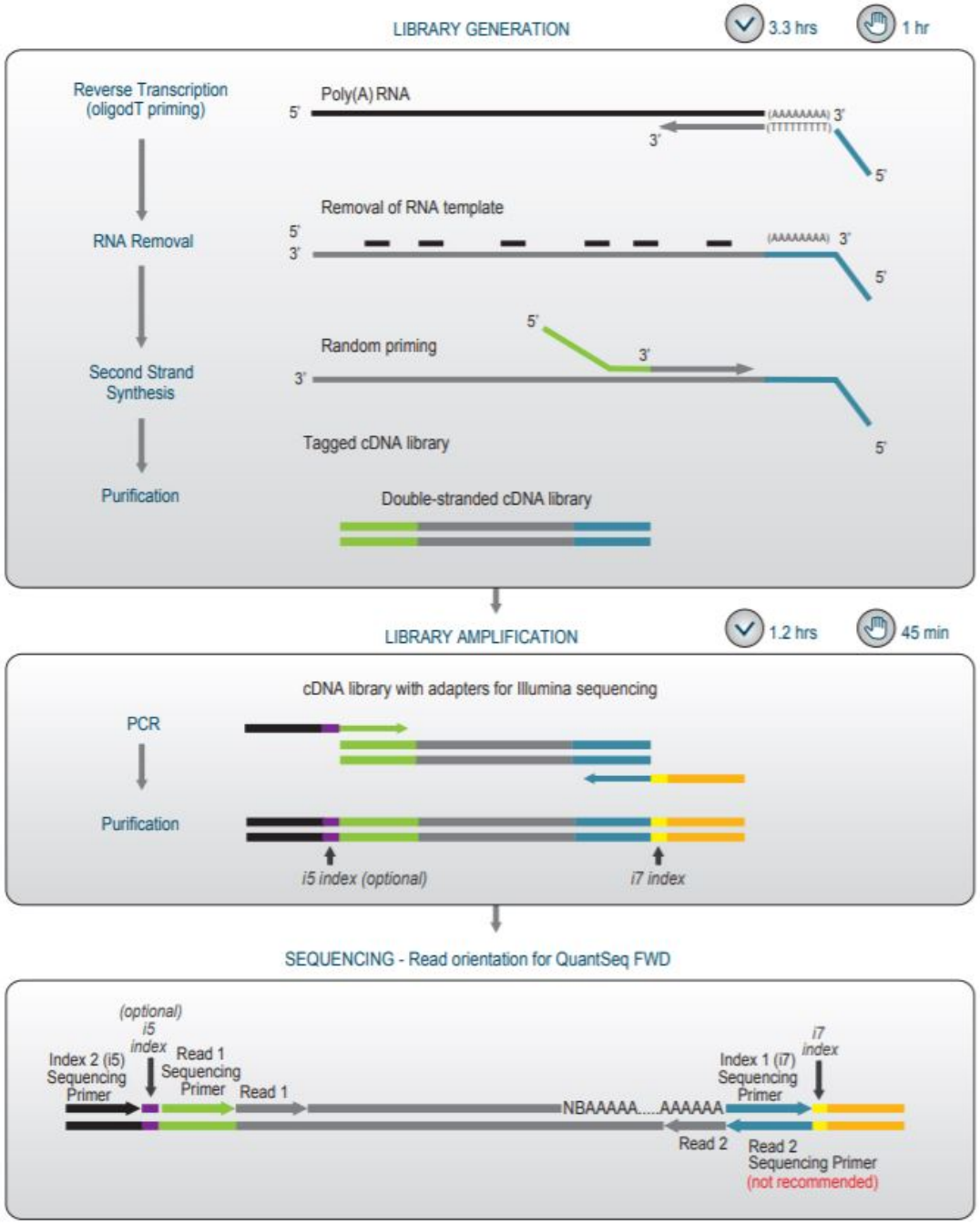

Figure S2: 3' mRNA-seq library preparation outline. During sequencing, read 2 is primed before the polyA tail, resulting in poor overall quality. Figure from [295].

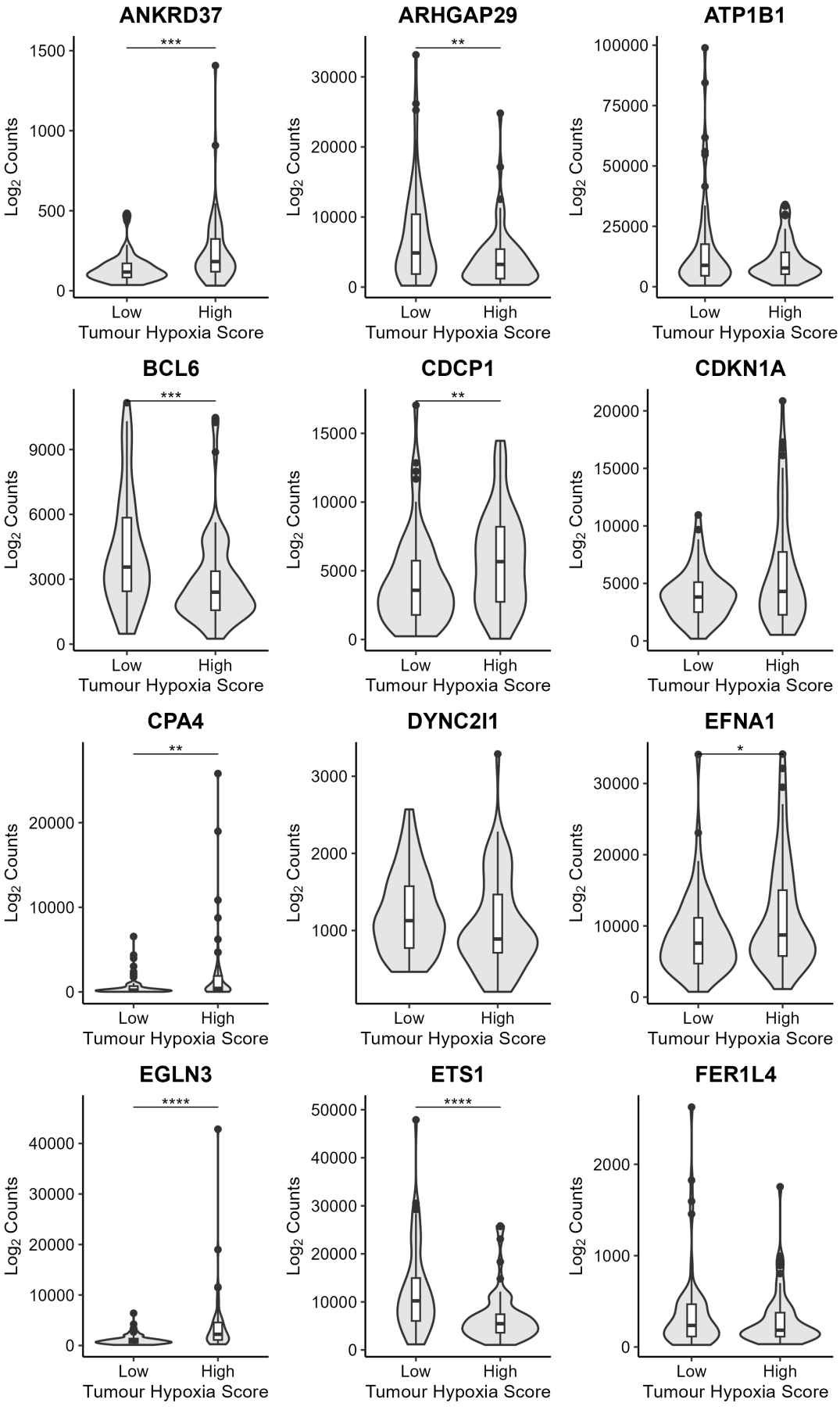

Figure S3: Expression of chronic hypoxia specific genes in breast cancer samples from TCGA.

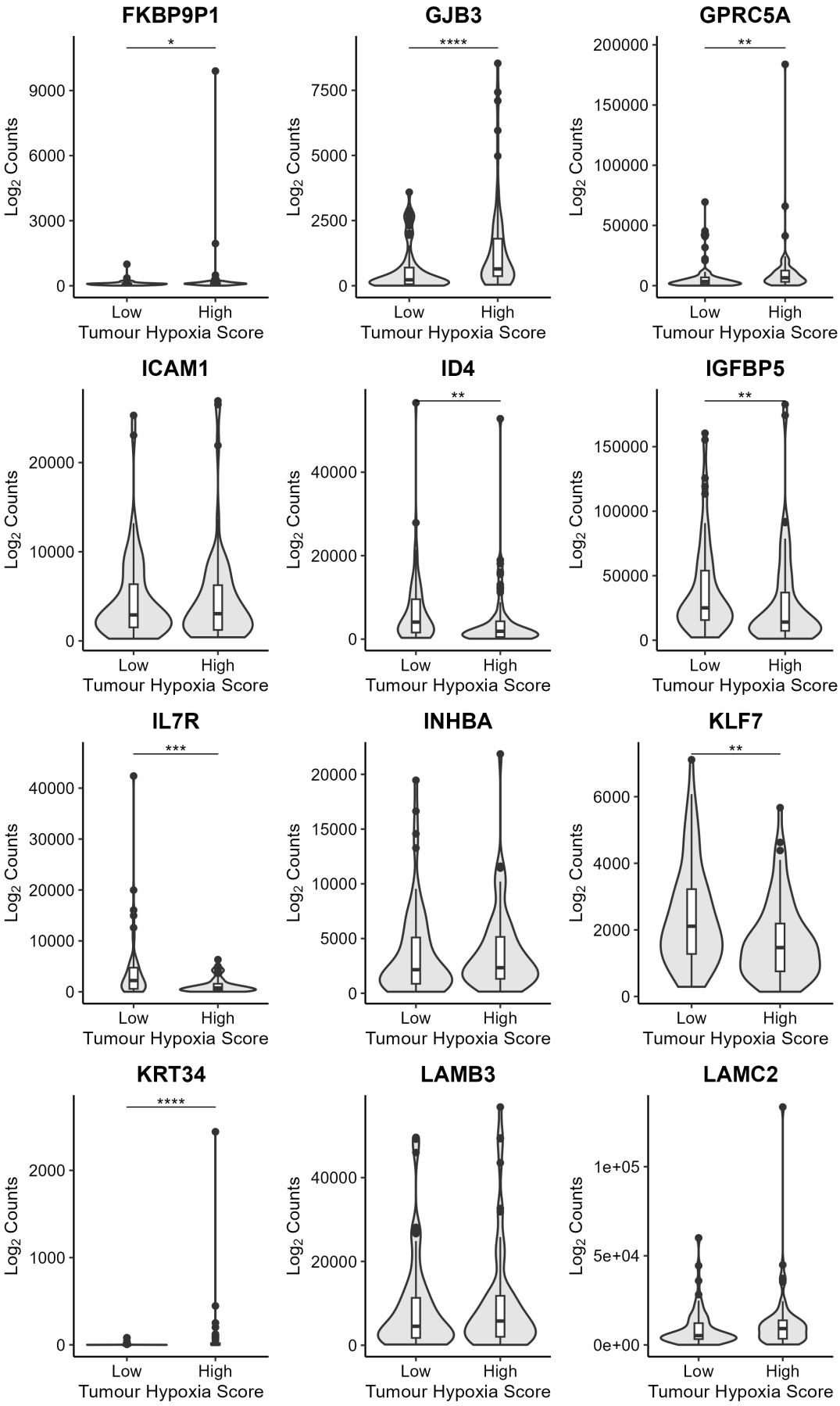

Figure S3: Expression of chronic hypoxia specific genes in breast cancer samples from TCGA.

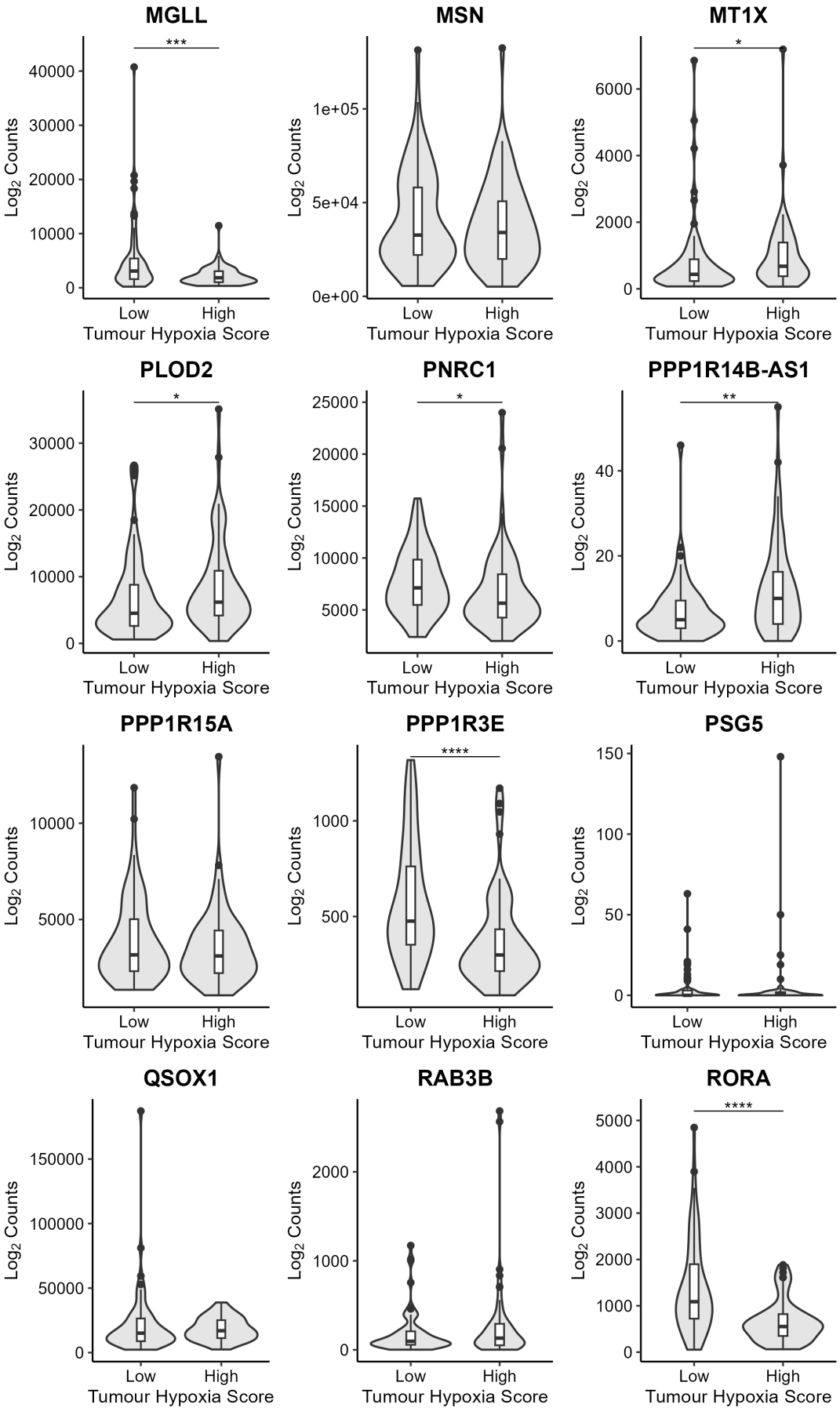

Figure S3: Expression of chronic hypoxia specific genes in breast cancer samples from TCGA.

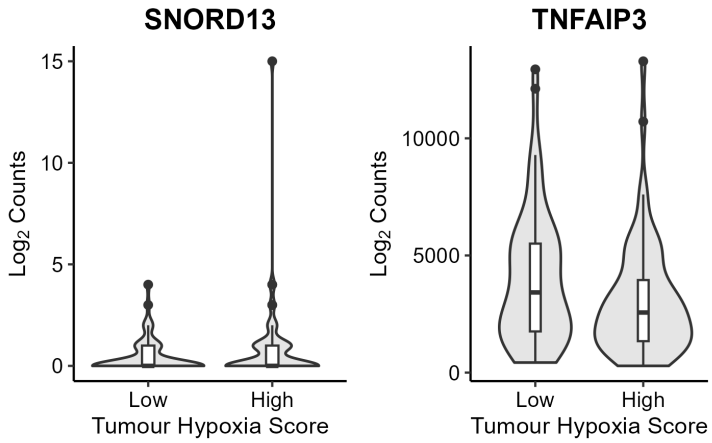

Figure S3: Expression of chronic hypoxia specific genes in breast cancer samples from TCGA. Expression of chronic hypoxia specific genes in breast cancer samples from TCGA. Only basal tumours (i.e. typically triple-negative) were used. Each sample was assigned a hypoxia score according to expression of genes in the Buffa hypoxia signature using sigQC. Samples with a hypoxia score in the upper tertile were labelled as ”high” and samples with a hypoxia score in the lower tertile labelled as ”low”. Significance was tested using an unpaired two-sample Wilcoxon test. *ANKRD37* is found in the Buffa hypoxia signature.

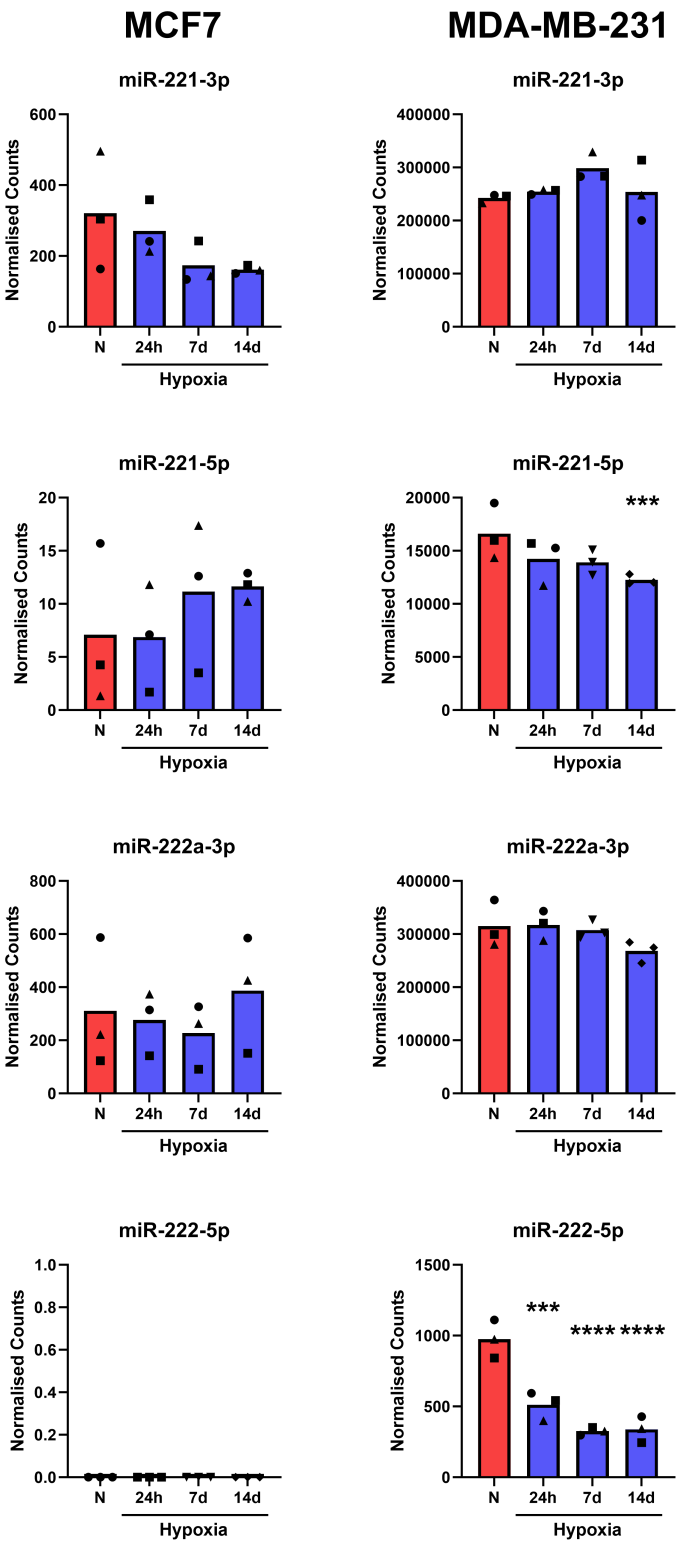

Figure S4: Expression of miR-221/222 in MCF7 and MDA-MB-231 cells. miR-221 and miR-222 are never significantly (adjusted p value $\leqslant 0.05$) upregulated under chronic hypoxia in MCF7 and MDA-MB-231 cells. Each shape represents a different biological replicate (* adjusted p value $\leqslant 0.05$, ** adjusted p value $\leqslant 0.01$, *** adjusted p value $\leqslant 0.001$, **** adjusted p value $\leqslant 0.0001$).

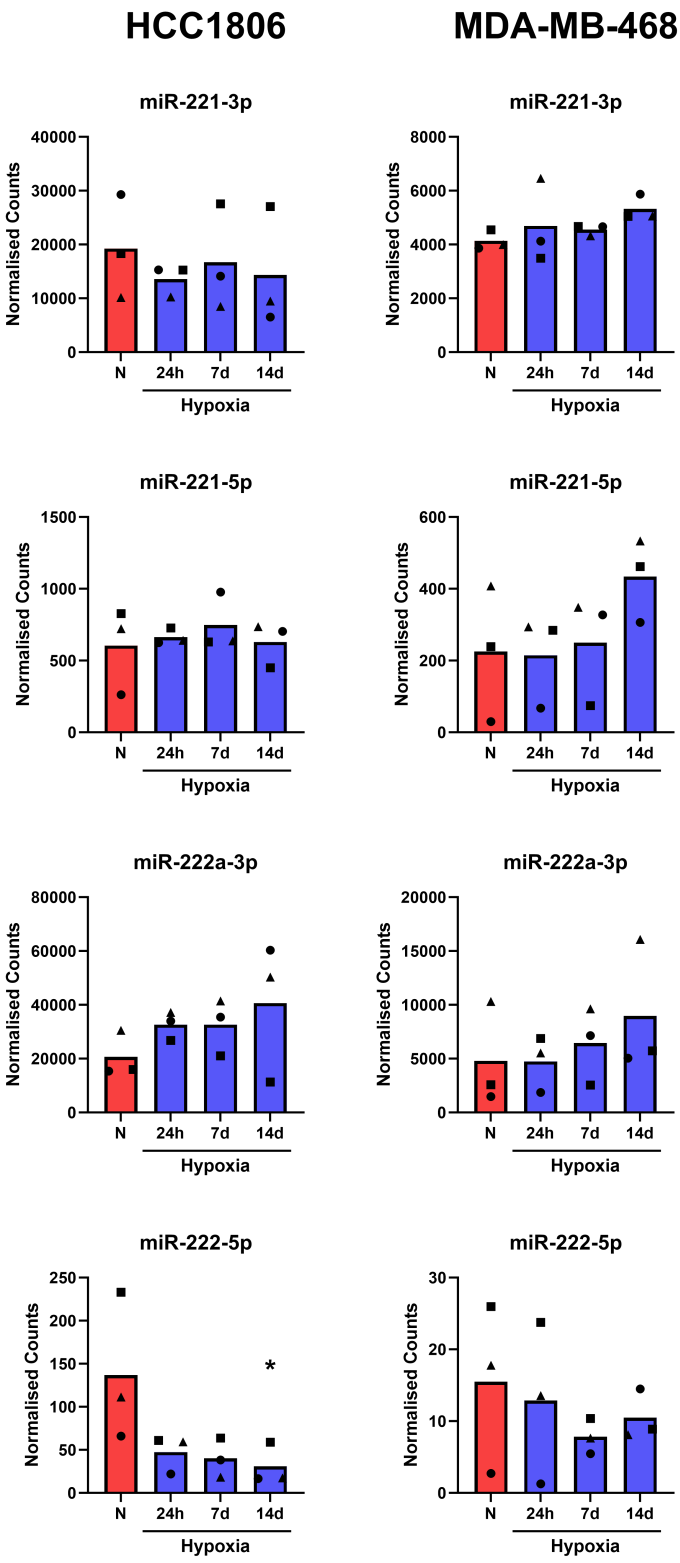

Figure S5: Expression of miR-221/222 in HCC1806 and MDA-MB-468 cells. miR-221 and miR-222 are never significantly (adjusted p value $\leqslant 0.05$) upregulated under chronic hypoxia in HCC1806 and MDA-MB-468 cells. Each shape represents a different biological replicate (* adjusted p value $\leqslant 0.05$).

9.4 Appendix: Gene essentiality under chronic hypoxia

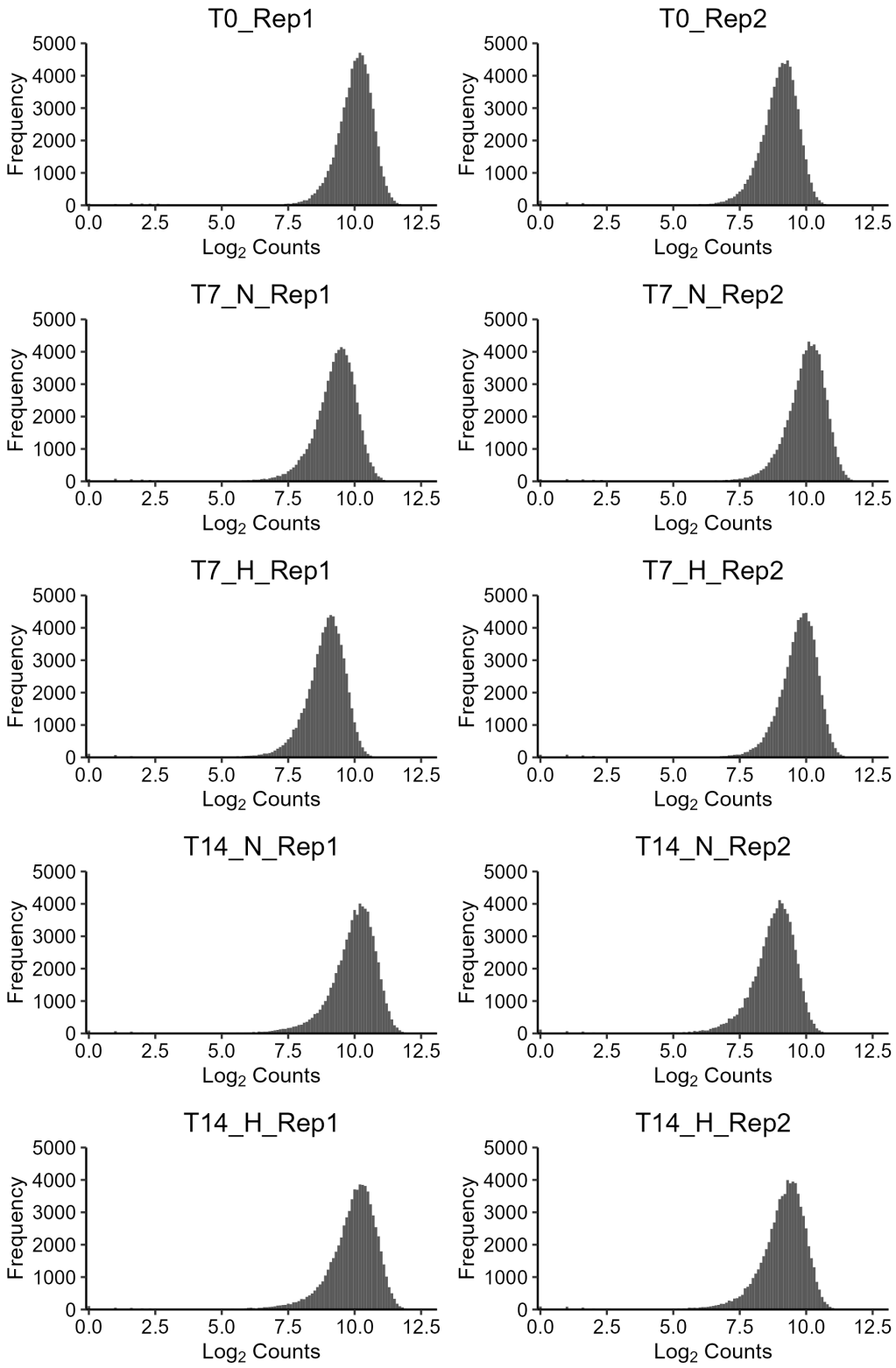

Figure S6: Distribution of guides in the TKOv3 CRISPR screen - MDA-MB-468. The guides produced from sequencing of the MDA-MB-468 CRISPR samples are roughly equally distributed

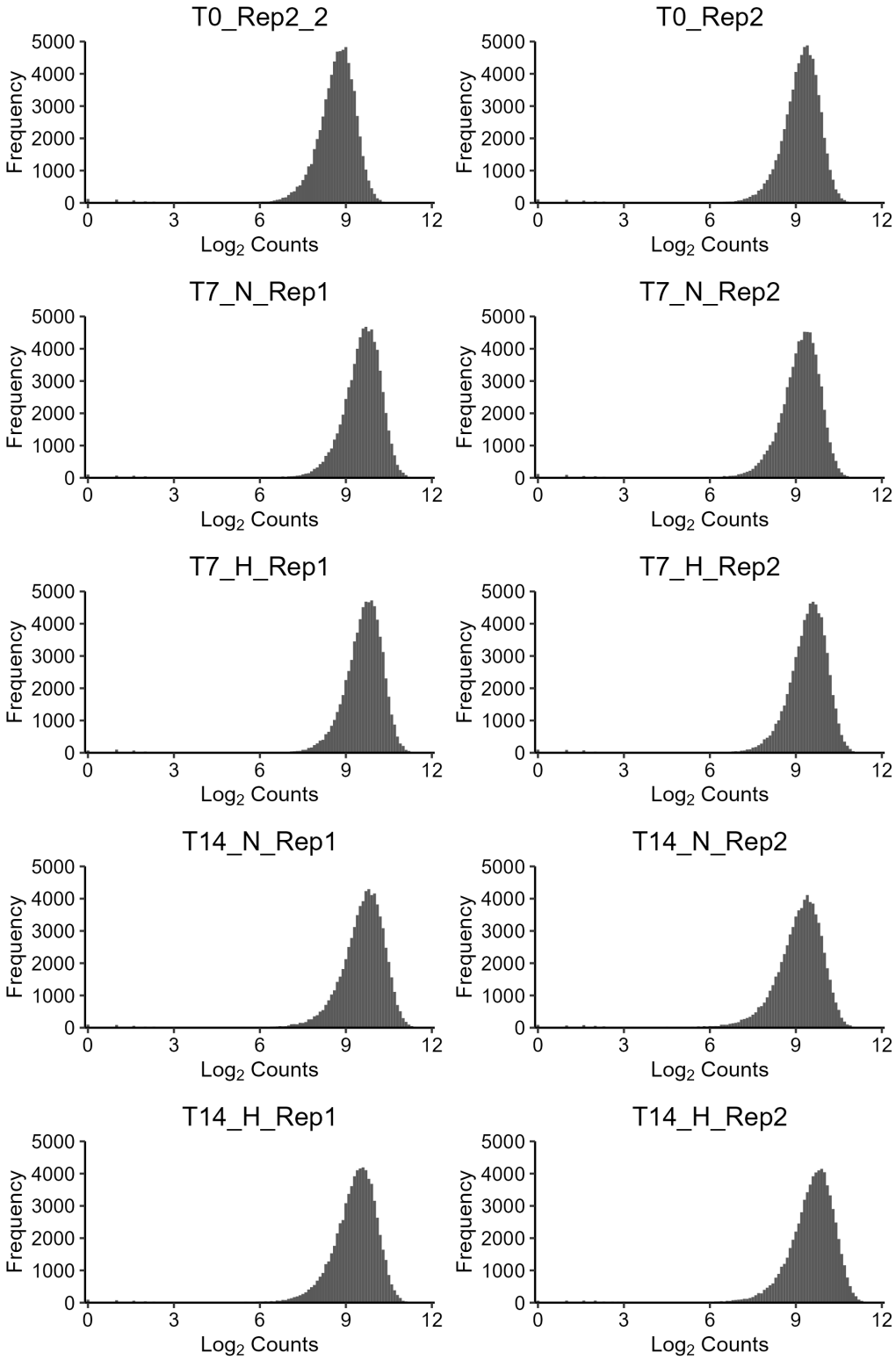

Figure S7: Distribution of guides in the TKOv3 CRISPR screen - MCF7. The guides produced from sequencing of the MCF7 CRISPR screen samples are roughly equally distributed

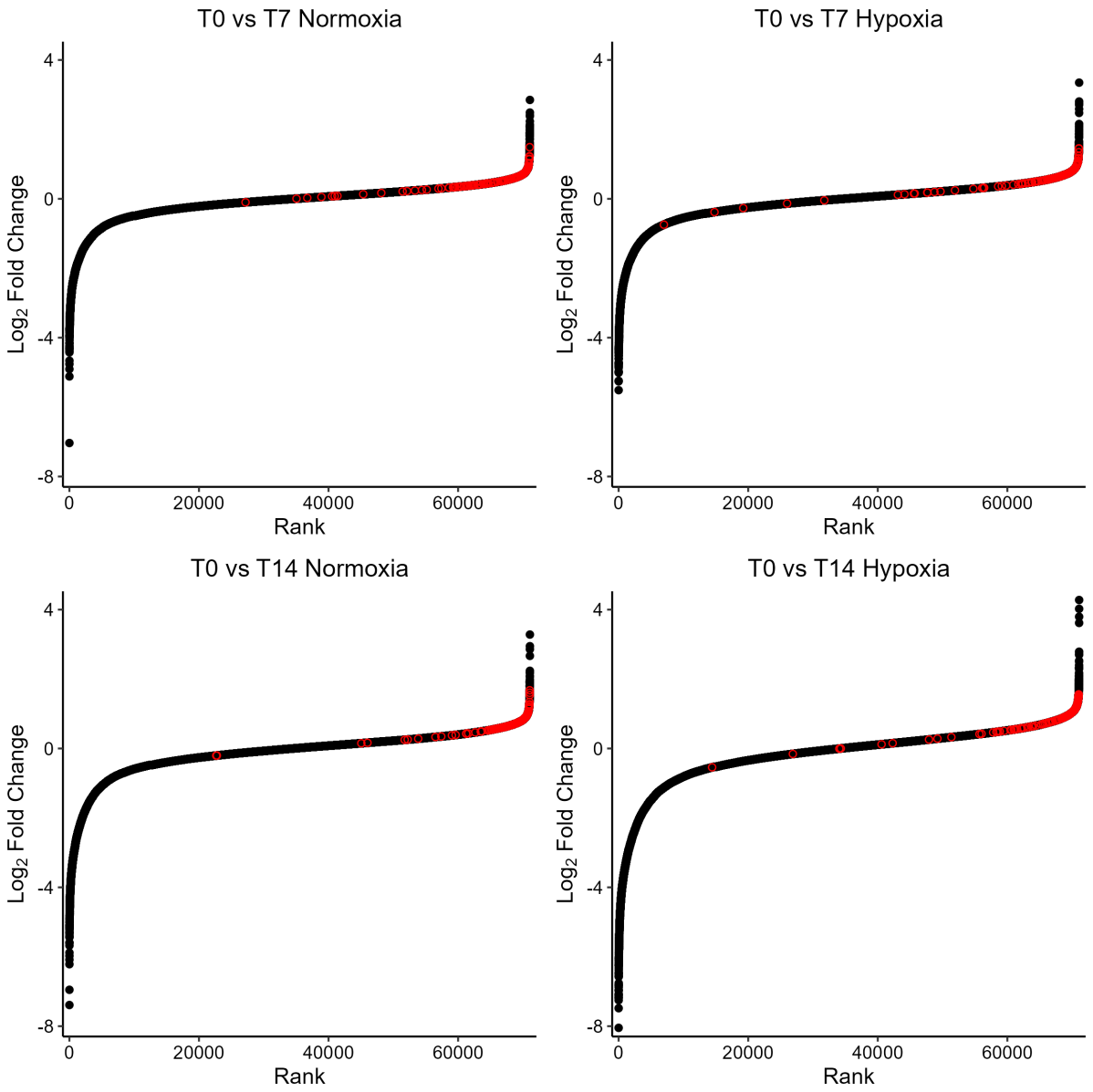

Figure S8: Log_2 fold change of non-targeting guides - MDA-MB-231 screen. Guides targeting EGFP, luciferase, and LacZ (i.e. non-targeting to the human genome) generally have a positive log_2 fold change, showing enrichment over time.

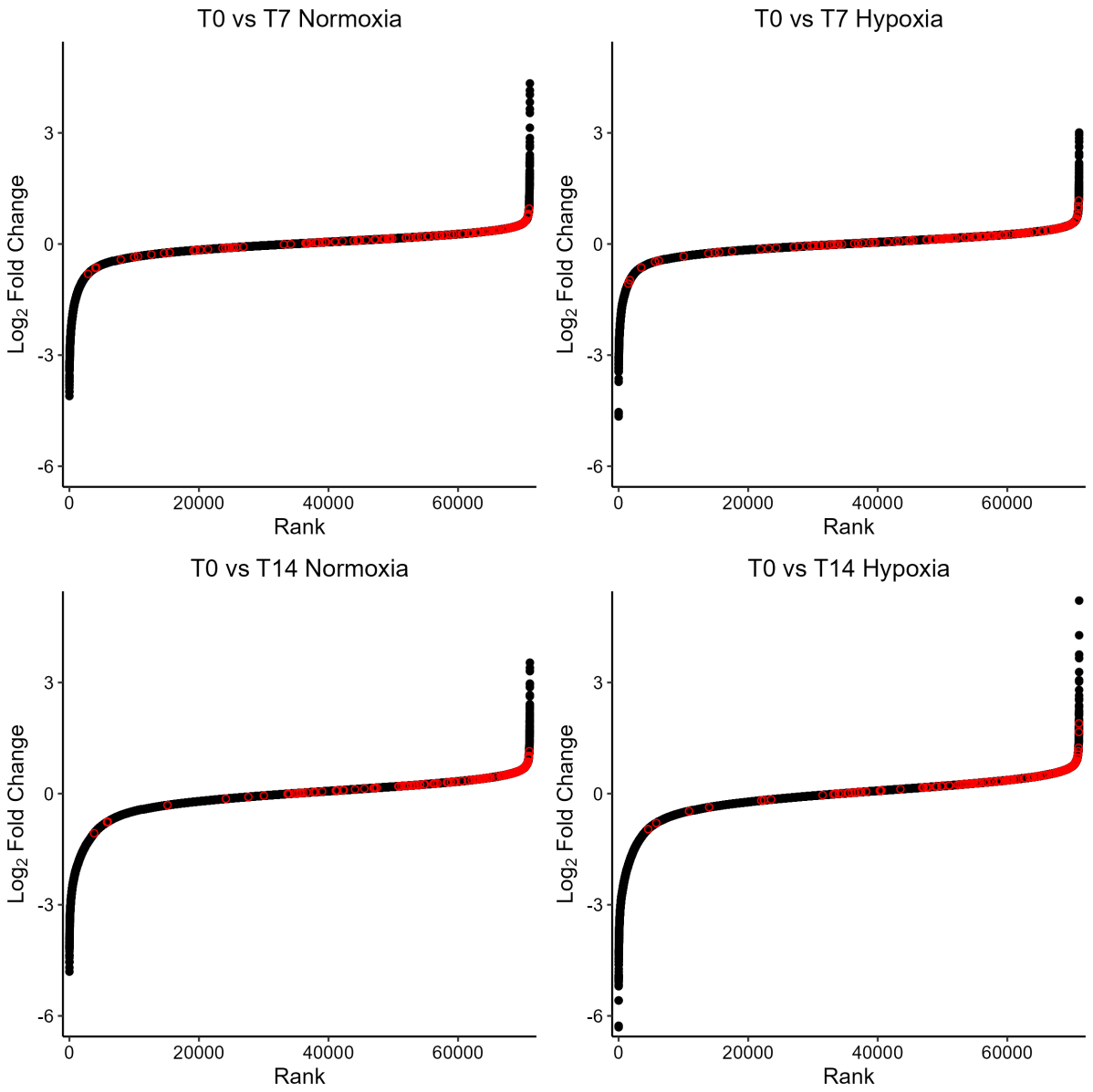

Figure S9: Log_2 fold change of non-targeting guides - MDA-MB-468 screen. Guides targeting EGFP, luciferase, and LacZ (i.e. non-targeting to the human genome) generally have a positive log_2 fold change, showing enrichment over time.

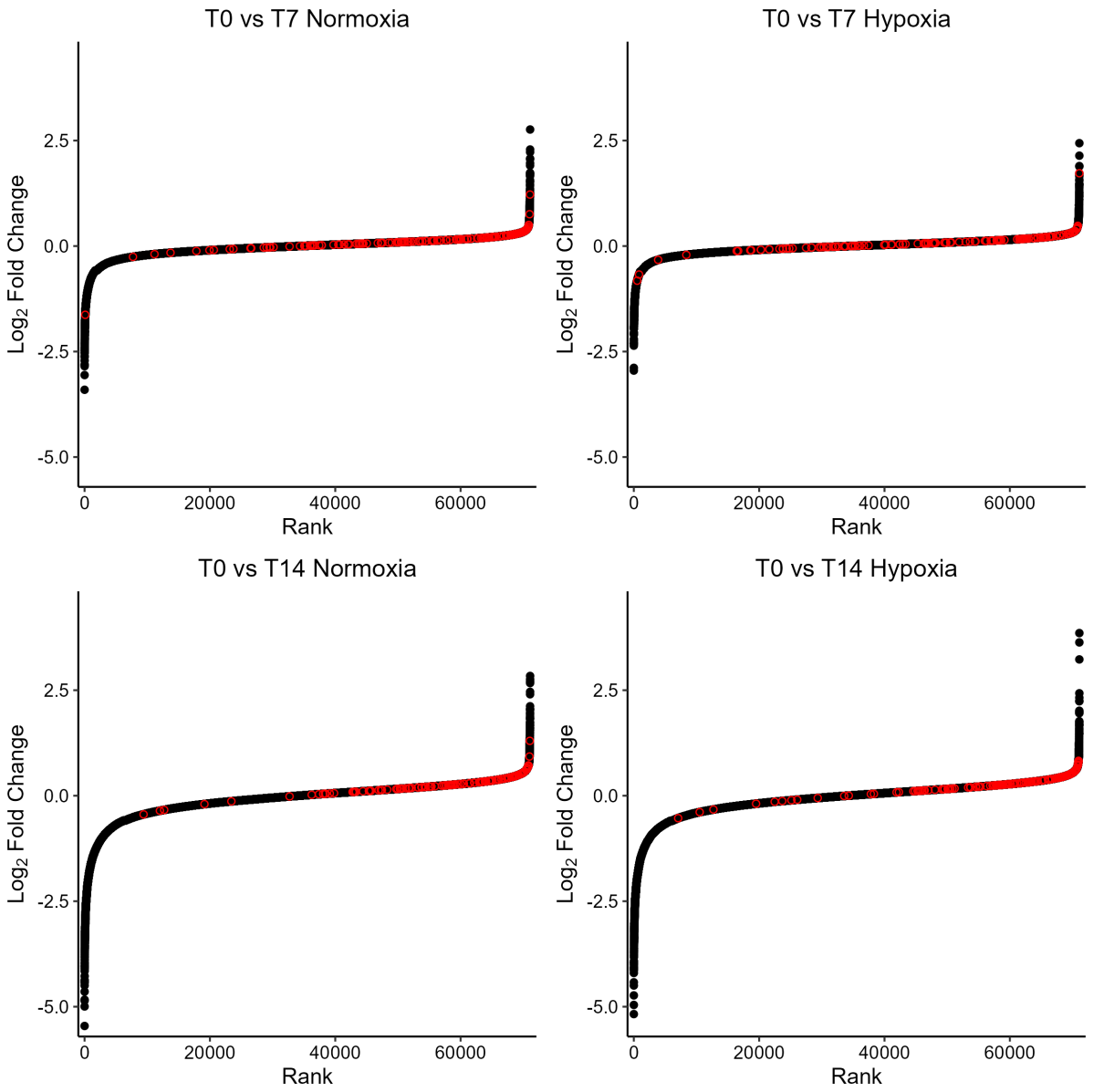

Figure S10: Log_2 fold change of non-targeting guides - MCF7 screen. Guides targeting EGFP, luciferase, and LacZ (i.e. non-targeting to the human genome) generally have a positive log_2 fold change, showing enrichment over time.

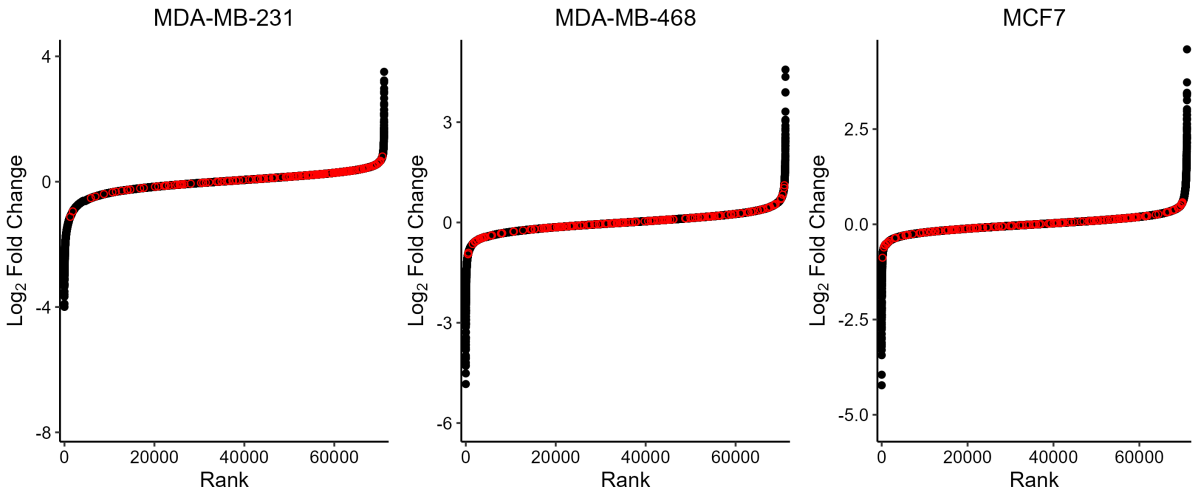

Figure S11: Log_2 fold change of non-targeting guides in T14 normoxia vs T14 hypoxia comparisons. Guides targeting EGFP, luciferase, and LacZ (i.e. non-targeting to the human genome) have a similar log_2 fold change distribution across cell lines when T14 normoxia is compared with T14 hypoxia.

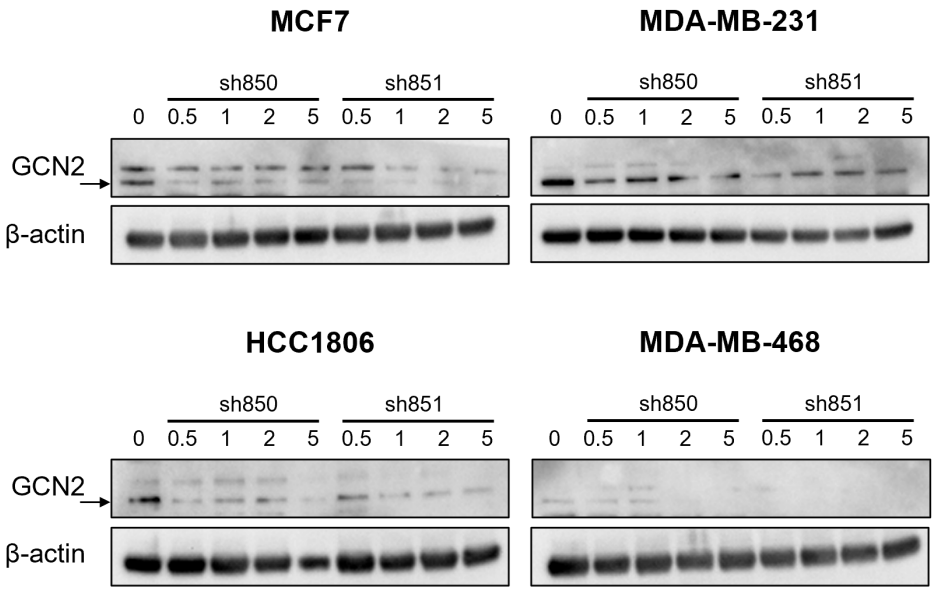

Figure S12: GCN2 knockdown western blot showing all MOIs tested. Western blots showing GCN2 in control (0) and shRNA knockdown cells at MOIs 0.5-5. sh851 appears better able to knockdown the protein. Knockdown level does not strongly correlate to MOI.

9.5 Appendix: A microRNA-ome wide chronic hypoxia CRISPR screen

The custom python script used to remove duplicated guide sequences targeting the same mature miRNA and label guide sequences which are identical but target different miRNAs.

```
import pandas as pd
import numpy as np
import string
with open("pLX-miR_guides.csv", "rt") as pLX:
    output = pd.DataFrame()
    sg = []
    seq = []
    miRNA = []
    for ln in pLX:
        a = ln.split(",")[2]
        a = a.rstrip("\n")
        if a.endswith("-1"):
            a = a[:-2]
        if a.endswith("-2"):
            a = a[:-2]
        if a.endswith("-3"):
            a = a[:-2]
        if a.endswith("-4"):
            a = a[:-2]
        if a.endswith("-5"):
            a = a[:-2]
        miRNA.append(a)
        b = ln.split(",")[1]
        seq.append(b)
        c = ln.split(",")[0]
        sg.append(c)
    mylist = [sg, seq, miRNA]
    output = pd.DataFrame(mylist).transpose()
    output2 = output.drop_duplicates()
    output2.columns = ["sgRNA", "Sequence", "Target"]
    output2["sgRNA"] = np.where(output2["sgRNA"].duplicated(keep=False),
        output2["sgRNA"] + "." +
```

```
    output2.groupby("sgRNA").cumcount().add(1).astype(str), output2["sgRNA"])
output2.to_csv("pLX-miR_guides_new.csv")
```

R code used to produce the curated pLX-miR FASTA file.

```
plx <- read.csv("pLX-miR_guides_curated_fa.csv")
fasta <- plx$Sequence
fasta <- as.data.frame(rep(fasta, each = 2))
fasta[c(TRUE,FALSE),] <- plx$sgRNA
fasta[c(TRUE,FALSE),] <- paste0(">", fasta[c(TRUE,FALSE),])
write.csv(fasta, "pLX-miR.fa")
```

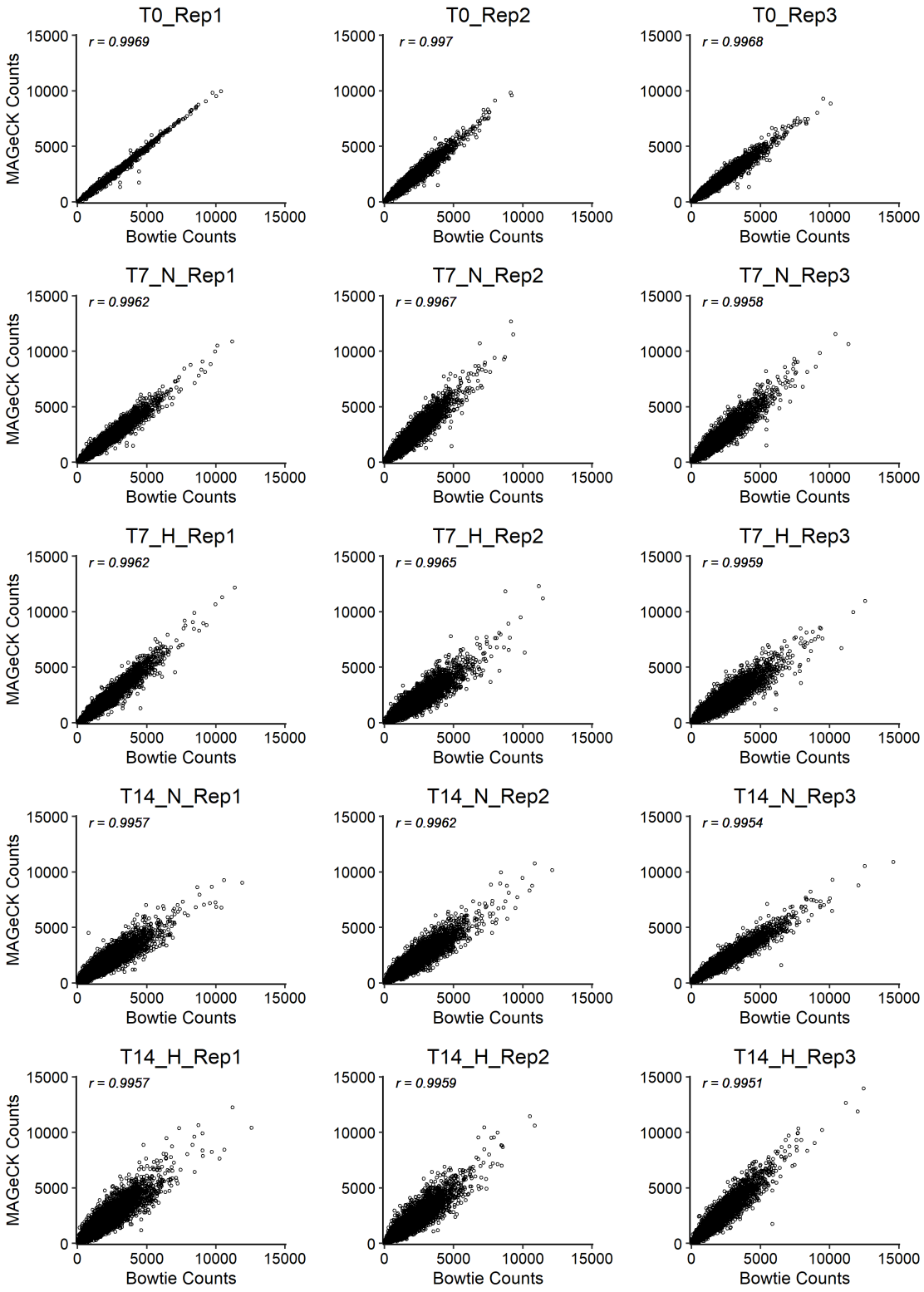

Figure S13: Correlation of MAGeCK and Bowtie mapping. The counts for each guide produced by MAGeCK and the Bowtie method are strongly correlated. The lowest Pearson correlation coefficient is 0.9951.

The modified `getPredictedTargets` functions used to predict the gene targets of miRNAs.

```
my|\_|getPredictedTargets <- function (mirna, sources = c("diana", "miranda",
    "mirdb", "pictar", "targetscan"), species = "hsa", min_src = 1, method = "geom",
    promote = TRUE, synonyms = TRUE, both_strands = FALSE, ...)
{
    if (!(species %in% c("mmu", "hsa", "rno",
        "dme"))) {
        warning(paste("species", species, " not supported in the current version",
            sep = ""))
        return(NULL)
    }
    n_sources <- length(sources)
    if (n_sources < min_src) {
        warning("min_src > n_sources!")
        return(NULL)
    }
    if (min_src <= 0) {
        warning("min_src changed to 1")
        min_src <- 1
    }
    if (nchar(mirna) < 3) {
        warning("unrecognised miRNA")
        return(NULL)
    }
    l_outputs <- list()
    cols <- c("GeneID", "score")
    if (substr(mirna, 1, 3) == species) {
        mirna = substr(mirna, 5, nchar(mirna)) # This was edited to prevent trimming
            of longer miRNA names
    }
    for (src in sources) {
        l_outputs[[src]] <- getTargetsFromSource(mirna, species,
            source = src, synonyms = synonyms, both_strands = both_strands)
    }
    merged.scores <- suppressWarnings(Reduce(function(...) .mergeRename(...,
        by = "GeneID", all = TRUE), l_outputs))
```

```
    if (is.null(merged.scores)) {
        warning(paste("no targets found for mirna ", mirna,
            sep = ""))
        return(NULL)
    }
    valid_srcs <- (1:(n_sources + 1))[colSums(merged.scores,
        na.rm = TRUE) > 0]
    if (length(valid_srcs) < 2) {
        warning(paste("no targets found for mirna ", mirna,
            sep = ""))
        return(NULL)
    }
    n_valid_srcs <- length(valid_srcs) - 1
    if (n_valid_srcs < min_src) {
        warning(paste("sources which returned a target list < min_src"))
        #min_src <- n_valid_srcs # This line was removed
      return(NULL)
    }
    #merged.scores <- merged.scores[, valid_srcs] # This line was removed to prevent
        removal of empty columns
    result <- my_aggregateRanks(merged.scores, n_sources, n_valid_srcs, min_src,
        method = method, promote = promote)
    return(result)
}


my_aggregateRanks <- function(ranks, n_sources, n_valid_srcs, min_src, # The
    n_sources option was added here
                          method = "geom", promote=TRUE) {

    #if (n_valid_srcs>1) {
        target_found <- rowSums(!is.na(ranks[,2:(n_sources+1)])) # Changed
            n_valid_srcs to n_sources to account for the table having more columns.
   # } else {
    #    target_found <- 1*(!is.na(ranks[,2:(n_valid_srcs+1)]))
    #}
```

```
    ranks <- ranks[target_found>=min_src,]
    ranks <- ranks[!duplicated(ranks$GeneID),]
    ranks <- ranks[!is.na(ranks$GeneID),]

    row.names(ranks) <- ranks$GeneID

  #  if (n_valid_srcs==1) {
   #     data1 <- data.frame(source1scores=ranks[,2], row.names = row.names(ranks))
  #  } else {
        data1 <- ranks[,2:(n_sources+1)] # Changed the n_valid_srcs to n_sources to
            account for the table having more columns.
        data1 <- as.data.frame(data1)
   # }
    num.gene <- dim(data1)[1]

    if (num.gene<1) {
        warning( "not enough targets overlapping between sources,
reduce min_srcs parameter or add sources")
        return(NULL)
    } else if (num.gene==1) {
        warning("only one target found")
        result <- data.frame(matrix(1,nrow=1,ncol=n_sources+2)) # Changed the
            n_valid_srcs to n_sources to account for the table having more columns.
        colnames(result) <- c(paste("source_",1:n_sources,sep=""),
                              "rank_product","rank_final") # Changed the n_valid_srcs to
                                  n_sources to account for the fact that the table having
                                  more columns.
        row.names(result) <- row.names(data1)
        return(result)
    }

    rank_ind <- apply(data1, 2,
    function(x) (num.gene+1) - rank(x, ties.method="average", na.last = FALSE))
```

```
    if (method=="geom") {
        result <- .aggregateGeom(data1, rank_ind, promote)
    }  else if (method=="max") {
        result <- .aggregateMinMax(data1, rank_ind, minmax=method)
    } else {
        result <- .aggregateMinMax(data1, rank_ind, minmax=method)
    }

    colnames(result) <- c(paste("source_",1:n_sources,sep=""), # Changed the
        n_valid_srcs to n_sources
                          "rank_product","rank_final")
    return(result)
}
}
```

Table S5: DNA extraction results

Sample	DNA Concentration (ng/µL)			A260/A280			A260/A230		
	Reading 1	Reading 2	Average	Reading 1	Reading 2	Average	Reading 1	Reading 2	Average
T0_Rep1	217.0	218.4	217.7	1.89	1.89	1.89	2.35	2.34	2.35
T7_N_Rep1	217.7	200.1	208.9	1.91	1.90	1.91	2.38	2.39	2.39
T7_H_Rep1	193.5	192.6	193.1	1.90	1.90	1.90	2.37	2.39	2.38
T14_N_Rep1	230.9	265.7	248.3	1.90	1.90	1.90	2.38	2.38	2.38
T14_H_Rep1	245.2	242.8	244.0	1.91	1.91	1.91	2.37	2.39	2.38
T0_Rep2	282.9	297.7	290.3	1.93	1.93	1.93	2.32	2.31	2.32
T7_N_Rep2	269.6	275.4	272.5	1.94	1.94	1.94	2.31	2.32	2.32
T7_H_Rep2	301.7	308.2	305.0	1.92	1.93	1.93	2.32	2.32	2.32
T14_N_Rep2	287.0	290.2	288.6	1.93	1.93	1.93	2.32	2.33	2.33
T14_H_Rep2	349.6	342.7	346.2	1.94	1.94	1.94	2.33	2.33	2.33
T0_Rep3	281.7	292.6	287.2	1.93	1.93	1.93	2.34	2.34	2.34
T7_N_Rep3	215.2	207.8	211.5	1.91	1.91	1.91	2.33	2.34	2.34
T7_H_Rep3	245.1	235.4	240.3	1.92	1.91	1.92	2.34	2.31	2.33
T14_N_Rep3	287.5	287.8	287.7	1.93	1.92	1.93	2.35	2.36	2.36
T14_H_Rep3	299.2	298.7	299.0	1.93	1.93	1.93	2.34	2.32	2.33
Neg_control	254.7	274.4	264.6	1.91	1.90	1.91	2.38	2.37	2.38

Table S6: PCR2 product results

Sample	DNA Concentration (ng/µL)			Nanodrop	
	Reading 1	Reading 2	Average	A260/A280	A260/A230
T0_Rep1	32.0	31.4	31.7	1.82	2.42
T7_N_Rep1	30.8	30.0	30.4	1.83	2.42
T7_H_Rep1	32.2	31.0	31.6	1.84	2.39
T14_N_Rep1	31.0	30.2	30.6	1.82	2.07
T14_H_Rep1	32.6	32.2	32.4	1.85	2.47
T0_Rep2	29.6	30.6	30.1	1.83	2.15
T7_N_Rep2	29.8	29.8	29.8	1.78	1.95
T7_H_Rep2	31.2	30.4	30.8	1.81	1.82
T14_N_Rep2	29.0	28.0	28.5	1.84	2.38
T14_H_Rep2	30.2	30.0	30.1	1.83	2.44
T0_Rep3	34.8	31.8	33.3	1.81	2.09
T7_N_Rep3	32.4	30.2	31.3	1.81	2.34
T7_H_Rep3	32.8	31.4	32.1	1.82	2.32
T14_N_Rep3	34.2	37.4	35.8	1.83	2.38
T14_H_Rep3	34.0	31.6	32.8	1.82	1.91

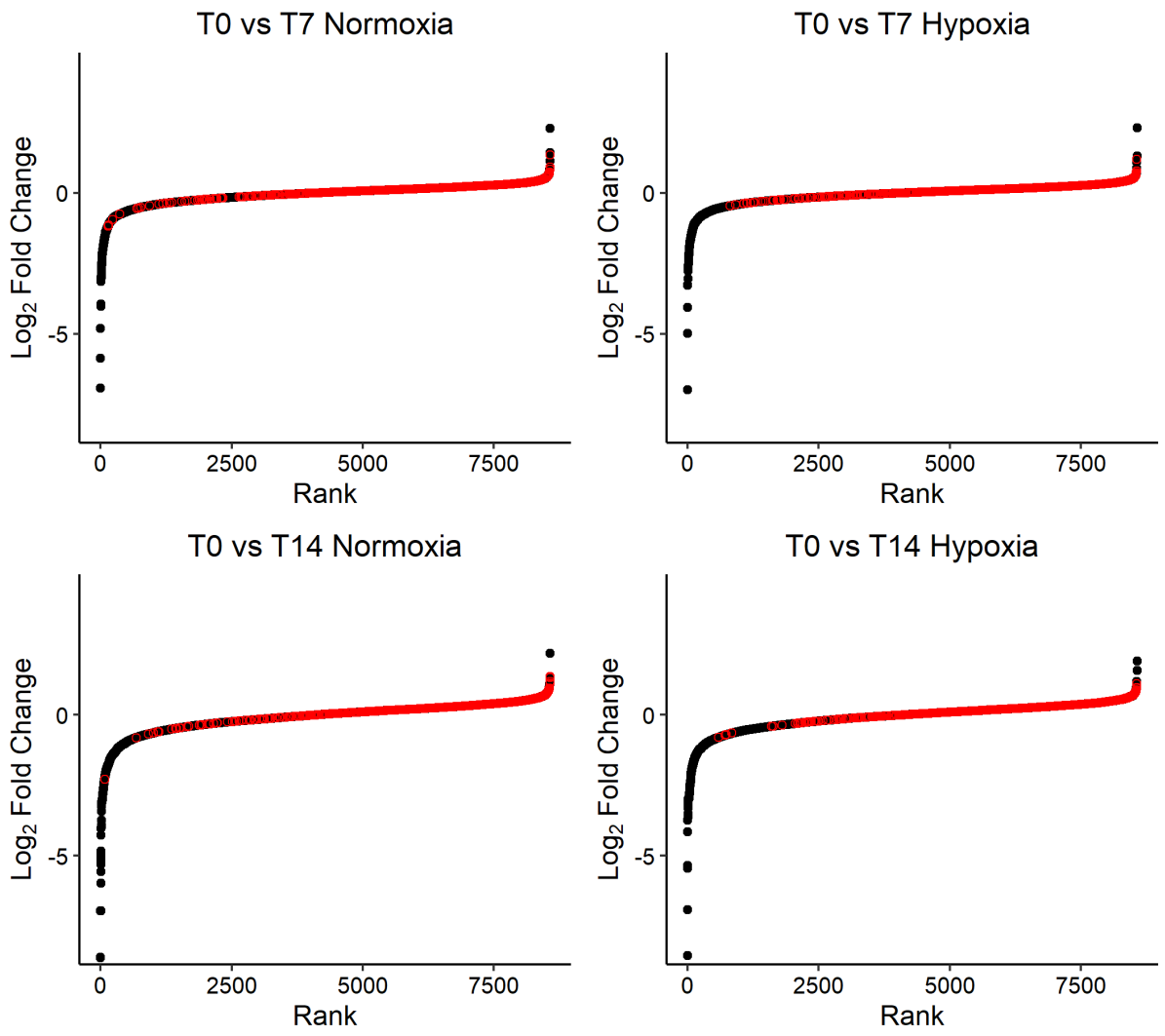

Figure S14: Log_2 fold change of non-targeting control guides. Non-targeting control guides are highlighted in red. The non-targeting guides preferentially have a positive log_2 fold change, indicating positive selection in the screen.