DEFINING THE MECHANISMS THROUGH WHICH ALDH1A3-INDUCED LONG NON-CODING RNA NRAD1 REGULATES GENE EXPRESSION IN BREAST CANCER

By

Justin Brown

Submitted in partial fulfilment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
May 2021

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Triple-negative breast cancer (TNBC) is an aggressive subtype with limited targeted treatment options and poor patient prognoses, owing in part to the subtype's enrichment in tumor-initiating cancer stem cells (CSCs) that also contribute to therapeutic resistance. Therapies targeting the CSCs within TNBCs therefore represent enticing strategies for the treatment of patients with this subtype. Accordingly, long non-coding RNAs (lncRNAs) have been shown to confer CSC attributes through induction of stemness-associated genes by interacting with regulatory proteins and microRNAs (miRNAs). The abrogation of CSC-associated lncRNAs may therefore exemplify a mechanism through which the CSCs in TNBC tumors can be targeted. Herein, we investigate the mechanisms through which CSC-enriched long non-coding RNA NRAD1 regulates gene expression in TNBC by exploring its interactions with putative protein and miRNA molecular partners. The use of NRAD1-directed proteomics assays identified the protein S100A8 as a possible NRAD1 binding protein, after which RNA immunoprecipitation was used to confirm the S100A8-NRAD1 interaction in TNBC cell lines. S100A8 is a member of the calcium-binding S100 protein family that contributes to modulation of the inflammatory response in numerous human pathologies. Importantly, S100A8 is implicated in the development and progression of several human cancers, including breast cancer. Short hairpin RNA-mediated knockdown of S100A8 in TNBC cells followed by QPCR analysis of NRAD1-regulated genes uncovered the role of S100A8 in NRAD1-mediated regulation of gene expression in TNBC. Furthermore, using miRNA microarrays and *in silico* analyses, we identified changes in miRNA expression following NRAD1 knockdown in TNBC cell lines and elucidated miRNAs with which NRAD1 may interact to regulate gene expression in TNBC. Taken together, we have constructed a foundation for the possible mechanisms through which NRAD1 regulates gene expression in TNBC.

LIST OF ABBREVIATIONS USED

HRAS	HRas proto-oncogene
BRCA1/2	Breast Cancer Type 1/2 Susceptibility Protein
HR	Homologous Recombination, Hormone Receptor
IDC	Invasive Ductal Carcinoma
ILC	Invasive Lobular Carcinoma
ER	Estrogen Receptor
PR	Progesterone Receptor
HER2	Human Epidermal Growth Factor Receptor Two
ERBB2	Erb-B2 Receptor Tyrosine Kinase 2
TNBC	Triple Negative Breast Cancer
EMT	Epithelial to Mesenchymal Transition
MET	Mesenchymal to Epithelial Transition
PARP	Poly ADP Ribose Polymerase
TIL	Tumor Infiltrating Lymphocyte
PD-1	Programmed Cell Death Protein 1
PD-L1	Programmed Death Ligand 1
PIK3CA	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha
SC	Stem Cell
CSC	Cancer Stem Cell
AML	Acute Myeloid Leukemia
SL-IC	SCID Leukemia-Initiating Cell
ABC	ATP-Binding Cassette Protein
ALDH	Aldehyde Dehydrogenase
ALDH1A1	Aldehyde Dehydrogenase 1 Family Member A1
ALDH1A2	Aldehyde Dehydrogenase 1 Family Member A2
ALDH1A3	Aldehyde Dehydrogenase 1 Family Member A3
ALDH2	Aldehyde Dehydrogenase 2 Family Member
CD24	Cluster of Differentiation 24
CD44	Cluster of Differentiation 44
CD34	Cluster of Differentiation 34
CD38	Cluster of Differentiation 38
CD133	Cluster of Differentiation 133
NF-κB	Nuclear Factor Kappa B
MAPK	Mitogen-Activated Protein Kinase
GSI	Gamma Secretase Inhibitor
NICD	Notch Intracellular Domain
GSK3-β	Glycogen Synthase Kinase 3 Beta
TCF	T-Cell Factor
LEF	Lymphoid Enhancer Binding Factor
PORCN	Porcupine O-Acyltransferase
Hh	Hedgehog
PTCH	Patched Cell Surface Receptor
GLI	Glioma-Associated Oncogene Homolog 1
SMO	Smoothened, Frizzled Class Receptor
CAR	Chimeric Antigen Receptor
ICB	Immune Checkpoint Blockade
CTLA-4	Cytotoxic T-Lymphocyte Associated Protein 4
APC	Antigen Presenting Cell
TME	Tumor Microenvironment

LOC554202	lncRNA LOC554202
LINC01133	Long Intergenic Non-Coding RNA 1133
LINC00617	Long Intergenic Non-Coding RNA 617
lncRNA-Hh	LncRNA Hedgehog
RP1-5O6.5	LncRNA RP1-5O6.5
LINC00511	Long Intergenic Non-Coding RNA 511
FEZF1-AS1	FEZ Family Zinc Finger 1 Antisense RNA 1
LINC01108 (ES1)	Long Intergenic Non-Coding RNA 1108
PART1	Prostate Androgen Regulated Transcript 1
ASO	Antisense Oligonucleotide
shRNA	Short Hairpin RNA
RNAi	RNA Interference
RISC	RNA-Induced Silencing Complex
siRNA	Small Interfering RNA
PS	Phosphorothioate
RNase	Ribonuclease
LNA	Locked Nucleic Acid
CNS	Central Nervous System
SMA	Spinal Muscular Atrophy
AML	Amyotrophic Lateral Sclerosis
SAMMSON	Survival Associated Mitochondrial Melanoma Specific Oncogenic RNA
PDX	Patient-Derived Xenograft
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
Cas9	CRISPR Associated Protein 9
UCA1	Urothelial Carcinoma Associated 1
gRNA	Guide RNA
MEST	Mesoderm Specific Transcript
SOC	Serous Ovarian Carcinoma
SOX9	SRY-Box Transcription Factor 9
MYB	MYB Proto-oncogene
ESRP1	Epithelial Splicing Regulatory Protein 1
PTC	Papillary Thyroid Cancer
MAFG	MAF bZIP transcription factor B
OSCC	Oral Squamous Cell Carcinoma
ChIRP	Chromatin Isolation by RNA Purification
MS	Mass Spectrometry
ChIRP-MS	Chromatin Isolation by RNA Purification and Mass Spectrometry
ChIP	Chromatin Immunoprecipitation
RIP	RNA Immunoprecipitation
S100A8	S100 Calcium-Binding Protein A8
S100A9	S100 Calcium-Binding Protein A9
CYP27A1	Cytochrome P450 Family 27 Subfamily A Member 1
PDZK1P1	PDZ Domain Containing 1 Interacting Protein 1
RARRES3	Retinoic Acid Receptor Responder 3
SLC15A1	Solute Carrier Family 15 Member 1
CXCL17	C-X-C Motif Chemokine Ligand 17
CYP4Z1	Cytochrome P450 Family 4 Subfamily Z Member 1
CTNNAL1	Catenin Alpha Like 1
EIF5A2	Eukaryotic Translation Initiation Factor 5A2
FOSL1	FOS Like 1
FSCN1	Fascin Actin-Bundling Protein 1

GDF15	Growth Differentiation Factor 15
GNG11	G Protein Subunit Gamma 11
ID4	Inhibitor of DNA Binding 4
IL7R	Interleukin 7 Receptor
SERPINE2	Serpin Family E Member 2
FGF2	Fibroblast Growth Factor 2
SLITRK6	SLIT And NTRK Like Family Member 6
IgG	Immunoglobulin G
HIST1H1C	Histone Cluster 1 H1 Family Member C
HIST1H1E	Histone Cluster 1 H1 Family Member E
HRNR	Hornerin

ACKNOWLEDGEMENTS

CHAPTER 1 – INTRODUCTION

1.1 Cancer

Cancer refers to an assortment of closely related genetic diseases, where dysregulated cellular proliferation promotes the formation of heterogenous tumors. These tumors, which can arise in numerous anatomical sites, can release tumorigenic cells capable of establishing novel tumors at distant locations. This process is known as metastasis. Metastatic spread constitutes the greatest risk of death by interfering with critical organs and biological processes, representing an urgent subject for cancer research.[1] Since 1990, cancer deaths in the United States and Canada have declined each year, owing to earlier detection and intervention, and advances in treatment[2]. Current cancer treatments often involve targeted therapies that act on cancer-specific genetic aberrations or proteins involved in different aspects of cancer formation and progression, often in combination with systemic chemotherapies, which represent pillars in cancer remediation[3]. The term chemotherapy was coined by Paul Ehrlich, a German chemist who described it as the use of chemicals to treat disease. While surgery and radiotherapies overshadowed the field of cancer therapy until the 1960s, response to treatment flattened due to micrometastases, small groups of metastatic cells[3,4]. During this time, combination chemotherapies demonstrated unprecedented clinical success in patients with advanced cancers. Chemotherapeutic agents were administered with surgery and/or radiation, reducing micrometastases and giving rise to the field of adjuvant chemotherapy[3]. Recent scientific attention has been directed towards immunotherapies that take advantage of the body's immune system to fight cancer by activating or suppressing immune response in a context-dependent manner[5]. Despite the compelling progress in the war on cancer, it continues to serve as the leading cause of death in Canada, representing 30% of national deaths[6]. Thus, significant research efforts are still required to gain understanding and elucidate novel treatment routes for this complex disease.

Cancers are diseases with a fundamentally genetic basis, whereby dynamic genomic alterations underly a multistep process driving the progression of normal cells into tumorigenic and malignant variants[7]. Several human cancers with an age-dependant incidence illustrate between four and seven stochastic rate-limiting steps in tumorigenesis[8]. The presence of multiple rate-limiting steps in

tumorigenesis has also been demonstrated in transgenic mouse models[9]. These findings have been supported by the observation that tumor cell genomes are altered at multiple sites through diverse genomic aberrations[10]. Together, these findings suggest that tumor formation resembles Darwinian evolution, where successive genetic alterations impart selective growth advantages that lead to the transformation of normal cells to those with a neoplastic phenotype[7]. The landmark paper, *The Hallmarks of Cancer* by Douglas Hanahan and Robert Weinberg (2000)[7] details six biological abilities shared by most or all tumors that are amassed during multi-step tumorigenesis. These hallmark capabilities include sustained proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis, and activating invasion and metastasis. Importantly, each hallmark embodies the successful surmounting of an anti-cancer defense system.

Advances in cancer research led to the characterization of novel enabling characteristics and emerging cancer hallmarks[11]. The emerging hallmarks included reprogramming of cellular metabolism and evasion of immunological destruction. Most notably, the two enabling characteristics, genome instability and mutation, and tumor-promoting inflammation promote the acquisition of both the original and emerging cancer hallmarks[11]. Briefly, genome instability confers cancer cells with genetic aberrations that facilitate tumor progression. Further, while cells of the innate immune system are known to be involved in eradicating infection and modulating wound repair, their intimate involvement in tumor formation is now widely accepted[11]. Cancers are highly heterogeneous, not only in tumor composition, but also in patient response to therapy. For example, two patients with the same cancer may respond differently to treatment and have different clinical prognoses. The development of treatment strategies is challenging as tumorigenesis cannot be attributed to a single mutation[12]. Importantly, human cells employ numerous defenses against cancer-promoting gene mutations. Invasive cancers are established only when multiple genes are defective. These mutations may be passed on through the germline, increasing cancer susceptibility, or may emerge via somatic mutation[12]. The first evidence for a somatic mutation in a human cancer gene was the conversion of amino-acid 12 of HRAS from glycine to valine in bladder carcinoma cells[13]. Since this discovery, a myriad of cancer genes have been identified and the proteins

2

they encode modulate cell proliferation, differentiation, death, and DNA-repair mechanisms[12]. Cancer-promoting mutations often arise in three gene types, proto-oncogenes, tumor suppressors, and stability genes[14].

Gain-of-function mutations in oncogenes contributes to cancer formation. Proteins encoded by oncogenes often regulate cell proliferation and/or apoptosis and can be classified into six groups including transcription factors, chromatin remodelers, growth factors, growth factor receptors, signal transducers, and regulators of apoptosis[15]. Oncogenes are mutated such that they are constitutively expressed or active in scenarios when wild-type genes are not. They can become active from chromosomal translocations, gene amplifications or point mutations, which provide a growth or survival advantage to cells carrying these aberrations[14,15].

To limit dysregulated cell growth and proliferation, effective genomic programs are in place, tumor suppressors, that induce programmed cell death, cell cycle arrest, or senescence when proliferation becomes aberrant[16]. The incredible effectiveness of tumor suppressors is observed in the low rate of cancer development throughout a human lifetime, despite the immense number of potentially affected cells, each carrying a myriad of susceptible oncogenes[16]. For example, elevated expression of the RAS oncogene in the presence of non-mutated tumor suppressors (p53 and p16) induced cell cycle arrest. However, loss of function mutations in these tumor suppressors allows RAS-mediated oncogenic activity and subsequent cancer progression[17]. Therefore, oncogene activation must be paired with tumor suppressor gene inactivation to support tumorigenesis. In contrast to oncogenes, tumor suppressors are associated with loss of function mutations during tumorigenesis. Tumor suppressor inactivation results from missense mutations in functional residues, mutations leading to protein truncation, indels, or epigenetic silencing[14].

Stability genes or caretaker genes maintain genomic stability by detecting DNA damage and repairing damaged DNA. These genes are often involved in processes such as mismatch repair, base-excision repair, and nucleotide-excision repair. Stability genes maintain genetic alterations at a low level; therefore, their inactivation leads to elevated mutations in other genes[18]. For example, the *BRCA1* stability

gene is involved in the repair of double strand breaks through homologous recombination (HR). Cells with mutated *BRCA1* are deficient in this repair program, which can promote expansive DNA mutation and cancer development[19]. Germline *BRCA1* mutations predispose women to both breast and ovarian cancers and hold important clinical implications[19].

Taken together, cancers represent a complex group of diverse pathologies that arise from a collection of stepwise mutations in critical proliferative regulators, generating neoplastic cells with selective growth advantages that can produce heterogeneous tumors. Importantly, mutations do not account for the entirety of selective advantages contributing to cancer formation. For example, cancer cells may alter normal cellular pathways to facilitate cancer formation and progression.

1.2 Breast Cancer

Breast cancer is currently the most common cancer among Canadian women, affecting 1 in 8 women in their lifetime. It is also the second leading cause of cancer-related death in Canadian women, and it is estimated that 1 in 33 patients will die from the disease[6]. Fortunately, patient outcomes have improved thanks to better screening, earlier diagnoses, and targeted therapies. However, due to the immense heterogeneity of the disease, not all patients respond to therapy, prompting novel research efforts. Breast cancer is a highly heterogeneous disease with several unique subtypes, each with respective gene expression portraits and clinical implications, complicating treatment efforts.

Breast cancers can be characterized based on clinical features, tumor gene expression, and histologic subtype[20]. Most breast tumors are derived from mammary ductal epithelium, specifically the terminal duct lobular unit. The most common histologic type of invasive breast cancer, invasive ductal carcinoma (IDC; 50-75% of patients), arises from milk ducts, while the second most common type, invasive lobular carcinoma (ILC; 5-15% of patients), arises from milk glands (lobules). Mixed ductal/lobular carcinomas and rare histologies constitute the remaining patients[21]. Preceding each of these infiltrating cancers, are carcinomas *in situ,* pre-invasive lesions resulting from clonal proliferation of

malignant cells that have not infiltrated through the epithelial basement membrane and into nearby breast tissue. These lesions, however, may become invasive and spread to additional anatomical regions through additional mutations and molecular alterations[22].

Breast cancer is categorized clinically based on expression of estrogen receptor (ER), progesterone receptor (PR), and/or human epidermal growth factor receptor 2 (HER2/ERBB2)[23]. Hormone receptor status in breast tumors holds important clinical ramifications. For example, patients with non-metastatic breast tumors that express either ER or PR in at least 1% of tumor cells are considered hormone receptor (HR) positive[24]. These patients receive endocrine therapies, sometimes in combination with chemotherapy. These tumors, which are typically HER2 negative, represent 70% of breast cancer patients. Endocrine agents that antagonize ER signaling are the main therapies for ER positive and PR positive breast cancers[25]. For example, HR positive tumors are often treated with tamoxifen, a selective ER modulator that binds competitively to estrogen receptors, preventing their interaction with estrogen, thereby silencing the expression of estrogen-regulated genes and reducing estrogen-mediated tumor growth[25–27]. In addition, aromatase inhibitors can be used as a treatment for postmenopausal women with HR-positive breast cancers, who form the majority of breast cancer patients[28]. Aromatase inhibitors diminish plasma estrogen levels via inhibition of aromatase, the enzyme that catalyzes the biosynthesis of estrogens from androgen substrates[28]. Like tamoxifen[29,30], aromatase inhibitors such as Anastrozole can prevent breast cancer recurrence and the development of new tumors in high-risk patients[31,32]. Notably, aromatase inhibitors are associated with reduced side effects, relative to tamoxifen, which can increase the risk of uterine cancers[33,34].

ERBB2-positive patients receive monoclonal antibodies (trastuzumab or pertuzumab) or small-molecule tyrosine kinase inhibitors (lapatinib or neratinib) against ERBB2, a transmembrane receptor tyrosine kinase that is overexpressed or amplified in 20% of breast cancers and a poor prognostic indicator[35]. Importantly, chemotherapy is often administered in combination with ERBB2-targeted treatments and endocrine therapy (when tumors are concurrently HR-positive)[36]. Recently, the FDA approved the antibody-drug conjugate, fam-trastuzumab deruxtecan-nxki, an anti-ERBB2 antibody with

an associated cytotoxic topoisomerase inhibitor, for patients with non-resectable or metastatic ERBB2-positive breast cancer[37].

While breast tumors can be characterized based on hormone receptor status, tumor subtyping based on gene expression profiles may provide more valuable prognostic information[38]. Breast tumors demonstrate significant heterogeneity in gene expression portraits, which can predict disease outcome and response to therapy[39]. Subtyping based on gene expression revealed five unique subtypes: basal-like, luminal A, luminal B, HER2-enriched, and claudin-low[38,40,41]. Luminal A and luminal B tumors are typically ER-positive and may respond to ER-directed therapies such as tamoxifen. Similarly, HER2-enriched tumors highly express ERBB2, and can benefit from ERBB2-targeted treatments. Basal-like breast cancers are characterized by a gene expression portrait reflective of that of normal breast basal-myoepithelia, where 75% fail to express any of the three clinically targetable receptors and are denoted as "triple-negative"[38]. Most claudin-low tumors are also triple-negative and relative to basal-like breast cancers, are more enriched in EMT attributes, immune responses, and cancer stem cells (CSC)[42].

Triple-negative breast cancers (TNBC), which represent 15-20% of breast tumors, are a particularly aggressive subtype characterized by the lack of therapeutic targets and adverse patient outlook[43–45]. Triple-negative and basal-like breast cancers are most common in young black and Hispanic women[46,47]. These tumors are of high histologic grade and do not express ER, PR, or ERBB2. Thus, TNBC patients do not respond to endocrine or anti-ERBB2 therapies and are treated primarily with chemotherapy[48,49]. TNBCs are associated with early and aggressive patterns of metastasis and recurrence[50]. Within five years of their diagnosis, women with TNBC are more likely to experience visceral metastasis relative to women with other breast cancer subtypes[51]. Patients with TNBCs are known to experience metastasis to several anatomical locations including the bone, lungs, liver, and brain. However, metastasis to the lungs and brain occurs at an elevated rate among patients with TNBCs[45,52].

While TNBCs exhibit adverse patient prognoses, many patients respond well to chemotherapy and demonstrate enhanced responsiveness to adjuvant therapies relative to other subtypes[44,53], a phenomenon referred to as the triple-negative paradox[48]. Importantly, most breast cancers with *BRCA1/2*

mutations are triple-negative. Thus, some TNBCs exhibit a deficiency in homologous recombination and the associated repair of double strand DNA breaks. Fortunately, patients with *BRCA1*-mutated cancers may respond to PARP inhibitors such as Olaparib, which have demonstrated clinical success in TNBC and ovarian cancers[54–56]. Moreover, TNBCs exhibit elevated immune cell infiltrate and an active immunological microenvironment[49]. Tumor-infiltrating lymphocytes (TILs) have been associated with improved prognoses and response to adjuvant chemotherapy in TNBC[57,58]. TILs are positively correlated with PD-1/PD-L1 expression in TNBC, demonstrating the susceptibility of this subtype to anti-PD-1/PD-L1 immunotherapies[59]. For example, a phase III clinical trial demonstrated that the anti-PD-L1 antibody atezolizumab plus nanoparticle albumin-bound paclitaxel increased progression-free survival in PD-L1-positive patients with metastatic TNBC[60]. Intriguingly, some TNBCs express androgen receptors, which may represent targets for endocrine therapies for patients with TNBC[61]. Androgen receptor antagonists have demonstrated success in clinical trials investigating their effectiveness in TNBCs[62–64]. For example, *PIK3CA* mutations in androgen receptor positive TNBC impart sensitivity to PI3K inhibitor and androgen receptor inhibitor combination therapy[63].

The aggressiveness of TNBC has been attributed to the enrichment of CSC populations within this subtype[65–72]. Within tumors, CSCs are the most tumorigenic cells, initiating new tumors with high efficiency[73]. Most concerning in terms of mitigating the risk of metastasis and recurrence in the treatment of TNBC is the resistance of CSCs to cancer therapies[74]. Thus, novel therapies that also target CSCs may better reduce the risk of relapse and improve the outcomes of TNBC patients.

1.3 Cancer Stem Cells

Historically, tumors were thought to be comprised of clonal cells with enhanced proliferative ability but equipotent tumor-forming potential. It is now widely accepted that tumors are highly heterogeneous and are comprised of diverse cells with differential tumor-initiating capacities[75,76]. Within tumors are cancer cells and non-cancer cells (i.e., resident fibroblasts, cancer associated fibroblasts and

immune cells). Of the cancer cells, CSCs are the most tumorigenic, initiating tumors with heightened efficiency relative to their non-CSC counterparts. CSCs are defined by key characteristics – enhanced tumorigenicity and the capacity for self-renewal and differentiation[73,77,78]. Self-renewal is responsible for maintenance of the CSC population, while differentiation permits changes into non-CSCs. Thus, CSCs can recapitulate tumors with both CSC and non-CSC populations[77]. Furthermore, CSCs are intimately linked to metastasis[79,80]. Evidence for the existence of a subset of cancer initiating cells (i.e. CSCs) was first demonstrated by Bonnet and Dick (1997) in human acute myeloid leukemia (AML)[77]. A subset of cells exclusively expressing the cell surface phenotype CD34[High]/CD38[Low], which comprised only 0.01-1% of the tumor population could induce human AML upon transplantation into NOD/SCID mice. These cells, termed SCID leukemia-initiating cells (SL-ICs) possessed characteristics reflective of CSCs, whereby upon transplantation into NOD/SCID mice, cells could proliferate and differentiate to re-produce the disease of the donor and could self-renew, supporting AML-recapitulation in secondary recipient mice.

In addition to their tumor initiating ability and metastatic capacity, the role of CSCs as barriers to successful cancer remediation is exacerbated by their resistance to chemotherapies and radiotherapies[81,82]. Thus, CSCs play a critical role in cancer recurrence. This is supported by several lines of evidence that illustrate enrichment of CSC populations following chemotherapy and Radiotherapy[81,83–88]. Specifically, CSCs are characterized by elevated expression of drug transporters, enhanced DNA repair mechanisms, immune evasion and adaptive metabolic plasticity, which contribute to their tumor-forming and resistance abilities[89–95]. For example, ABC (ATP-binding cassette) transport proteins (ABCB1, ABCC1, and ABCG2) upregulated in CSCs, promote chemotherapeutic drug efflux (i.e. doxorubicin and paclitaxel) and contribute to multi-drug resistance among cancers[96–103]. Moreover, detoxifying enzymes such as aldehyde dehydrogenase (ALDH) are upregulated in CSCs and participate in drug inactivation[74,104–107].

CSCs are also characterized by metabolic plasticity and can be distinguished from bulk tumor cells based on their dependence on specific metabolic pathways[108,109]. Bulk tumor cells exhibit metabolic reprogramming to facilitate the use of aerobic glycolysis as their primary energy source in lieu oxidative

phosphorylation, which is more efficient in ATP generation. Termed the Warburg effect, this process supports the elevated energetic needs of highly proliferative cells[110]. Conversely, as CSCs represent a distinct subpopulation of cells with enhanced capacity for tumorigenicity, metastasis, and proliferation, they also differ in their energetic and metabolic requirements[95].

Treatment of tumors with conventional therapies fails to target resistant CSCs. Conventional therapies efficiently target bulk tumor cells (non-CSCs), but leave behind CSCs, which contribute to disease relapse and metastasis[111,112]. Therapies that target both CSCs and bulk tumor cells would thus promote tumor reduction and lower the risk of disease relapse[74,113–116].

CSCs are often identified by flow cytometry, sorting cells based on their cancer-specific cell surface markers or by elevated ALDH expression assessed by the Aldefluor assay, a flow cytometry-based method that provides a readout of ALDH enzyme activity[105,106,117]. Importantly, Aldefluor positive cell populations (ALDH[high]) from several cancers can efficiently initiate tumors *in vivo* at low concentrations[118–120]. Upon sorting of CSC populations, putative CSCs are implanted into mice at low concentrations to confirm their tumor-initiating ability. Tumors are subsequently harvested and re-passaged to observe tumor formation in secondary recipients, providing support for self-renewal and differentiation. Furthermore, CSCs can be identified based on their formation of spheroids in suspension cultures and generation of three-dimensional organoids[121–128].

Some controversy exists surrounding the mechanism from which CSCs are derived. It has been suggested that CSCs arise from normal tissue stem cells (SCs) that experience transformative oncogenic somatic mutations. This model suggested that a genetically altered normal SC sub-population would generate mutated SC progeny, which would undergo successive heritable genetic alterations, ultimately giving rise to the final SC population (CSCs), in the absence of non-SC contribution[129]. An updated model was proposed to account for the biological and mathematical shortcomings of the initial model. This model accounted for the emerging evidence that supported a higher degree of plasticity in hierarchically composed cell populations than was initially conceived[130]. The current model suggests that initial heritable alterations (genomic mutations or epigenetic alterations) are sustained by non-SCs,

9

transit-amplifying cells (progenitor cells) in particular, which re-enter the SC pool by de-differentiation before the process is repeated. Intermediate SC products of de-differentiation may generate unique populations of differentiated cells downstream, contributing to tumor heterogeneity[130]. Furthermore, several studies have shown that epithelial to mesenchymal transition (EMT) promotes the acquisition of CSC-like characteristics in different human malignancies[131–133]. EMT refers to the process through which epithelial cells phenotypically transform into a mesenchymal-like state. During this transformation, epithelial markers are lost while mesenchymal markers become gradually more expressed. The transition to a mesenchymal state via EMT is linked to cell invasiveness and thus, EMT is commonly associated with migration and metastasis[134].

1.4 Breast Cancer Stem Cells

Breast CSCs are commonly identified by the cell surface markers CD44 and CD24, specifically the CD44^{high}/CD24^{low} phenotype, and elevated ALDH enzyme activity (Aldefluor^{high}). Breast CSCs were the first to be identified in solid tumors[73]. Al-Hajj and colleagues (2003)[73] identified tumor-initiating breast CSCs from non-CSCs based on the cell surface marker phenotype CD44^{high}/CD24^{low}. Notably, as few as 100 cells of this phenotype were capable of initiating tumors in mice, while tens of thousands of non-CSCs did not form tumors. Like the CD34^{High}/CD38^{Low} SL-ICs from AML, CD44^{high}/CD24^{low} breast CSCs could recapitulate the original tumor heterogeneity after serial passaging. Similarly, Ginestier and colleagues (2007)[106] showed that elevated ALDH activity defines the most tumorigenic cells in xenotransplantation models of breast cancer. Specifically, cells with elevated ALDH activity were capable of self-renewal and recapitulation of the parental tumor, characteristics indicative of CSCs. These CSC markers hold clinical value, as they can predict clinical outcomes. For example, ALDH1 is a marker of poor prognosis in breast cancer[135] and Aldefluor^{high} cells mediate metastasis and recurrence[136–138]. Similarly, breast cancer cells expressing the CD44^{high}/CD24^{low} phenotype may help predict the presence of lymph node metastases[139].

The stem cell marker ALDH is a detoxifying enzyme that oxidizes intracellular aldehydes and contributes to stem cell differentiation via metabolism of retinal to retinoic acid[140]. The ALDH superfamily is comprised of 19 different isoforms with unique functions. While ALDH1A1 was thought to be the primary contributor to ALDH activity of CSCs, the focus was recently shifted to another ALDH isoform, ALDH1A3[117]. CSC identification based on ALDH1A1 was not correlated with metastasis, raising the question of the main isoform in Aldefluor positivity among breast CSCs. Marcato and colleagues (2011) showed that ALDH1A3, and not ALDH1A1, correlates best with the ALDH activity of patient breast tumor-derived CSCs[117]. Knockdown of ALDH1A3 reduced Aldefluor[high] breast cancer cells, and ALDH1A3 expression in patient tumors significantly correlated with tumor grade, metastasis, and cancer stage, marking it as a critical breast CSC marker and potentially valuable prognostic indicator[117]. Later, ALDH1A3 was shown to influence breast cancer progression by inducing retinoic acid signaling in TNBC. Retinoic acid-inducible gene expression was shown to depend on ALDH1A3[141].

While CSCs are important tumor-initiating cells and are therefore implicated in the development of all breast tumor subtypes, they are particularly enriched in TNBCs. Several studies have shown that the enrichment of CSCs based on the expression of the CD44[high]/CD24[low] and/or ALDH[high] phenotypes within this subtype is associated with its adverse clinical prognosis[65–72]. Specifically, the enrichment of CSCs within TNBCs may contribute to its high rates of metastasis and recurrence. Thus, treatments that target CSCs may reduce the risk of relapse among patients with TNBC.

The CD44[high]/CD24[low] and ALDH[high] CSC phenotypes describe two distinct populations of breast CSCs[67]. Ginestier and colleagues[106] (2007) demonstrated that breast cancer cells with a CD44[high]/CD24[low]/ALDH[low] phenotype were less tumorigenic relative to CD44[high]/CD24[low]/ALDH[high] cell populations. Similarly, CD44[high]/CD24[low]/ALDH[high] breast cancer cells demonstrate increased metastasis and chemo/radio-resistance relative to their CD44[high]/CD24[low]/ALDH[low] counterparts[142]. Thus, an important link exists between these two breast CSC populations.

1.5 Towards clinical success: Therapeutic targeting of breast CSCs

Breast CSCs are a major barrier to the effective remediation of breast cancer. Thus, treatment strategies that involve the concurrent targeting of breast CSCs and non-CSCs will be the most effective in mitigating the risk of metastasis and recurrence among breast cancer patients. However, clinical targeting of breast CSCs is not trivial. Several attempts to inhibit breast CSCs have recently focused on highly conserved signaling pathways that regulate aspects of tissue stem cells and CSCs and are deregulated and highly active in CSCs. These pathways, (Notch, Wnt, and Hedgehog), regulate stem cell self-renewal and differentiation, and play important roles in embryonic development[143–145]. Importantly, these pathways communicate with other signaling pathways such as NF-κB, MAPK, and PI3K, which may also represent targets for the eradication of CSCs[146].

Inhibition of the Notch pathway in breast CSCs has been the focus of several pre-clinical studies[143,147–150]. Specifically, γ-secretase inhibitors (GSIs) represent a common mechanism of interest in Notch inhibition. GSIs target γ-secretase, which cleaves Notch receptors, liberating the Notch intracellular domain (NICD) transcriptional activator, which translocates to the nucleus and initiates Notch activity[151]. GSIs have demonstrated pre-clinical success, particularly in combination with chemotherapies. For example, GSIs in combination with trastuzumab have led to tumor regression and elimination of recurrence by targeting CSCs in mouse models of ERBB2+ breast cancers[152]. Similarly, GSIs in combination with trastuzumab completely abrogated tumor recurrence by targeting CSCs in ERBB2+ breast cancers, while trastuzumab alone was associated with a 50% recurrence rate[153]. Thus, Notch pathway inhibitors should be investigated with traditional chemotherapeutics. For example, Notch pathway inhibitors should be administered with endocrine therapies in HR-positive breast cancers, anti-HER2 treatments in ERBB2-enriched breast cancers, and MET inhibitors and taxanes among patients with TNBC[154]. To this end, a phase Ib clinical trial investigating the GSI, RO4929097 in combination with exemestane demonstrated some success in patients with metastatic ER-positive breast cancer[155]. The GSI, MK-0752 in combination with docetaxel has been shown to decrease CD44[high]/CD24[low] and Aldefluor[high] populations of breast CSCs in breast tumor xenografts in mouse models and from tumors of

patients undergoing serial biopsies[156]. However, this treatment may be associated with gastrointestinal toxicity. Currently, no GSIs have been approved for clinical use.

The canonical Wnt/β-catenin pathway is highly active in TNBCs where it enhances cell proliferation and stem cell maintenance[157,158]. In brief, extracellular Wnt ligands bind to membrane-bound frizzled receptors and low-density lipoprotein receptor-related protein, promoting phosphorylation of disheveled, which inactivates GSK3-β, a negative regulator of β-catenin. Accumulation of β-catenin in the cytoplasm followed by its nuclear translocation allows it to form a complex with TCF/LEF transcription factors and activate downstream genes[159]. Jang and colleagues (2015)[160] used a murine model of metastatic breast cancer to show that inhibition of the Wnt/β-catenin pathway attenuates metastasis via suppression of CSCs, illustrating the importance of this pathway in CSC activity. Inhibitors of the Wnt pathway are also under clinical investigation in the treatment of breast cancer. For example, LGK-974 is an inhibitor of porcupine (PORCN), an acyltransferase necessary for the function/secretion of Wnt ligands[161]. De-regulated Wnt ligands promote aberrant Wnt/β-catenin signaling, mediating breast cancer progression[157,162]. Pre-clinical experiments have shown that LGK-974 exhibits strong dose-dependent effects in murine models of Wnt-dependent breast tumors, reducing tumor growth in *vivo*[161]. LGK-974 can reduce tumor growth and induce tumor regression as a single agent and in combination with taxol in human primary breast tumor models[163]. Notably, LGK-974 is currently under investigation in clinical trials for TNBC (NCT01351103)[164]. In addition, Vantictumab (OMP-18R5), an anti-frizzled receptor monoclonal antibody is also under investigation in combination with paclitaxel in patients with locally recurrent or metastatic breast cancer (NCT01973309)[165].

Hedgehog (Hh) signaling is a key player in breast cancer progression, particularly in HR-positive breast cancers and TNBCs[166]. Hh signaling has been implicated in tamoxifen resistance in HR-positive breast cancers and chemotherapy resistance in TNBCs via CSC maintenance[144,167–170]. Hh signaling is highly expressed in breast CSCs relative to non-CSCs where it contributes to CSC activity[144,169–171]. For example, Koike et al. (2017)[170] demonstrated that the Hh inhibitor, GANT61 decreased cell growth and diminished the CSC population in TNBC cells. Further, GANT61 enhanced the anti-cell growth activity

of Paclitaxel. Briefly, Hh ligands bind the Patched (PTCH) cell surface receptor, inducing accumulation of Smoothened (SMO), which activates GLI transcriptional regulators that activate transcription of Hh target genes[172]. Vismodegib an inhibitor of SMO, was the first FDA-approved anti-Hh treatment, approved for treatment of patients with advanced and metastatic basal-cell carcinoma[173]. Vismodegib is currently under investigation in TNBC in combination with Paclitaxel, Epirubicin, and cyclophosphamide (NCT02694224)[174].

In summary, crosstalk between the Notch, Wnt and Hedgehog pathways and with other pathways significantly complicates treatment strategies. On the other hand, this crosstalk may allow the inhibition of multiple pathways by targeting only one. Furthermore, combination approaches targeting multiple pathways simultaneously may prove effective against CSCs that escape targeting.

The interactions between CSCs and the immune system may allow targeting breast CSCs with immunotherapies, which activate endogenous anti-tumor immune responses[175–178]. Attractive mechanisms of immunotherapy targeting CSCs include chimeric antigen receptor (CAR)-engineered T cells and immune checkpoint blockade (ICB)[179,180]. Moreover, therapies targeting breast CSC-specific antigens may prove clinically valuable[181–183]. ICB has shown the most promise, with anti-CTLA4 and anti PD-1/PD-L1 therapies demonstrating unprecedented therapeutic success. In accordance, ICB-based therapies have been FDA-approved for treatment of several cancers, resulting in long-lasting therapeutic responses[184,185]. ICB employs antibodies to block immune regulatory checkpoints, removing signals that inhibit T cell activation, promoting an anti-tumor immune response by tumor-reactive T cells. For example, CTLA4 (cytotoxic T lymphocyte-associated protein 4) negatively regulates T cell co-stimulation by antigen presenting cells (APCs), required for induction of tumor-reactive T cells in the tumor-draining lymph node[186–188]. Ipilimumab, a clinically approved anti-CTLA4 antibody blocks CTLA4 activation and alleviates its negative influence on T cell activation[189,190]. This therapy has been particularly effective in patients with metastatic melanoma[191,192].

ICB targeting the PD-1/PD-L1 axis has been a major focus more recently. In normal cells, the PD-1 (programmed cell death-1) receptor is expressed on the surface of activated T cells. The PD-L1

ligand is expressed by dendritic cells or macrophages. This immune checkpoint axis ensures appropriate activation of T cells to prevent autoimmunity. In cancer, immune checkpoints allow cancer cells to evade detection and elimination by the immune system[193,194]. Elevated expression of PD-L1 by tumor cells and activated immune cells within the TME is associated with aggressive disease and poor patient prognosis[195–197]. Accordingly, PD-L1, induced by cytokines produced in the TME such as IFN-γ, is highly expressed on tumor cells in the tumor microenvironment (TME) and interacts with PD-1 receptors on CD8+ cytotoxic T cells, leading to T cell exhaustion and suppression of anti-tumor activity. Anti PD-1/PD-L1 therapies prevent the formation of this immune checkpoint and allow T cell-mediated targeting of tumor cells[193]. Importantly, several lines of evidence have illustrated that CSCs, including breast CSCs, express higher levels of PD-L1 relative to non-CSCs, potentially rendering them susceptible to ICB[111,198–201]. For example, Gupta and colleagues (2016) showed that PD-L1 knockdown in murine B16 melanoma cells significantly reduced CSC populations[202].

While Nivolumab and pembrolizumab were the first FDA-approved anti-PD-1 agents for melanoma and non small-cell lung cancer (NSCLC)[203,204], several clinical trials have been conducted investigating immune checkpoint inhibitors in breast cancer targeting CTLA-4, PD-1, and PD-L1. These therapies have been investigated as single agents or in combination with trastuzumab in ERBB2+ breast cancers, and chemotherapies in TNBCs. In 2019, the anti-PD-L1 antibody, Atezolizumab (ATZ) was FDA approved for use in combination with nanoparticle albumin-bound (nab) paclitaxel for patients with unresectable locally advanced or metastatic PD-L1+ TNBC[205].

Moreover, dendritic cell (DC) vaccines can induce specific anti-CSC immune responses[182,206–208]. Specifically, mice vaccinated with DCs loaded with lysates from NSCLC-derived CSCs exhibited enhanced anti-tumor immunity relative to those vaccinated with bulk tumor cell-loaded DCs, demonstrating that CSCs may represent a better source of antigens[207]. The eradication of breast CSCs via DC vaccination requires the identification of breast CSC-specific antigens or isolation of CSC populations from patient tumors to generate whole-cell lysates[182]. Currently, DC vaccines targeting CSCs are under investigation in clinical trials for treatment of several cancers, including breast cancer. For

example, the trial conducted by Ning et al. (2012) investigated the safety of immunization with Aldefluor[high] breast CSC-loaded DCs (NCT02063893)[207]. Another phase I trial investigating the use of a multiantigen DNA plasmid-based vaccine to target immunogenic proteins expressed in breast CSCs is recruiting (NCT02157051)[209].

CAR T cells are another way to target CSC-specific antigens, where CAR T cells can recognize specific antigens expressed on the surface of CSCs[180,210]. CAR T cells targeting specific CSC antigens are under investigation in pre-clinical trials targeting CSC-specific antigens such as CD133 or NKG2D in glioblastoma[211,212]. ALDH and CD44 have been proposed as attractive targets for CAR T cell therapy, potentially benefitting patients with breast cancer[178]. Byrd and colleagues (2018) recently demonstrated that CAR T cells targeting the TNBC-associated antigen TEM8 could also target breast CSCs, reducing mammosphere formation[213]. Further, Seitz and colleagues (2020) recently showed that CAR T cells targeting the breast CSC antigen, GD2 could suppress tumor progression and prevent lung metastasis *in vivo* via elimination of breast CSCs[214].

The aforementioned functional breast CSC markers CD44 and ALDH represent attractive targets for CSC inhibition and have been the assessed in several pre-clinical studies. CD44 is a transmembrane glycoprotein that modulates CSC characteristics such as cell survival, migration, invasion, angiogenesis, and metastasis[215,216]. Pre-clinical studies have evaluated inhibition of CD44 as a treatment for breast cancer, demonstrating that CD44 antagonization decreases breast cancer progression[217,218]. Specifically, combination treatment with nanoparticles containing Salinomycin (CSC inhibitor targeting CD44) and nanoparticles containing paclitaxel to target both CSCs and cancer cells provided enhanced cytotoxic effects against both CD44[high] CSCs and bulk cancer cells, demonstrating that combination therapies involving CSC inhibitors and traditional chemotherapies may represent promising treatment strategies[218].

In addition, other studies have assessed inhibition of the functional CSC marker, ALDH in breast cancer. The ALDH inhibitor, Disulfiram, which is normally used to treat chronic alcohol consumption, was used to inhibit ALDH1A3, reducing CSC populations and reversing chemoresistance in paclitaxel resistant TNBC cells[219]. Moreover, Yip et al. (2011) showed that Disulfiram/copper treatment inhibited

mammosphere formation among CD44[high]CD24[low]ALDH[high] breast CSCs and enhanced Paclitaxel cytotoxicity in breast cancer cell lines by inducing ROS production and inhibiting NF-κB[220]. Intriguingly, a recent study showed that Disulfiram/copper could induce immunogenic cell death (ICD) in breast CSCs and non-breast CSCs and sensitized ionizing radiation (IR)-resistant breast CSCs to IR-induced ICD to a level comparable to non-breast CSCs[221]. Citral has also been shown to be a specific inhibitor of ALDH1A3, decreasing Aldefluor activity of TNBC cells and ALDH1A3-mediated breast tumor growth, colony formation, and gene expression[222]. Similarly, the ALDH inhibitor, Diethylaminobenzaldehyde (DEAB) has been shown to suppress breast tumor growth *in vivo*, however, demonstrates limited specificity, targeting ALDH1A1, ALDH1A2, and ALDH2[223,224].

Accumulating evidence has suggested that non-coding RNAs (ncRNAs) may serve as functional regulators of breast CSCs. Thus, inhibition of these ncRNAs may eradicate CSCs within breast tumors and improve patient outcomes. Specifically, a recent focus has been placed on long non-coding RNAs (lncRNAs) for their roles in CSC dynamics and breast cancer progression.

1.6 Long non-coding RNAs: Unravelling the code

The historical observation that the size of the genome is poorly correlated with organism size and developmental complexity was supported by the finding that the majority of the genome does not encode protein-coding genes. To this end, the 2001 human genome project demonstrated that only 1-2% of the human genome encodes proteins, while the rest is non-coding[225]. Apart from the approximately 20,000 protein-coding genes, the remaining non-coding genomic regions were known as "junk DNA" or transcriptional noise as a result of their enrichment in pseudogenes, transposons, and simple repeats[226]. Recent genomic technologies allowing deep genomic and transcriptomic analyses have demonstrated that as much as 85% of the human genome is transcribed, with variable ranges and levels of expression for different RNA types[227]. Specifically, these analyses have shown that as a class, protein-coding genes demonstrated higher expression levels than lncRNAs, where ~25% of expressed protein-coding genes, but

~80% of lncRNAs were present in samples at 1 or fewer copies per cell. The pervasive transcription of the human genome was of interest, given the low abundance of genomic protein-coding RNA transcripts. These findings suggested that the majority of genomic RNA transcripts are non-coding, and therefore, necessitate an expanded appreciation for their diverse biological functions[228].

Non-coding RNAs are RNA species that are unable to encode functional proteins and can be broadly categorized as small ncRNAs, such as microRNAs (miRNAs) or lncRNAs. MiRNAs have formed the bulk of historic attention in cancer research; however, the recent spotlight has been placed on lncRNAs, which are one of the most prevalent genomic RNA species, yet lack sufficient understanding[229,230]. LncRNAs are independently transcribed RNA species (genes) greater than 200 nucleotides in length that exhibit structural similarity to mRNAs; however, lack open reading frames (ORFs)[229,231,232]. Genes that transcribe lncRNAs are found throughout the human genome. The genomic context of lncRNAs can be discerned through the classification of lncRNAs according to their location in the genome, meaning the location from where they are transcribed relative to protein-coding genes[226]. This method of classification reveals five major categories of lncRNAs: Stand-alone lncRNAs[233,234], natural antisense transcripts[235], pseudogenes[236], long intronic ncRNAs[237,238], and divergent transcripts, promoter-associated transcripts, and enhancer RNAs[226]. The classification of lncRNAs based on their genomic context, however, does not imply function or evolutionary context[226]. The focus of this work will be stand-alone lncRNAs, which are distinct transcription units that do not overlap protein-coding genes. Those transcribed from intergenic regions are known as large intergenic/intervening non-coding RNAs (lincRNAs). These lincRNAs are often transcribed by RNA polymerase II, are polyadenylated and spliced with alternative transcript variants, although they exhibit a bias towards two-exon transcripts[229].

Importantly, it is estimated that the human genome contains approximately 50,000 lncRNA genes and 107,000 transcripts, far surpassing the number of protein-coding genes (20,000), although the functions of only hundreds are known[239,240]. Recent studies have shown that thousands of lncRNAs are evolutionarily conserved; however not to the level of protein-coding genes[233,241]. Guttmann and colleagues (2009) identified approximately 1,600 lincRNAs across four mouse cell types, with >95%

demonstrating evolutionary conservation, supporting their functional biological roles. While lncRNA transcripts are less conserved than the mRNAs of protein-coding genes, lncRNA promoter regions are often conserved at a level that mirrors those of protein-coding genes. Further, many lncRNAs have arisen within the primate lineage[229]. Importantly, lncRNA conservation may be represented by the interactions of lncRNAs with proteins and other RNA types instead of sequence stretches[242]. Together, while lncRNAs are less conserved than protein-coding genes, this does not indicate the absence of biological function[242].

A Census of cancer-promoting lncRNAs recently illustrated functional conservation of lncRNAs in tumorigenesis[243]. Conservation of gene function between human and mouse tissues provides compelling evidence for a gene's importance. This is exemplified by well-characterized protein-coding genes involved in human cancers such as TP53 and MYC, whose mutations in murine models mirror human disease formation. A recent study investigating human and mouse orthologues of LINC-PINT demonstrated common tumor suppressor activity in human and mouse cell lines through a highly conserved sequence element[244]. Carlevaro-Fita et al. (2020) generated an assortment of lncRNAs with direct functional or genetic data supporting their roles in human cancers. Strikingly, they showed that cancer-associated lncRNA genes have conserved roles in cancer between humans and mice[243]. Specifically, eight core lncRNAs were found with independently identified mouse orthologues possessing similar functional roles in cancer[243]. Thus, the likelihood that these lncRNAs play important roles in cancers is dramatically increased.

1.7 Functional roles of lncRNAs

The altered expression levels of specific lncRNAs are associated with developmental processes and diseases. While the functions of some lncRNAs are understood, the biological functions of most lncRNAs have yet to be uncovered[245]. LncRNAs are known to epigenetically modulate allelic expression in biological mechanisms such as dosage compensation[246] and genomic imprinting[247], in addition to other

critical developmental processes[248–251]. For example, the lncRNA Evf2 regulates genes in the embryonic

brain that are critical players in the development of interneurons producing the inhibitory

neurotransmitter, gamma-aminobutyric acid (GABA) and the formation of GABA-dependent neuronal

circuits in the adult brain[252]. Importantly, defective GABA-regulated circuits have been linked to

conditions such as autism[253], Tourette's syndrome[254], epilepsy[255], and schizophrenia[256].

Notably, lncRNAs represent fundamental transcriptional regulators whose cellular functions are

often determined by their sub-cellular localization (cytoplasm vs. nucleus)[257]. Nuclear-localized lncRNAs

can regulate gene expression in cis or trans, acting as transcriptional guides, transcription factor decoys,

scaffolds for molecular interactions, enhancer RNAs (eRNAs), regulators of chromatin interactions and

chromatin remodelling complexes. Conversely, cytoplasmic lncRNAs often modulate protein expression,

functioning as competing endogenous RNAs (ceRNAs) that sequester specific miRNAs and prevent their

degradation of target mRNAs[228,231,258].

The earliest identified lncRNAs exhibited functions in human development. For example,

lncRNA XIST (X-inactive specific transcript) is a well-studied lncRNA involved in X chromosome

inactivation (XCI). Identified in the 1990s, XIST was one of the first non-coding RNAs to be actively

studied as it provided crucial insights into the gene dosing mystery between male and female mammals[259].

The initiation of XCI is regulated by XIST, which coats the X chromosome in *cis* to mediate gene

silencing, maintaining its inactive state[246]. LncRNAs HOTAIR and H19 are also developmentally

important lncRNAs. HOTAIR serves as a guide for PRC2 (Polycomb repressive complex 2), leading to

the epigenetic silencing of HOX family genes essential for specification of body plans during embryonic

development[260]. LncRNA H19 is highly expressed during embryogenesis. H19 and Insulin-like growth

factor 2 (Igf2) are co-expressed in tissues derived from the endoderm and mesoderm during embryonic

development. LncRNA H19 modulates growth by controlling the expression of Igf2 in *cis*[261]. LncRNAs

are now most extensively studied in the context of disease with an emphasis on their roles in human

cancers.

LncRNAs exhibit elevated tissue-specific expression and aberrant regulation in many human cancers[262–264]. Many lncRNAs have been shown to stimulate cancer development[265], metastasis[266], and drug resistance[267]. In addition, some lncRNAs are predominately expressed in CSCs and are therefore implicated in tumorigenesis, tumor recapitulation, and resistance to chemo-, radio-, and immunotherapies[268].

Each of the developmentally important lncRNAs discussed here (XIST, HOTAIR, and H19) also play important roles in cancer progression. For example, XIST functions as a ceRNA for the tumor suppressor miRNA, miR-200c, promoting stemness characteristics and tumorigenicity in bladder cancer cells[269]. Similarly, lncRNA H19 sponges miR-200a, preventing its repression of target transcription factors ZEB1 and ZEB2, enhancing lung cancer proliferation and metastasis[270]. The lncRNA HOTAIR promotes the progression of different cancers via interactions with chromatin[271]. Notably, HOTAIR alters the chromatin organization to promote breast cancer metastasis. Specifically, HOTAIR recruits PRC2 to genome-wide target genes, inducing Histone H3 lysine 27 (H3K27) methylation, altering gene expression programs to promote cancer invasiveness and metastasis[272]. Further, HOTAIR similarly re-programs genome-wide PRC2 localization and chromatin modification in colorectal cancer[273]. Intriguingly, H19 interacts with EZH2 (enhancer of zeste homolog 2), a member of the PRC2 complex, inducing Wnt/β-catenin signaling and downregulation of E-cadherin which enhances bladder cancer metastasis[274].

1.8 LncRNA-miRNA interactions

The most commonly described functional lncRNA interactions leading to gene expression changes are with miRNAs. LncRNAs interact with miRNAs via multiple mechanisms in cancers. Importantly, lncRNA-miRNA axes can promote tumor suppression or oncogenesis and thus, characterization of the lncRNA-miRNA axis is important for determining its targetability[275]. LncRNA-miRNA interactions are critical mediators of the molecular mechanisms regulating EMT, CSCs, and drug resistance in breast cancer and therefore, possess significant potential as novel therapeutic targets (Figure 1).

For example, lncRNA MALAT1, which is upregulated in breast cancer tissues and associated with poor survival, promotes EMT in breast cancer by sponging miR-1. MALAT1 sponging of miR-1 has been shown to upregulate miR-1 target proteins Slug and Vimentin, contributing to metastasis in two different studies[276,277]. MALAT1 also promotes EMT and metastasis by increasing levels of the ZEB2 protein, a master regulator of EMT by sponging miR-204 in breast cancer[278]. Lung cancer associated transcript 1 (LUCAT1) promotes stemness among breast CSCs via sponging of miR-5582-3p, increasing the expression of TCF7L2 and Wnt signaling[279]. Colon cancer-associated transcript-1 (CCAT-1), first identified in colorectal cancer, is highly expressed in spheroid cultures and promotes stemness, migration and invasiveness in breast CSCs by functioning as a ceRNA to miR-204/211 and regulating TCF4[280,281]. LncRNA-miRNA-mRNA axes have also been implicated in the development of paclitaxel resistance in breast cancer by multiple studies. The lncRNA FTH1P3 (ferritin heavy chain 1 pseudogene 3) is upregulated in paclitaxel-resistant cell lines and sponges miR-206, attenuating its targeting of ABCB1 (ATP Binding Cassette Subfamily B Member 1)[282]. Similarly, CASC2 (cancer susceptibility candidate 2) promotes paclitaxel resistance via sponging miR-18a-5p and upregulating CDK19 (Cyclin-dependent kinase 19)[283]. Moreover, the Linc00518/miR-199a/MRP1 and AC073284.4/miR-18b-5p/DOCK4 axes have also been implicated in resistance to paclitaxel in breast cancer[284,285].

In addition, to their roles as miRNA sponges, lncRNAs can also act as miRNA precursors that regulate miRNA biogenesis, or as competitors against miRNAs for binding sites on target mRNAs[239]. Accordingly, the lncRNA LOC554202 is the host gene for the breast cancer metastasis-suppressing miRNA, miR-31 and regulates its transcriptional activity. Importantly, LOC554202 and miR-31 experience promoter hypermethylation in TNBC cell lines, downregulating their expression and contributing to invasion and metastasis[286]. Intriguingly, miRNAs can also target lncRNAs for degradation[287].

The number of lncRNAs and miRNAs functionally annotated in cancer remains modest relative to the overall number in the human genome[288]. Thus, the characterization of novel oncogenic lncRNAs and miRNAs, and specifically the lncRNA-miRNA interactions with critical importance in cancer progression, will reveal novel therapeutic targets. Ideally, precise targeting strategies will be developed that specifically

inhibit lncRNA-miRNA interactions that contribute to breast cancer progression. Targeting the interactions (instead of the individual non-coding RNA) may result in greater therapeutic impact with reduced chance of off-target effects.
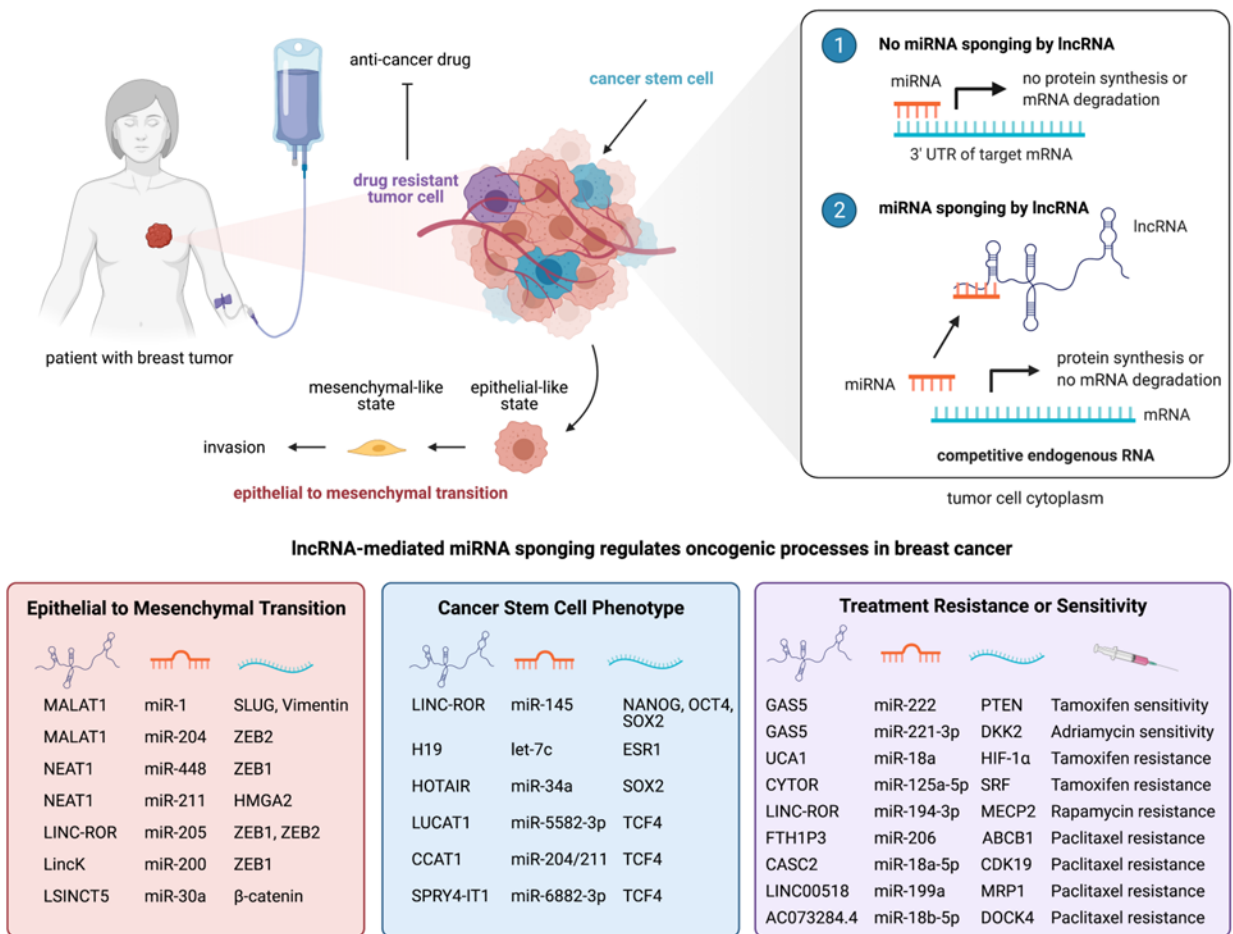
Figure 1. LncRNA-miRNA sponging interactions regulate oncogenic pathways in breast cancer. Key lncRNA-miRNA sponging interactions leading to altered levels of specific mRNAs (lncRNA-miRNA-mRNA axes) regulate the major oncogenic pathways of EMT, CSC maintenance, and drug response in breast cancer.

1.9 LncRNA protein partners mediate oncogenic functions

Oncogenic lncRNAs, while variable in their RNA archetypes, often interact with protein binding partners to carry out their regulatory gene expression functions, promoting cancer progression, regulating the localization, stabilization, or modification of their protein binding partners[289]. Importantly, numerous lncRNAs interact with proteins to facilitate metastasis in breast cancer. For example, the lncRNA BCAR4, upregulated in TNBC, contributes to breast cancer metastasis by binding to SNIP1 (SMAD nuclear interacting protein 1) and the phosphatase, PNUTS, serving as a molecular scaffold[290]. The lncRNA HOTAIR mediates breast cancer metastasis through interactions with PRC2, serving as a guide[272]. The lncRNA BORG promotes the metastasis of latent breast cancer cells and breast cancer recurrence through interactions with TRIM28[291]. Further, BORG mediates chemoresistance against doxorubicin in TNBC in part by binding and activating RPA1 (Replication Protein A1)[292]. Taken together, the identification and targeting of lncRNA interactive proteins with oncogenic functions may represent a promising treatment route to antagonize oncogenic lncRNAs.

1.10 The missing lnc: functional relevance of lncRNAs in breast CSCs

In recent years, increasing numbers of lncRNAs have been identified as playing an important role in breast cancer progression and some of these have been specifically associated within the CSC populations of breast cancers. Thus, the identification of novel lncRNAs with functional consequence in breast CSC populations is intriguing in terms of treatment. Inhibiting these lncRNAs may represent a promising novel approach for treating breast cancer and targeting therapy-resistant CSCs within these tumors. However, the identification of lncRNAs enriched in breast CSCs remains only modest.

My review of the literature identified 18 lncRNAs reported to be associated with breast CSC populations and/or promoting features associated with breast CSCs. These included HOTAIR, H19, NEAT1, MALAT1, BCAR4, DANCR, NRAD1, LINC-ROR, LINC01133, LINC00617, CCAT1, SPRY4-IT1, LncRNA-Hh, RP1-5O6.5, LINC00511, FEZF1-AS1, LncRNA-ES1, & LncRNA-HAL. A summary of their functions in breast CSCs is illustrated in table 1 and figure 2. As most of these

associations were reported in the context of cell lines, we investigated their clinical relevance by exploring patient tumor gene expression correlations. Using RNA sequencing (RNA-seq) data from the TCGA (The Cancer Genome Atlas) datasets available from cBioPortal[293,294], we extracted Spearman's correlation values between each identified CSC-associated lncRNA and CSC markers, stemness factors, EMT genes, and players in CSC-associated signaling pathways in either TNBC or all breast cancer subtypes. These results uncovered relationships between specific lncRNAs and CSC-associated genes, which were stronger in TNBC, supporting the importance of CSC-enriched lncRNAs in TNBC[268]. The literature suggested that most of the CSC-associated lncRNAs were enriched in TNBC, possibly due to the subtype's enrichment in CSCs. Our subtype enrichment analysis revealed that two CSC-enriched lncRNAs, NRAD1 (LINC00284) and DANCR are significantly enriched in TNBCs and basal-like breast cancers. Our adjacent analysis of lncRNA association with protein coding targets supports the further analysis into the potential clinical relevance of NRAD1 and DANCR in parallel to potential protein targets in the treatment of TNBC/basal-like breast cancers. Finally, we assessed the hazard ratios of CSC-associated lncRNAs and relevant protein-coding genes in different breast cancer subtypes, revealing potential targets for therapeutic intervention. Together, these analyses revealed novel correlations that may provide a foundation for further investigation of the clinically relevant correlations between lncRNAs and CSC-associated gene expression[268].

Most studies investigating CSC-enriched lncRNAs in breast cancer have not taken unbiased large-scale discovery approaches. Instead, they have interrogated the stemness-associated functions of previously identified lncRNAs. For example, HOTAIR and LINC-ROR, among others were discovered using such an approach[295,296]. While far less common in the literature, the exploration of CSC-enriched lncRNAs using large-scale discovery methods has identified novel CSC-enriched lncRNAs. For example, lncRNA-Hh (GAS1RR) was identified using a microarray, where it was subsequently upregulated in ALDH1-expressing mammospheres[297]. Further, tumors with reduced GAS1RR expression exhibited slower growth and reduced expression of stemness markers. Similarly, lncRNA H19 was identified via

microarray and was shown to be upregulated in ALDH1+ breast CSC populations. H19 expression was found to be essential for breast CSC survival *in vivo*, supporting its potential as a therapeutic target[298–300].

Some other studies investigating CSC-enriched lncRNAs have also taken an unbiased approach. First, Zhang et al. (2016) conducted an analysis of patient tumor RNA-seq data (TCGA), unveiling more than 50 lncRNAs that are upregulated in TNBCs and basal-like breast cancers, most of which were uncharacterized, representing potential novel targets in these subtypes[301]. Later, the Marcato lab screened for lncRNAs enriched in TNBCs and CSCs (ALDH^high cell populations) and associated with poor patient outcomes to identify a novel oncogenic lncRNA associated with CSCs in TNBC. This analysis revealed that approximately 25% of the TNBC/basal-like breast cancer-enriched lncRNAs identified by Zhang and colleagues were enriched in ALDH^high CSC populations of TNBCs[302]. Among these lncRNAs was prostate androgen regulated transcript 1 (PART1).

We recently identified PART1 as a novel CSC-associated lncRNA in TNBC, whose inhibition may improve patient outcomes. PART1 is enriched in TNBC/basal-like patient tumors and CSC populations, is androgen-inducible in breast cancer cells, and is associated with worse patient outcomes among basal-like breast cancer patients (Cruickshank, Wasson et al., submitted). Knockdown of PART1 decreased cell proliferation, increased apoptosis, and reduced mammosphere formation potential in TNBC cell lines. Furthermore, knockdown of PART1 reduced tumor growth of TNBC cells *in vivo.* Transcriptome analyses revealed that PART1 affects the expression of up to hundreds of genes in TNBC cell lines. Together, PART1 may represent a novel target in TNBC.

Most importantly, the Marcato lab's screen identified NRAD1 (non-coding RNA in the aldehyde dehydrogenase 1A pathway), a novel lncRNA which we showed to be regulated by CSC marker ALDH1A3 and contribute to its changes in gene expression to mediate breast cancer progression. From a clinical perspective, targeting NRAD1 with antisense oligonucleotides reduced cell viability and tumor growth of TNBC cells from a patient-derived xenograft, illustrating a novel CSC-associated target in TNBC of potential clinical value[302].

Table 1. Summary of long non-coding RNA functions in breast CSCs.

lncRNA	Mechanism of action in breast CSCs
HOTAIR	Regulates stemness in TNBC by sponging miR-34a, promoting upregulation of SOX2. HOTAIR promotes proliferation of CSCs within breast cancer cell lines by binding the promoters of p53 and p21[296].
H19	Sponges miRNA tumor suppressor, let-7, promoting an increase in breast CSC-enriched pluripotency factor LIN28 and glycolytic enzyme PDK1[299,303].
NEAT1	Promotes tumor growth, chemoresistance and maintenance of CD44^{high}CD24^{low}, ALDH^{high}, and SOX2^{high} CSC populations in TNBC[304].
MALAT1	ceRNA of miR-1 allowing MALAT1 interaction with Slug, prompting TNBC progression[277]. Also promotes TNBC progression by regulating KDM5B which regulates formation/maintenance of breast CSCs[305]. Binds RPB HuR, forming a repressive complex regulating CD133[306].
BCAR4	Regulates non-canonical Hh cascade gene transcription in a GLI2-dependent manner to promote metastasis in TNBC[307].
DANCR	Involved in positive regulation of stemness factors CD44, ABCG2, and ALDH1 in TNBC[308].
NRAD1	Oncogenic chromatin-binding lncRNA regulated by ALDH1A3 that contributes to ALDH1A3-mediated gene expression[302].
LINC-ROR	Induces EMT to promote the formation of breast CSCs[309].
LINC01133	Induced by mesenchymal stem cells in TNBC cells and associated with generation of breast CSC-like cells and modulation of the miR-199a-FOXP2 pathway. Regulates pluripotency factor, KLF4, which promotes stemness[310].
LINC00617	Induces EMT in TNBC cell lines, promoting an increase in CD44+/CD24- cells, increased mammosphere formation, and metastasis through regulation of the SOX2 stemness factor[311].
CCAT1	Regulates stem factors NANOG, SOX2, OCT4, and ALDH1A1; Acts as a ceRNA for miR-204/211 which targets TCF4, a transcription factor in the Wnt/β-catenin pathway[280].
SPRY4-IT1	Promotes stemness by sequestering miR-6882-3p which targets TCF4, allowing initiation of canonical Wnt signaling[312].
lncRNA-Hh	Twist-induced lncRNA that directly targets GAS1 to initiate Hh signaling in breast cancer, promoting SOX2 and OCT4 expression, EMT, tumorigenesis, and cells with CSC properties[297].
RP1-5O6.5	Regulated by KLF5, induces breast cancer growth and metastasis by inhibiting translation of cell cycle inhibitor p27kip1, promoting stemness[313].
LINC00511	Functions as a ceRNA, sequestering miR-185-3p to upregulate NANOG via E2F1[314]. Also, among several lncRNAs enriched in Aldefluor+ CSC populations in TNBCs[302].
FEZF1-AS1	Acts as a ceRNA for miR-39a, which targets NANOG, OCT4 and SOX2[315].
LINC01108 (ES1)	Acts as a ceRNA for miR-106b to decrease expression of E-cadherin and miR-200; Regulates stemness factors SOX2 and OCT4 and their downstream targets, miR-306, miR-106b[316].
lncRNA-HAL	Regulates the expression of CD44, CD24 and NANOG[317].
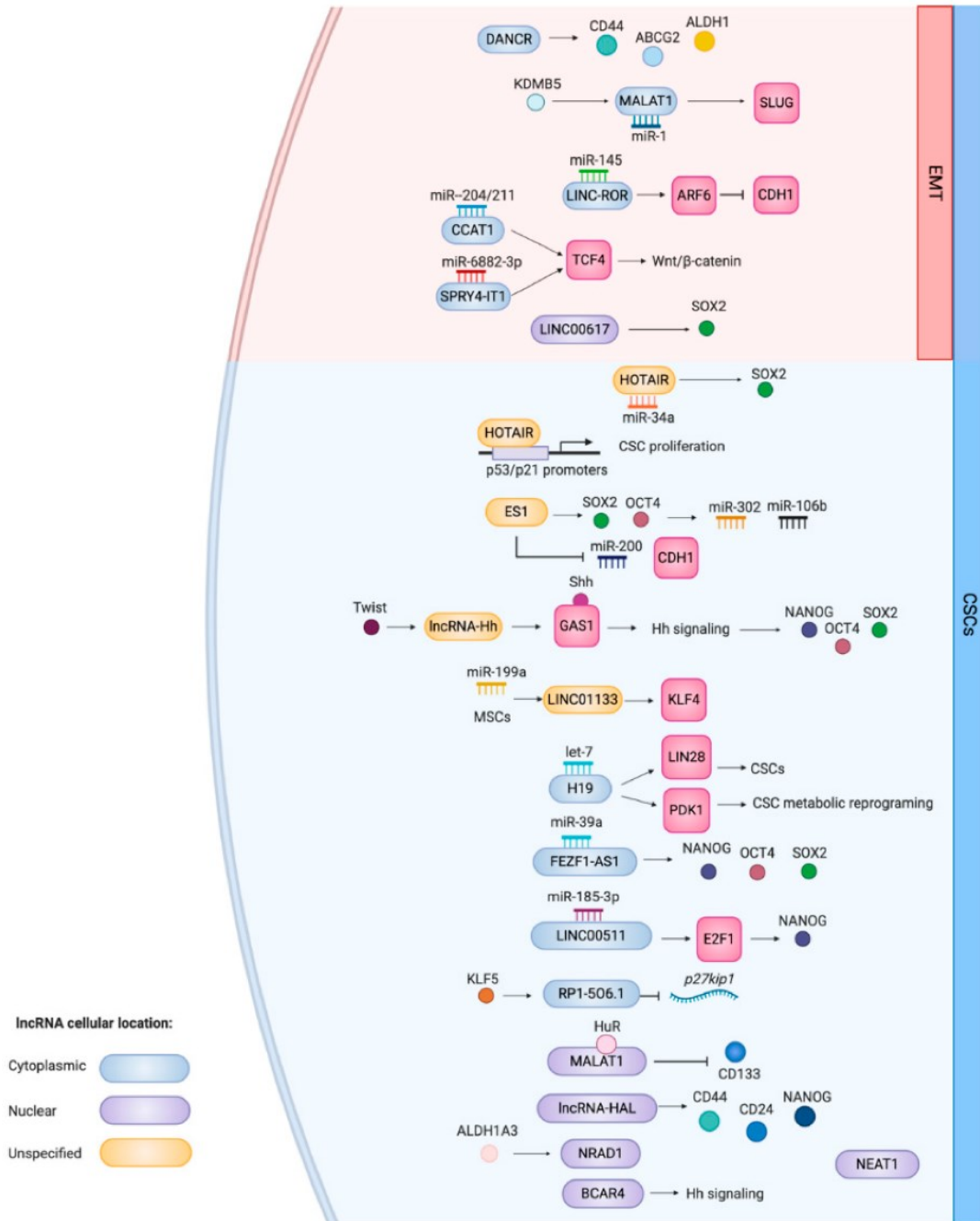
Figure 2. LncRNAs associated with breast CSCs and their functions. Schematic overview of the CSC and EMT-related functions of lncRNAs that have been associated with breast CSCs. This is a representation of functions summarized in Table 1. LncRNAs implicated with EMT genes and the processes are located in the red area of the cell. Those associated with development of the CSC phenotype are located in the blue area of the cell. If known, the cellular localization of the lncRNA is indicated by the color coding (blue = cytoplasmic; purple = nuclear; orange = unspecified).

1.11 Therapeutic Inhibition of non-coding RNAs and potential challenges

The multifaceted roles of lncRNAs in cancer progression marks them as intriguing therapeutic targets, especially in cancers in which such targets are absent, such as TNBCs. LncRNA expression is often experimentally inhibited *in vitro* with anti-sense oligonucleotides (ASOs) or with shRNAs (short-hairpin RNAs), a form of RNAi (RNA interference). RNAi represents an effective method in the analysis of gene functions in mammalian cells. RNAi is based on the cytoplasmic delivery of double-stranded RNA identical to the target sequence, inducing sequence specific degradation of host mRNA and reducing target gene expression through a pathway involving the RNA-induced silencing complex (RISC)[318]. RNAi accomplished via shRNAs is a popular tool for gene knockdown, using cellular shRNAs synthesized from DNA vectors. These shRNAs can be introduced into mammalian cells via infection with viral vectors, facilitating the stable DNA integration of shRNAs and allowing long-term target gene knockdown[319]. While shRNA-mediated gene knockdown is a useful tool for the inhibition of lncRNAs *in vitro*, its use *in vivo* is not effective from a clinical perspective[320]. One reason for this is the variable subcellular localization of lncRNAs. ASO and RNAi-based gene knockdown methods differ in their efficacy with respect to the subcellular compartmentalization of the target lncRNA[321]. Nuclear-localized lncRNAs are more effectively suppressed using ASOs, while the suppression of cytoplasmic lncRNAs is better accomplished using RNAi-based methods[321]. In addition, the delivery of shRNAs is often accomplished using lentiviral vectors, raising the concern of off-target effects.

One group recently demonstrated that the introduction of tumor-targeting siRNA nanoparticles against the oncogenic lncRNA DANCR showed effective therapy in TNBC[322]. These nanoparticles showed promising therapeutic efficacy both *in vitro* and *in vivo* using TNBC xenograft models. Thus, the delivery of siRNA nanoparticles targeting oncogenic lncRNAs may be a promising therapeutic approach for the treatment of TNBC and other cancers.

More commonly used in the modulation of disease-related lncRNAs are ASOs, which are DNA oligonucleotides that bind mRNA molecules in an antisense fashion, inhibiting lncRNA transcription through the degradation of the target mRNA by recruitment of RNase H[323]. In addition, ASOs can

function by modulating alternative splicing modifications to include or exclude exons or by miRNA inhibition which prevents miRNA binding to target mRNAs[323]. While early ASOs could successfully prevent RNA maturation, they were rapidly degraded by nucleases[324].

Current ASOs possess modifications that improve their strength and stability *in vivo*. The ASOs most investigated for their ability to silence oncogenic lncRNAs are GapmeRs[321]. GapmeRs possess a phosphorothioate (PS) backbone to prevent nuclease-mediated degradation, allowing more potent ASO activity. In addition, the PS modification promotes high binding affinity to proteins, allowing efficient cellular uptake and supporting target degradation by RNase H[321]. GapmeRs consist of 2' modified sugar ring bases separated by an intervening region of unmodified nucleotides, which facilitates target degradation without a heightened inflammatory response[325].

Locked nucleic acid (LNA) GapmeRs use synthetic nucleotides containing a methylene bridge between the 2'-O of the ribose ring and 4'-C of the nucleic acid, generating a "locked" formation and exhibit greater potency along with RNase H activity. Together, the central DNA gap, PS backbone and locked nucleic acids form a highly efficient, stable and nuclease resistant ASOs with potent RNase H activity for degradation of target mRNA[326].

ASOs have been investigated in numerous human clinical trials including those for cancers, diabetes, and muscular dystrophy[327]. In addition, ASOs have generated significant interest for their potential use in human neurogenerative diseases, where they have demonstrated safety and effectiveness in animal models and human clinical trials[323]. For example, ASOs effectively disseminate throughout the central nervous system (CNS) upon delivery to the cerebral spinal fluid, allowing their application to CNS diseases which lack sufficient treatments, akin to many human cancers[323]. Clinical trials investigating the treatment of spinal muscular atrophy (SMA) with ASOs have demonstrated striking success, leading to the FDA approval of these ASOs for treatment of the disease[328]. In addition, ASOs have demonstrated promising clinical results in the treatment of amyotrophic lateral sclerosis (ALS)[329,330] and Huntington's disease[331,332].

In addition to their activity in neurodegenerative diseases, LNA GapmeRs have demonstrated promising preclinical results for their *in vivo* inhibition of oncogenic lncRNAs. For example, LNA GapmeRs have been used to successfully reduce lung metastases *in vivo* via inhibition of lncRNA MALAT1[333] and breast cancer metastases via inhibition of lncRNA BCAR4[334]. Similarly, targeting of MALAT1 with LNA GapmeRs has been shown to reduce cell proliferation and induce apoptosis in *in vitro* and *in vivo* models of human multiple myeloma[335]. LNA GapmeR-mediated inhibition of lncRNA SAMMSON, which contributes to melanoma progression, reduced tumor size in mice bearing human melanoma PDXs. Further, inhibition of SAMMSON desensitized PDXs to conventional melanoma treatments[336].

In addition to the inhibition of oncogenic lncRNAs with ASOs, these lncRNAs may be inhibited with other methods including CRISPR/Cas9. Targeting ncRNAs with CRISPR/Cas9 has been reported in cancer[337]. Specifically, lncRNA UCA1 (urothelial carcinoma-associated 1), upregulated in bladder cancer, could be inhibited by transcript-specific CRISPR/Cas9-associated gRNAs (guide RNAs) *in vitro* and *in vivo*, demonstrating its potential therapeutic value[338]. These findings suggest that there are strategies in development that will enable the clinical targeting of functionally characterized oncogenic lncRNAs that are specific to certain breast cancer subtypes (e.g., TNBC/basal-like breast cancer) and the CSCs within these tumors.

While the lncRNA silencing methods are promising therapeutic strategies, a thorough assessment of off-target effects, potential toxicity, and drug delivery/precision targeting is required. Additionally, the multiple transcript variants of lncRNAs is more of a potential confounding factor than when considering strategies for targeting proteins. Depending on the specifics of the lncRNA silencing strategy, not all transcript variants will be targeted, possibly reducing the efficacy of the treatment if that variant is functional. In addition, the functions of most lncRNAs remain uncharacterized, further complicating the design of effective treatments and delivery strategies. Finally, whether lncRNAs represent more suitable targets over protein-coding genes is unknown and would require comparative analyses in clinical trials.

1.12 Non-coding RNA in the Aldehyde Dehydrogenase 1a Pathway (NRAD1) is a potential target for breast CSCs

Formerly known as LINC00284, NRAD1 is an oncogenic lncRNA involved in the progression of several human cancers. First characterized as oncogenic by the Marcato laboratory in breast cancer[302], the lncRNA was later shown to increase proliferation, migration, invasion, angiogenesis, and tumorigenic ability and inhibit apoptosis by gene regulation in ovarian cancer[339]. Specifically, the lncRNA recruits and activates NF-κB signaling and down-regulates mesoderm-specific transcript (MEST) in ovarian cancer cells[339]. In addition, NRAD1 was recently found to be upregulated in serous ovarian carcinoma (SOC) and associated with the upregulation of the SOX9 transcription factor[340]. Elevated NRAD1 expression was associated with poor prognosis and was shown to be an independent risk factor in patients with SOC. Thus, NRAD1 may represent a novel prognostic marker in SOC. The authors propose that NRAD1 transcription is regulated by SOX9, where NRAD subsequently functions as a ceRNA, sequestering miR-195-5p and miR-497-5p. This sponging activity was proposed to antagonize the miRNA-mediated inhibition of their target genes, thereby allowing their expression (i.e., MYB, SOX9 & ESRP1) and facilitating SOC initiation and progression (i.e. proliferation, invasion & metastasis)[340].

NRAD1 also promotes the progression of other human cancers. One report found that NRAD1 expression is closely related to decreased survival and high risk in gastric cancer[341]. In addition, NRAD1 has been implicated in the development and progression of muscle-invasive bladder cancer[342]. Conversely, the lncRNA has been proposed to function as a competitive endogenous RNA associated with improved clinical outlook in papillary thyroid cancer (PTC) where its expression was positively correlated with patient survival[343]. Specifically, Zhao and colleagues (2015) provided evidence to support the function of NRAD1 as a ceRNA of miR-205[343]. In breast cancer and melanoma, miR-205 functions as a tumor suppressor with greatly reduced expression in these cancers, with particularly reduced expression in TNBCs, which may speak to its sponging by NRAD1[344–346]. Most recently, NRAD1 was shown to sponge miR-211-3p, leading to upregulation of MAFG (MAF bZIP transcription factor B) and promotion of malignant characteristics in oral squamous cell carcinoma (OSCC) cells[347].

Vidovic and colleagues (2020) first reported that NRAD1 is highly enriched in TNBC and breast CSCs and was associated with worse patient outcomes[302]. They determined that the lncRNA is predominately nuclear-localized and is regulated by the CSC marker ALDH1A3 and its product, retinoic acid. Importantly, NRAD1 has enriched chromatin interactions among the genes it regulates, suggesting it may regulate at least some gene expression through interactions with chromatin. Further, NRAD1 contributes significantly to gene expression changes induced by ALDH1A3. Thus, the expression of this lncRNA may represent a novel mechanism of CSC gene regulation. Targeting NRAD1 using ASOs reduced cell viability and tumor growth in TNBC cells and a patient-derived xenograft (PDX) model and reduced mammosphere formation. Together this data suggested that NRAD1 is a novel oncogenic CSC-associated lncRNA and potential clinical target in the treatment of TNBC and breast CSCs.

1.13 Rationale and Hypothesis

NRAD1 is a novel oncogenic lncRNA involved in breast cancer progression induced by the functional CSC marker ALDH1A3 and has a largely uncharacterized mechanism[302]. While NRAD1 is known to contribute to regulation of over 20% of ALDH1A3-mediated gene expression, tumor growth, cell survival, and stemness, the mechanisms through which it exerts these functions remain to be elucidated[302]. Further, functional characterization of NRAD1 have implicated both gene regulation through chromatin interactions in breast cancer and through miRNA interactions in other cancers[302,340,343,347]. It is unclear if the described chromatin binding of NRAD1 in breast cancer requires protein binding partners for gene regulation and if NRAD1 also regulates gene expression in breast cancer via miRNA interactions. Moreover, it is unclear to what extent the ALDH1A3 induced gene expression changes mediated by NRAD1 is connected to changes in miRNAs.

I hypothesize that NRAD1 regulates gene expression via multiple mechanisms in breast cancer. To explore the different mechanisms of NRAD1-mediated gene regulation, I proposed two key aims. In Aim 1, I investigated the putative roles of NRAD1-binding proteins in NRAD1-mediated gene regulation. The use of lncRNA-directed proteomics methods (lncRNA pulldown and chromatin isolation by RNA

purification both coupled with mass spectrometry) helped to identify proteins binding to NRAD1. Short-hairpin RNA-mediated knockdowns of identified protein-coding mRNAs followed by downstream gene expression assays highlighted the potential roles of these proteins in NRAD1-mediated gene regulation. Together, the identification of NRAD1 protein partners with regulatory functions may represent attractive therapeutic targets for breast cancer. The purpose of this work was to further uncover novel mechanisms for ALDH1A3 and NRAD1 in breast cancer.

In Aim 2, I investigated changes in the miRNA landscape in response to NRAD1 or ALDH1A3 knockdown in two different TNBC cell lines. To inhibit NRAD1, cells were treated with two different anti-NRAD1 GapmeRs or the negative control GapmeR. ALDH1A3 knockdowns were previously generated using two different shRNA clones or the control. Total RNA enriched for miRNAs was extracted from NRAD1 and ALDH1A3 knockdown cells and sent for miRNA microarray analyses to assess genome wide changes in miRNA expression in response to NRAD1 or ALDH1A3 knockdown. These analyses sought to reveal novel interactions between NRAD1 and miRNAs which may contribute to its regulation of gene expression in breast cancer. Moreover, these analyses sought to reveal miRNAs commonly regulated by NRAD1 and ALDH1A3, elucidating novel roles for NRAD1 and ALDH1A3 in regulation of gene expression in breast cancer.

Together, my goal was to provide a framework for how NRAD1 regulates gene expression in breast cancer. Advancing our understanding of NRAD1-mediated breast cancer progression can inform novel avenues for NRAD1-targeting in breast cancer.

CHAPTER 2 – MATERIALS AND METHODS

2.1 Cell lines and Cell Culture Conditions

Throughout this body of work, MDA-MB-468 and SUM149PT (SUM149) human breast cancer cell lines, and HEK293T human embryonic kidney cells were used. MDA-MB-468 cells were first isolated by Cailleau et al. (1977) from a pleural effusion of a 51-year-old black female with metastatic adenocarcinoma of the breast[348]. SUM149 cells were isolated from a female patient with ER/PR-negative, HER2-positive (unactivated) inflammatory breast cancer[349]. HEK293T cells are highly transfectable derivatives of human embryonic kidney 293 (HEK293) cells used for generation of lentiviral gene knockdowns. Stable S100A8 lentiviral shRNA knockdowns were generated in MDA-MB-468 cells. Stable MDA-MB-468 and SUM149 ALDH1A3 retroviral shRNA knockdowns were previously generated[117,141]. MDA-MB-468 and HEK293T cells were purchased from the American Type Culture Collection (ATCC) and SUM149 cells were purchased from BioIVT. Both cell lines were cultured according to the supplier's recommendations as follows. MDA-MB-468 cells were cultured in Dulbecco's Modified Eagle Medium (DMEM), supplemented with 10% fetal bovine serum (FBS) and 1X antibiotic/antimycotic (AA). MDA-MB-468 cells with stable gene knockdown were cultured in the added presence of 0.25ug/mL puromycin. SUM149 cells were cultured in Ham's F-12 Nutrient Mix (F-12), supplemented with 5% FBS, 1X AA, 1X HEPES buffer, 1ug/mL hydrocortisone, and 5ug/mL human insulin. Similarly, SUM149 cells with stable gene knockdown were cultured with the addition of 0.25ug/mL puromycin. HEK293T cells were cultured in DMEM supplemented with 10% FBS (no AA). All cell lines were incubated at 37°C with 5% CO_2.

2.2 Quantitative Reverse Transcription PCR

All quantitative reverse transcription polymerase chain reactions (RT-QPCR) were conducted in two steps, where cDNA was first reverse transcribed from total RNA before proceeding with the QPCR reaction. First, mammalian cells were collected in TRIzol reagent (Invitrogen) and total RNA was

extracted using a PureLink RNA kit (Invitrogen) according to the manufacturer's instructions with the incorporation of an on-column DNase I treatment step. RNA was reverse transcribed into complimentary DNA (cDNA) using the iScript reverse transcription supermix (Bio-Rad), per the manufacturer's instructions. RT-QPCR was conducted using SsoAdvanced Universal SYBR Supermix (Bio-Rad) and gene-specific primers (Table 2) according to the manufacturer's recommended procedure using a CFX Connect or CFX384 Touch Real Time PCR Detection System (Bio-Rad). Standard curves were generated for each primer pair to ensure efficiency and precision of primer target detection. Generated primer efficiencies were incorporated into the CFX manager software (Bio-Rad) when assessing gene expression. Gene expression was calculated relative to at least two reference genes using CFX manager software.

2.3 Generation of pcDNA3-NRAD1

This procedure was complete by previous Marcato lab MSc. student, Dejan Vidovic. Standard molecular cloning techniques were used to generate the pcDNA3-NRAD1 plasmid. Restriction enzymes were obtained from New England Biolabs (NEB). The size of restriction enzyme products was verified using a 1% agarose gel and when appropriate, was extracted from the gel matrix using a gel extraction kit (Qiagen). All plasmids were isolated from bacteria using a plasmid miniprep kit (Qiagen), according to the manufacturer's protocol. All constructs were characterized using PCR and sequences were confirmed using sanger sequencing (GeneWiz). To generate the NRAD1 insert, first-strand cDNA was synthesized from RNA derived from MDA-MB-468 cells using SuperScript III reverse transcriptase and oligo(dT)$_{12-18}$ primer (Invitrogen), according to the manufacturer's instructions. The NRAD1 sequence was retrieved and amplified with NRAD1-specific primers (Table 2) with flanking 5'-*Hind*III (forward) and 3'-*Bam*HI (reverse) sites, in the presence of Platinum SuperFi DNA polymerase (Invitrogen). The amplified NRAD1 cDNA with primer-introduced restriction sites were digested with *Hind*III and *Bam*HI and ligated into the pcDNA3 vector, which was digested with the same two restriction enzymes using T4 DNA ligase. Competent TOP10 *Escherichia coli* cells were transformed and incubated overnight at 37°C. The

pcDNA3-NRAD1 plasmid was isolated from bacteria using a plasmid miniprep kit and the cloning success was verified using Sanger sequencing (GeneWiz).

2.4 RNA-Protein pulldown assays

An optimized RNA pulldown protocol coupled with mass spectrometry (MS) was used to identify NRAD1-assoicated proteins, as previously described, with some modifications[350]. First, 1.2×10^8 MDA-MB-468 cells were collected and washed in PBS. Cells were briefly centrifuged, and the pellet was resuspended in a 20mL solution comprised of a 1:1:3 ratio of PBS, nuclear isolation buffer, and double-distilled water (ddH$_2$O) respectively (components in Table 5). Cells were lysed on ice for 20 minutes using cell lysis buffer (Table 5) and nuclei were pelleted at 600RCF. The nuclear pellet was resuspended in buffer A (Table 5) and nuclei were lysed via sonication under the following conditions: temperature = 4°C, amplitude = 70%, ON time = 7 min 30 seconds; cycles = 15S ON, 45S OFF (to ensure constant 4°C temperature was maintained). Sonicated nuclei were centrifuged, and the supernatant was saved as purified nuclear lysate. NRAD1, cloned into the pcDNA3 vector, was *in vitro* transcribed in the presence of biotin (Biotin labelling mix, Roche) using either T7 polymerase or Sp6 polymerase, generating two uniquely biotinylated NRAD1 transcripts for comparison. Biotinylated NRAD1 transcripts were incubated with DNase I and purified using G50 sephadex columns (Roche). Next, the biotinylated transcripts were denatured at 65°C, then slowly cooled to 4°C to allow proper folding. The nuclear lysate was pre-cleared with buffer A-washed magnetic C1 streptavidin beads. After removal of beads, 25 ug of each biotinylated NRAD1 transcript was added to the lysate and allowed to incubate overnight at 37°C with rotation. Biotinylated NRAD1 bound to proteins were isolated by addition of 30 uL of buffer A-washed streptavidin beads. Collected beads were washed five times with buffer A with fresh protease inhibitor cocktail and RNase inhibitors. NRAD1-protein complexes were eluted from beads using 3mM biotin (Sigma) in PBS at room temperature for 30 minutes with rotation. Proteins were trichloroacetic acid (TCA)-precipitated and subjected to mass spectrometry to identify NRAD1-bound proteins (identified proteins listed in Appendix 1).

2.5 Chromatin Isolation by RNA Purification and Mass Spectrometry (ChIRP-MS)

ChIRP-MS (chromatin isolation by RNA purification followed by mass spectrometry) experiments were performed as previously described by Chu et al. (2018)[351]. First, tiling antisense DNA probes spanning the sequence of the NRAD1 long transcript variant (3208bp) were generated using the Stellaris FISH Probe Designer (https://www.biosearchtech.com/support/tools/design-software/stellaris-probedesigner), according to the following parameters: 20-mer probe every 100 bp of RNA, spacing between probes of 80 nt, target GC% = 45. The probes were purchased from BioSynthesis with HPLC purification and 3'-biotin-TEG modification (Table 4). Probes were separated into odd and even probesets and were pooled together within their respective groups to a final concentration of 100uM.

Approximately 250 million MDA-MB-468 cells were harvested and crosslinked using 3% formaldehyde (diluted in PBS) for 30 minutes at room temperature. Crosslinking was quenched with 0.125M glycine for 5 minutes. Cells were washed with PBS, pelleted, and weighed.

Cell pellets were resuspended in cell lysis buffer (Table 5) before sonication (QSonica Q800R2) in polystyrene tubes using the following conditions (temperature = 4°C, amplitude = 70%, ON time = 35 mins; cycles = 15S ON, 45S OFF (to ensure constant 4°C temperature is maintained). Chromatin size (1-2 kb) was confirmed using gel electrophoresis according to the protocol (Chu et al. 2018). The cell lysate was centrifuged at 16 000 RCF for 10 minutes (4°C) and supernatant was transferred to a fresh tube to begin ChIRP.

Some lysate (30 uL each) was saved for RNA and protein input samples. The sonicated cell lysate was pre-cleared with cell lysis buffer washed magnetic C1 streptavidin beads and incubated at 37°C for 30 minutes with rotation. Following pre-clearing (performed concurrently with addition of RNase A to the lysate for RNase treatment control at a final concentration of 10ug/mL) and separation of beads from the lysate using a magnet, the supernatant was transferred to a new tube. The lysate was supplemented by 2X its volume with hybridization buffer (Table 5) and either odd or even pooled probesets (1uL of 100uM probes per 1mL of lysate) and incubated for 16 hours at 37°C with constant end to end rotation. Next, cell lysis buffer washed magnetic C1 streptavidin beads were added to the lysate (100uL beads per 1uL of

100uM probes) before another incubation at 37°C for 30 minutes. The beads were subjected to five washes with wash buffer (Table 5), performed at 37°C for 5 minutes each wash. At the last wash, beads were transferred to a fresh tube for RNA extraction to later confirm NRAD1 enrichment over reference genes using QPCR.

RNA samples on beads were resuspended in 100uL pK buffer (Table 5) and the RNA input was topped up to 100uL with pK buffer. Samples were incubated at 50°C for 45 minutes with constant rotation. Then, the samples were heated at 95°C for 10 minutes to complete reversal of cross-links and denature the streptavidin beads. Finally, 1mL TRIzol reagent was added to each sample and total RNA was extracted using a PureLink RNA kit (Invitrogen), according to the manufacturer's instructions, but with a substitution of 100% ethanol. The extracted total RNA was used to generate cDNA and perform QPCR (protocol described in section 2.2) using NRAD1-specific primers. NRAD1 fold enrichment over reference transcripts was calculated.

Protein samples on beads were resuspended in 100uL biotin elution buffer (Table 5). Proteins were eluted from beads at room temperature with mixing for 20 minutes followed by an incubation at 65°C for 10 minutes. Upon separation from beads, the protein containing supernatant was transferred to a fresh tube and proteins were eluted from beads a second time. The two protein eluents were pooled together for protein precipitation overnight at 4°C using TCA. The following day, proteins were pelleted at 16 000 RCF for 30 minutes at 4°C, washed with cold acetone and pelleted once more for 5 minutes. Following the removal of acetone, protein pellets were subjected to a SpeedVac (Thermofisher) to dry the pellets. Protein pellets were then solubilized in Laemmli sample buffer (Bio-Rad) and boiled at 95°C with occasional mixing to reverse cross-links. Finally, the protein samples were subjected to polyacrylamide gel electrophoresis (PAGE) using mini-PROTEAN TGX gels (Bio-Rad) and silver staining to identify proteins for mass spectrometry.

2.6 Silver Staining of SDS-PAGE Gels for Mass Spectrometry

To visualize protein bands for Mass Spectrometry analysis, protein samples run through polyacrylamide gels were first fixed twice in fixation solution (Table 5) for 30 minutes. The gel was then placed in sensitizing solution (Table 5) for 30 mins followed by three 5-minute washes with nuclease-free water. Next, the gel was placed in silver nitrate solution (Table 5) for 20 minutes followed by two 1-minute washes with nuclease-free water. The gel was then placed in developing solution (Table 5) for 10 minutes followed by 0.5M EDTA stopping solution (Table 5) for 10 minutes and was finally washed 3 times for 5 minutes with nuclease-free water.

2.7 Lentiviral knockdown generation of MS-identified proteins

Stable short hairpin ribonucleic acid (shRNA) knockdown clones were generated for each MS-identified protein of interest. Each clone was generated using the lentiviral vector GIPZ (EGAD, Dalhousie University, NS, Canada) with either the shRNA scramble sequence or two different shRNA sequences specific to each protein of interest. Briefly, HEK293T cells were cultured in DMEM supplemented with 10% FBS and no AA and were transfected (Lipofectamine, Invitrogen) with packaging plasmids (0.75ug pSPAX2 and 0.25ug pMDG.2) and control or protein-specific shRNA clones (1ug). Forty-eight hours later, lentivirus-rich HEK293T cell supernatants were added to cultured MDA-MB-468 cells. Forty-eight hours later, stable transfectants were selected using 1.5ug/mL puromycin (Sigma). Selected clones were maintained in DMEM supplemented with 10% FBS, 1X AA, and 0.25ug/mL puromycin. Total RNA was extracted from cells using the PureLink RNA kit (Invitrogen) according to the manufacturer's instructions. All knockdowns were confirmed using RT-QPCR (section 2.2) and western blotting (Section 2.11).

2.8 GapmeR Treatments

  To generate transient *in vitro* NRAD1 knockdowns, GapmeR treatments were used. GapmeRs were mixed with OptiMEM reduced serum media (Invitrogen Thermo Fisher Scientific) and TransIT-BrCa transfection reagent at a ratio of 10:1:2 parts OptiMEM:GapmeR:TransIT, to a final treatment concentration of approximately 15 nM GapmeR. MDA-MB-468 or SUM149 cells received treatment. Quantification of GapmeR-induced NRAD1 knockdown was performed following treatment of MDA-MB-468 or SUM149 cells with either negative control GapmeR (5'-AACACGTCTATACGC-3'), GapmeR #3 (5'-GCTGAACGCTGCCTTT-3') or GapmeR #4 (5'-CTTTGCTGAACTGATG-3'). Following a 48-hour incubation, cells were collected in TRIzol reagent for RNA extraction and RT-QPCR.

2.9 RNA Immunoprecipitation with Formaldehyde Crosslinking

  Nuclear immunoprecipitation of RNA-binding proteins (RIP) allows for the isolation and identification of RNA species associated with chromatin. Thus, to validate the possible interaction of S100A8 with nuclear lncRNA NRAD1, we performed RIP with cross-linking in MDA-MB-468 cells using the Magna Nuclear RIP Nuclear RNA-Binding Protein Immunoprecipitation Kit (Millipore) with some optimization.

  Briefly, MDA-MB-468 cells grown to 60% confluency in 15cm dishes were cross-linked by adding 381uL 16% methanol-free formaldehyde directly to 20 mL of 1X DMEM (10% FBS) growth media followed by incubation for 10 mins at room temperature with shaking. Excess formaldehyde crosslinking was quenched by adding 2mL 10X glycine to each dish and incubating at room temperature with shaking for 5 minutes. Cells were then washed twice with 10mL ice-cold PBS which was replaced with 1mL 1X PBS (with 200X protease inhibitor cocktail) for cell scraping into a microfuge tube. Cells were pelleted by centrifugation at 800 x g and 4°C for 5 minutes. Upon removal of the supernatant, cross-linked pellets were lysed and sonicated according to the following protocol.

Cross-linked cell pellets were resuspended in nuclear isolation buffer (with fresh PIC and RNase inhibitor) and incubated on ice for 15 mins, vortexing for 10 seconds at high speed every 5 minutes. Cells were then pelleted by centrifugation at 800 x g and 4°C for 5 minutes. Upon removal of the supernatant, nuclear pellets were resuspended in RIP Cross-linked lysis buffer (with fresh PIC and RNase inhibitor). The lysate was then sonicated according to optimized sonication conditions.

To optimize the sonication protocol such that chromatin was sheared from 200bp – 1kb, the cell lysis protocol above was followed until the sonication step with a new cross-linked cell pellet. After cell lysis, a 5 uL aliquot of unsonicated sample was removed for comparison to sonicated samples using gel electrophoresis. The sample was sonicated initially for 10 mins (15S ON, 45S OFF, 70% amplitude, 4°C) and then for an additional 5 minutes up to 45 minutes, removing a 5 uL aliquot after the initial 10 minute sonication and each additional 5 minute sonication. Each 5 uL sample was diluted with 90 uL proteinase K buffer (Table 5) containing no proteinase K enzyme and centrifuged for 10 minutes at 16,000 x g. The supernatant was removed and 5 uL proteinase K enzyme was added to each sample before incubation at 50°C for 45 mins. DNA was extracted using a Qiagen mini elute PCR purification kit according to the manufacturer's protocol and the samples were subjected to gel electrophoresis using a 1% agarose gel.

Following optimization, lysates were sonicated for a total of 10 minutes (15S ON, 45S OFF, 70% amplitude, 4°C). The sonicated lysate was centrifuged at 10,000 x g for 10 mins at 4°C and the supernatant was aliquoted to be used for immunoprecipitation. Five microlitres of the supernatant was removed to analyze chromatin size via gel electrophoresis according to the sonication optimization protocol.

Magna ChIP protein A/G magnetic beads (10uL per RIP reaction) were transferred to a microfuge tube and washed twice with nuclear RIP dilution buffer using a magnetic stand. After removal of the supernatant at the second wash, beads were resuspended in RIP dilution buffer (100uL per RIP) and the appropriate antibodies were added as follows: 5ug anti-S100A8 antibody (Table 3), 5 ug rabbit IgG (S100A8 negative control, Table 3), 1 ug anti-EZH2 antibody (positive control, Table 3) and 1 ug mouse IgG (EZH2 negative control, Table 3). Tubes with antibodies were incubated with rotation for 30 mins at

room temperature, centrifuged briefly and supernatant removed. Beads were washed three times with nuclear RIP dilution buffer and the supernatant was removed at the final wash.

Sheared and cross-linked chromatin was rapidly thawed at 37°C and centrifuged at 10,000 x g for 10 minutes at 4°C. The supernatant was transferred to a new tube and a 1:1 mixture of DNase I and DNase I supplement was added to each tube and was allowed to incubate at 37°C for 20 minutes. EDTA (0.5M) was added to each tube to stop the reaction and the chromatin was centrifuged once more (10,000 x g for 10 mins at 4°C).

To each tube containing antibody bound to protein A/G magnetic beads, 450 uL RIP immunoprecipitation buffer (with fresh PIC and RNase inhibitor) was added to resuspend the beads. Next, 45 uL of DNase-treated crosslinked and sheared chromatin was added to each tube and 5 uL of the remaining chromatin was removed for use as 10% input sample. All tubes were incubated over night at 4°C with rotation. The next day, tubes were centrifuged briefly, and beads were separated from the supernatant on a magnetic stand. Upon removal of the supernatant, beads were washed twice with ice-cold low salt wash buffer, once with ice-cold lithium chloride (LiCl) wash buffer and once with ice-cold Tris-EDTA (TE) buffer. The supernatant was removed at the last wash.

Nuclear RIP elution buffer with 10% SDS and proteinase K was added to the washed beads which were then incubated at 60°C for 30 mins with rotation. Samples were then cooled to room temperature, centrifuged briefly, and placed on the magnetic stand. The supernatant was added to a new tube for RNA purification.

2.10 MicroRNA Extraction

The mirVana miRNA isolation kit (Thermo Fisher Scientific) was used to isolate total RNA from MDA-MB-468 and SUM149 cells treated with anti-NRAD1 GapmeRs and with retrovirus-mediated ALDH1A3 knockdown to assess changes in the miRNA landscape in the context of NRAD1 or ALDH1A3 knockdown. Knockdowns were confirmed using RT-QPCR and the mirVana-purified RNA

was sent for microarray analyses (Affymetrix GeneChip miRNA 4.0 Array) at The Centre for Applied Genomics (TCAG) at the Hospital for Sick Children (Toronto, ON).

2.11 Western Blotting

For S100A8 knockdown quantification at the protein level, approximately 5.0 x 10^5 MDA-MB-468 and SUM149 cells harbouring each shRNA and their respective GipZ controls were seeded in 6-well dishes, trypsinized, and washed twice with ice-cold PBS. Cell pellets were resuspended in 100 uL RIPA buffer (Table 5) with fresh protease inhibitors and lysed on ice for 10 minutes. Cell lysates were then sonicated for 5 minutes (15S ON, 45S OFF, 40% amplitude, 4°C) and boiled in Laemmli buffer at 95°C for 10 minutes. 20uL of each sample was loaded into a 4-15% Mini-PROTEAN TGX Stain-free gel (BioRad) and run at 100V for 1 hour. The gel was then transferred to a polyvinylidene fluoride (PVDF) membrane using a BioRad TurboBlot system (low molecular weight setting: 1.3A-25V, 5 minutes). Following the transfer, the PVDF membrane was blocked overnight at 4°C with 5% skim-milk solution. The membrane was then probed using the primary antibody (anti-Rabbit S100A8 antibody, Abcam, ab92331, 1/5000) overnight at 4°C (Table 3). Secondary peroxidase-conjugated AffiniPure goat anti-rabbit IgG (Jackson ImmunoResearch Laboratories, Inc., 1/1000) was used to detect protein levels via chemiluminescence after staining with ECL substrate (BioRad) on a BioRad ChemiDoc imaging system (Table 3).

2.12 Microarray Analysis

For microarray analysis, MDA-MB-468 cells were treated with either control GapmeR or GapmeR#4 for 48 hours, collected in TRIzol reagent, and RNA purification was performed to isolate total RNA (n =3). For control shRNA vs. shALDH1A3 MDA-MB-468 cells, the cells were grown to confluency, collected in TRIzol reagent, and RNA purification performed (n=3). Samples were reverse transcribed into cDNA and hybridized to an Affymetrix Human Gene 2.0 ST microarray platform, where

gene expression differences were quantified. Data was collected in raw CEL format and differential gene expression was visualized using the Transcriptome Analysis Console (Affymetrix).

For miRNA microarray analysis, MDA-MB-468 or SUM149 cells were treated with either the control GapmeR, GapmeR#3, or GapmeR#4 for 48 hours, at which time, total RNA enriched for miRNAs was immediately extracted using the mirVana miRNA isolation kit (n=3). For ALDH1A3 knockdown, MDA-MB-468 or SUM149 ALDH1A3 knockdown cells (control, 61, or 63) were grown to confluency and were seeded in 6-well plates for mirVana miRNA isolation the following day (n=3). Samples were reverse transcribed to an Affymetrix GeneChip miRNA 4.0 Array platform, where miRNA expression differences were quantified. Data was collected in raw CEL format and differential miRNA expression was visualized using the Affymetrix Transcriptome Analysis Console.

2.13 Subcellular Fractionation

To fractionate cells into nuclear and cytoplasmic fractions, 2×10^6 MDA-MB-468 cells were grown to confluency, trypsinized, and resuspended in ice-cold fractionation lysis buffer (100 uL per 1 000 000) cells; components in Table 5). The tube was inverted gently to mix, then placed on ice to incubate for 5 min. The sample was then centrifuged at 5000 RCF for 5 min at 4°C to pellet nuclei. The supernatant (cytoplasmic fraction) was transferred to a new RNase-free tube, and EDTA was added to a final concentration of 10 mM. An appropriate amount of TRIzol reagent was added to each fraction (1 mL for nuclei, 0.8 mL for cytoplasmic fraction). RNA purification was then performed as previously discussed, and lncRNA expression was assessed in the fractions by RT-QPCR, using gene specific primers for NRAD1, NEAT1 (positive nuclear control), and DANCR (positive cytoplasmic control). Primer sequences are in Table 2.

2.14 Bioinformatics

RNA-sequencing expression data (RNA-seq RSEM V2 format) from patient breast tumors (breast invasive carcinoma) was extracted from the TCGA (The Cancer Genome Atlas) Cell, 2015 dataset via

cBioPortal[293,294]. RNA-sequencing data was accessed from 816 patients with any tumor subtype, 107

patients with basal-like tumors, or 97 patients with TNBC tumors via cBioPortal. Similarly, RNA-

sequencing data from patient ovarian tumors (serous ovarian cystadenocarcinoma) was extracted from the

TCGA PanCancer Atlas dataset (n=585 patients) via cBioPortal. Data from gene expression profiling via

microarray (Affymetrix Human Genome U133 Plus 2.0 Array) assessing genome-wide changes in gene

expression between normal patient samples and ovarian tumor patient tissues (serous ovarian cancer) was

extracted from the NCBI Gene Expression Omnibus (GEO) (GSE38666)[352]. NRAD1 transcript sequences

were accessed using the Ensembl genome browser[353]. MiRNA sequences were accessed using

miRBase[354]. TargetScan was used to identify predicted mRNA targets of miRNAs[355].

2.15 Statistics

All statistical analyses were performed in GraphPad Prism 9 or using R-studio. Statistical tests

varied based on the experimental setup. In all cases where two samples are compared, a student's t test

was performed. When three or more samples were compared, a one-way ANOVA with Dunnett's post-

test was performed. An alpha of 0.05 was used for all statistical tests and correlations. Significance is

listed as follows: * = p< 0.05, ** = p<0.01, *** = p<0.001, **** = p<0.0001.

Table 2. QPCR primer sequences.

Gene	Forward primer sequence (5'-3')	Reverse primer sequence (5'-3')
PUM1	GGCGTTAGCATGGTGGAGTA	CATCCCTTGGGCCAAATCCT
ARF1	GTGTTCGCCAACAAGCAGG	CAGTTCCTGTGGCGTAGTGA
B2M	AGGCTATCCAGCGTACTCCA	CGGATGGATGAAACCCAGACA
18S rRNA	GGATGTAAAGGATGGAAAATACA	TCCAGGTCTTCACGGAGCTTGTT
GAPDH	GGAGTCAACGGATTTGGTCGTA	TTCTCCATGGTGGTGAAGAC
NRAD1 common region	CCAGGGGATAAAACCCGCT	TAAGCACCAAGTCACGCTGC
NRAD1 Long transcript	CAGATAAGCCTCTCGGCACC	TTTGCCCTTCCCTGTTACCC
NRAD1 Short transcript	GCAGGCACTCGAGGCTTTA	TAGACCCATATCCTCAAGGCTTTT
DANCR	AGGAGTTCGTCTCTTACGTCT	TGAAATACCAGCAACAGGACA
NEAT1	CCTCCCTTTAACTTATCCATTCAC	TCTCTTCCTCCACCATTACCA
ALDH2	ACAATGGCAAGCCCTATGTC	ACAGGTTCATGGCGTGTGTA
CYP27A1	GAGGGCAAGTACCCAGTACG	GTACCAGTGGTGTCCTTCCG
PDZK1IP1	TCCTGACCGTCGGAAACAAG	TCGGGCACATTCTCATAGGC
RARRES3	GGCTGTTGCTATCGGGTCAA	GACCAACCATCTCCTTCGCA
SLC15A1	CAAGTGCATCGGTTTTGCCA	CTCTTTAGCCCAGTCCAGCC
CXCL17	AGGCCAGGCTTCTAGGAGAT	GGGGCTCTCAGGAACCAATC
CYP4Z1	GAATCCTGGGTTGGTCGAGG	AGGTTTCACAATCTGGCGGT
CTNNAL1	CCATGATGGCTCTCTTAGTCCA	ACCCATCCGTTATTTTCCATCTGA
EIF5A2	AGAACGGCTTCGTGGTACTG	CGTGCTTTCCCGTCTTGGA
FOSL1	CTGGTGCCAAGCATCAACAC	ACTGAGGGTAGGTCAGAGGC
FSCN1	GCAAGAATGCCAGCTGCTAC	ACAAACTTGCCATTGGACGC
GDF15	TCCAGATTCCGAGAGTTGCG	CGAGGTCGGTGTTCGAATCT
GNG11	CCTGCCCTTCACATCGAAGAT	TCTCTGCAACTTCACTTCTTTGC
ID4	AGTGCGATATGAACGACTGCTAT	TTGCTGACTTTCTTGTTGGGC
IL7R	TTCTCTGTCGCTCTGTTGGTC	ACTGGGCCATACGATAGGCT
SERPINE2	ATTGAACTGCCCTACCACGG	GTGTGGGATGATGGCAGACA
FGF2	GCGACCCTCACATCAAGCTA	AGCCAGGTAACGGTTAGCAC
SLITRK6	TGCTGCAGGGATAGTGGTTC	TGCACAGGACTGTTGTCTCTC
S100A8	TGCTAGAGACCGAGTGTCCT	TGCCACGACATCTTTATCA
ALDH1A3	TCTCGACAAAGCCCTGAAGT	TATTCGGCCAAAGCGTATTC
NEAT1 (RIP)	CTTCCTCCCTTTAACTTATCCATTCAC	CTCTTCCTCCACCATTACCAACAATAC

48

Table 3. Antibody information.

Antibody	Details	Concentrations used
Anti-S100A8	Anti-Rabbit S100A8 antibody. Abcam. Catalogue number: ab92331	WB: 1/5000 RIP: 5ug
Goat Anti-Rabbit IgG (HRP-conjugated secondary antibody)	Peroxidase-Affinipure Goat Anti-Rabbit IgG. Jackson ImmunoResearch. Catalogue number: 111-035-003	WB: 1/1000
Rabbit IgG	Normal Rabbit IgG. Millipore Sigma. Catalogue number: 12-370.	RIP: 1ug
Anti-EZH2 (positive control)	Anti-EZH2 Clone AC22. Millipore Sigma. Catalogue number: 03-900.	RIP: 5ug
Normal Mouse IgG	Normal Mouse IgG. Millipore Sigma. Catalogue number: 12-371.	RIP: 1ug

Table 4. ChIRP-MS probe sequences.

Probe number	Sequence (5'-3') + (3' biotin-TEG)
1	AAATTCTCCTGGTTTGCTTT [3BIOTEG]
2	AGGCTTATCTGAGGGAGATT [3BIOTEG]
3	GTTCCAGCAATCATTGACAC[3BIOTEG]
4	CACAGTCAGATCGTAACACC[3BIOTEG]
5	CTGGGAAAGGACATCATGGT[3BIOTEG]
6	ACTTGACCTCTATGACTCTG[3BIOTEG]
7	TAACAGAACACATACGCCCT[3BIOTEG]
8	ATAAGTAGGTGTGAGGCTGT[3BIOTEG]
9	TAGTGACTGCCTTTGACATT[3BIOTEG]
10	CTATCTTCCGTATGTAGTTT[3BIOTEG]
11	AACTGCCATAACTTTGGACC[3BIOTEG]
12	CATGAAGTCCCAAATTCCAC[3BIOTEG]
13	CTCAGAAGTCCATCTTTGTT[3BIOTEG]
14	CCCTGAACCAATGAGGAAGA[3BIOTEG]
15	AGATTTCTTTCTGCATTGGC[3BIOTEG]
16	CCAGGAACCACGAAGACAAG[3BIOTEG]
17	TTGGCTCCAGTCTTGATTTC[3BIOTEG]
18	GACAGAAGAAATACCCCTGC[3BIOTEG]
19	GTGGTTTGCTTGAAGAAGCA[3BIOTEG]
20	ATGCCTAGAATTGAGGAGAA[3BIOTEG]
21	AGGCACATACTATGTGCATT[3BIOTEG]
22	ATGACTTCTCAGAGCTTAGG[3BIOTEG]
23	TAACAGGTTGTTTGGTGACC[3BIOTEG]
24	GAGCAAAGGACCTGATGTGA[3BIOTEG]
25	TGGAGAGGCAAATGGAGGTA[3BIOTEG]
26	AGATACATTCTGATCCCATA[3BIOTEG]

Table 5. Components of buffers generated in the lab.

Protocol	Buffer	Components
ChIRP-MS	Cell Lysis Buffer	50mM Tris-HCl pH 7.0, 10mM EDTA, 1% SDS. Add fresh: 1mM AEBSF, 100X PIC, 100X RNase inhibitor.
	Hybridization Buffer (made fresh each use)	750mM NaCl, 1% SDS, 50mM Tris-HCl pH 7.0, 1mM EDTA, 15% formamide. Add fresh: 1mM AEBSF, 100X PIC, 100X RNase inhibitor.
	Wash Buffer	2X SSC (diluted from 20X SSC), 0.5% SDS. Add fresh: add 1mM AEBSF, 100X PIC, 100X RNase inhibitor.
	Proteinase K Buffer	100mM NaCl, 10mM Tris-HCl pH 7.0, 1mM EDTA, 0.5% SDS. Add fresh: 1mM AEBSF, 100X PIC, 100X RNase inhibitor, 5% v/v proteinase K.
	Biotin Elution Buffer	12.5mM D-biotin, 7.5mM HEPES pH 7.5, 75mM NaCl, 1.5mM EDTA, 0.15% SDS, 0.075% sarkosyl, 0.02% Na-Deoxycholate.
LncRNA Pulldown	Nuclear Isolation Buffer	1.28M sucrose, 40mM Tris-HCl pH 7.5, 20mM $MgCl_2$, 4% Triton X-100
	Buffer A	150mM KCl, 25 mM Tris pH 7.4, 5mM EDTA, 0.5mM DTT, 0.5% NP40 (Igepal). Add fresh: 1mM AEBSF, 100X PIC, 100X RNase inhibitor.
Western Blotting	RIPA Buffer	50mM Tris-HCl pH 7.4, 1% Igepal, 0.25% Na-Deoxycholate, 150mM NaCl, 1mM EDTA. Add fresh: 100X PIC.
Subcellular Fractionation	Hypertonic Lysis Buffer	10 mM Tris pH 7.5, 10 mM NaCl, 3 mM MgCl2, 0.3% v/v NP-40, 10% v/v glycerol. Add fresh: 100X RNase inhibitor.
Silver Staining for Mass Spectrometry	Fixation Solution	40% Ethanol (95%), 10% Glacial acetic acid.
	Sensitizing Solution	37.5mL 95% Ethanol, 5mL (0.05g/mL) Sodium thiosulphate, 8.5g Sodium Acetate, ddH_2O up to 125mL total vol.
	Silver Nitrate Solution	12.5 mL Silver Nitrate (0.025g/mL), ddH_2O up to 125mL total vol.
	Developing Solution	3.125g Sodium Carbonate, 0.025mL 37% Formaldehyde, ddH_2O up to 125mL total vol.
	Stopping Solution	10 mL 0.5M EDTA, ddH_2O up to 125mL total vol.

CHAPTER 3 – RESULTS DATA CHAPTER 1

3.1 NRAD1 is a predominately nuclear lncRNA with transcripts that have distinct subcellular localization

LncRNA genes typically encode multiple transcript variants that exhibit unique tissue expression profiles and associated biological functions. Understanding the roles of specific lncRNA transcript variants will critically inform targeted efforts to abrogate the functions of oncogenic lncRNAs. The lncRNA NRAD1 has two transcript variants expressed with high confidence that the Marcato lab has also cloned and sequenced[302], which from here will be referred to as the short (2521bp) and long (3208bp) transcript variants (Fig. 3).

The initial manuscript that characterized the functions of NRAD1 in TNBC by Vidovic and colleagues (2020) did not distinguish between the long and short transcript for the phenotypic analyses but did focus on the long transcript variant for the chromatin binding characterization assays. The subcellular localization of a lncRNA often defines its biological activity[257]. Here, we showed that subcellular fractionation assays coupled with RT-QPCR revealed that the long NRAD1 transcript variant is primarily nuclear, mirroring the localization of positive control lncRNA NEAT1 and opposing that of lncRNA DANCR, the negative control in MDA-MB-468 cells (Fig. 4A, B, C, and D). The nuclear localization of NRAD1 agreed with chromatin binding function associated with the long transcript[302] . However, the short transcript variant, although mostly nuclear, exhibited significantly higher expression in cytoplasmic fractions relative to the long transcript variant (Fig. 4B and D). Thus, the presence of different transcript variants, particularly the short transcript, is consistent with potential NRAD1 function in the cytoplasm.

The first data chapter of this work will focus on further exploration of the roles of the long NRAD1 transcript variant in nuclear modulation of gene expression. Changes in the potential NRAD1-associated miRNA landscape in response to GapmeR-mediated NRAD1 knockdown will be investigated in the second data chapter of this manuscript.

To illustrate the potential for GapmeR-mediated knockdown of each NRAD1 transcript variant,

MDA-MB-468 and SUM149 cells were treated with two separate LNA GapmeRs (GapmeR #3 and GapmeR #4) and a negative control (scramble) GapmeR followed by RT-QPCR using primers specific to the short transcript variant and the genomic region common to both sequences (general) (n=3). GapmeR-mediated knockdown of the long transcript was previously validated by Vidovic and colleagues[302]. Treatment of MDA-MB-468 and SUM149 cells with each GapmeR resulted in significant reduction of short and general NRAD1 transcripts relative to the negative control GapmeR, showing that anti-NRAD1 GapmeRs #3 and #4 can inhibit each NRAD1 transcript variant (Fig. 5A, B, C, and D).
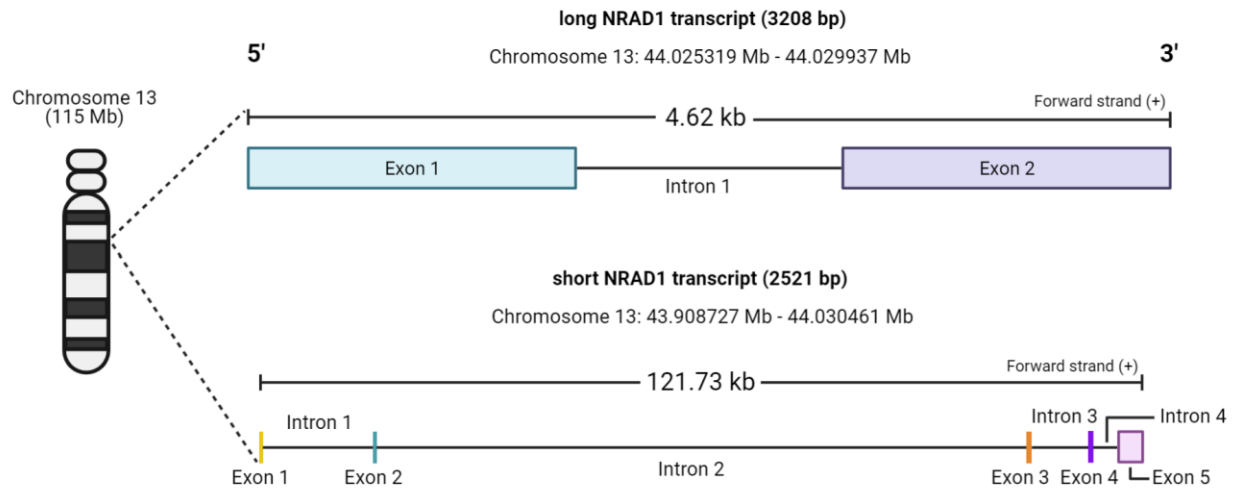
Figure 3. Localization of long (3208bp) and short (2521bp) NRAD1 transcripts on human chromosome 13. Transcript data was modeled from information accessed through the Ensembl database. Figure was created using *Biorender*.
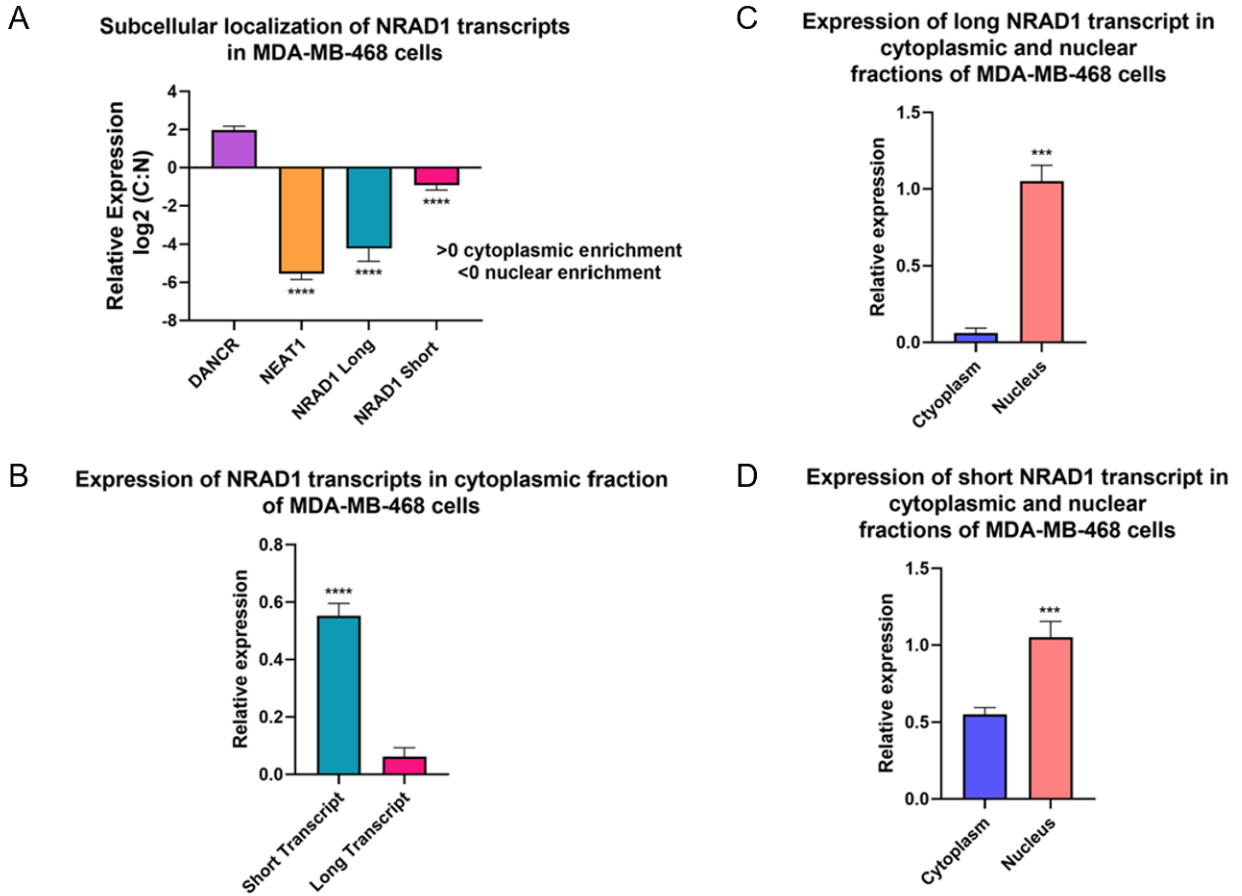
Figure 4. NRAD1 transcripts exhibit distinct subcellular localization in MDA-MB-468 cells. (A) QPCR analysis of nuclear and cytoplasmic fractions of MDA-MB-468 cells using primers specific to the cytoplasmic lncRNA DANCR, the nuclear lncRNA NEAT1, and both long and short NRAD1 transcripts. Significance was determined using a one-way ANOVA. **(B)** Expression levels of short and long NRAD1 transcripts were assessed in MDA-MB-468 cell cytoplasmic fractions via QPCR. **(C)** QPCR analysis of long NRAD1 transcript variant in cytoplasmic and nuclear fractions of MDA-MB-468 cells. **(D)** QPCR analysis of short NRAD1 transcript in cytoplasmic and nuclear fractions of MDA-MB-468 cells. Significance was determined using the student's t-test with Welch's correction for B, C and D (*** = p<0.001, **** = p<0.0001). Relative expression versus GAPDH is shown for each analysis (n=4). Error bars represent standard deviation.

NRAD1 expression following GapmeR treatment in MDA-MB-468 and SUM149 cells
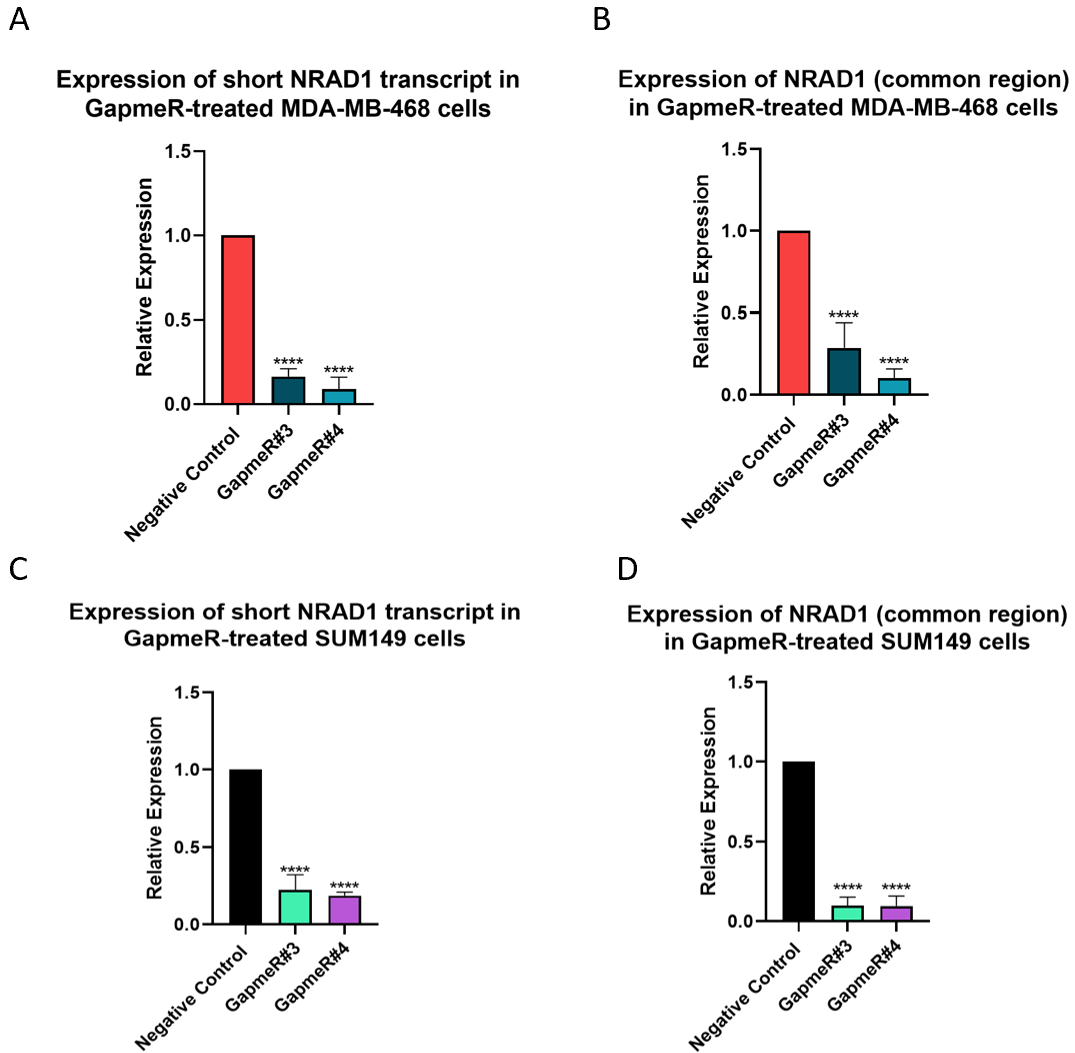
Figure 5. NRAD1-specific GapmeRs successfully knockdown NRAD1 transcripts in MDA-MB-468 and SUM149 cells. QPCR analysis of short (n=3) and general (n=7) NRAD1 transcripts treated with NRAD1-specific GapmeRs or control GapmeR in MDA-MB-468 and SUM149 cells. Expression is shown relative to the negative control GapmeR, where ARF and PUM were used as reference genes. Significance was determined using a one-way ANOVA (** = $p<0.01$, *** = $p<0.001$, **** = $p<0.0001$). Error bars represent standard deviation.

3.2 lncRNA directed proteomics methods identify putative NRAD1-binding proteins

Several studies have previously shown that some oncogenic lncRNAs promote cancer progression through interactions with protein binding partners[289]. While NRAD1 was shown to promote cell survival, tumor growth and cancer stemness in TNBC, the mechanism through which it exerts these functions remains enigmatic[302]. Intriguingly, NRAD1 mediated gene regulation was associated with NRAD1 binding to chromatin in 60% of instances[302]. I hypothesize that NRAD1 chromatin binding and gene regulation is mediated by NRAD1-binding proteins. I therefore performed two RNA-directed proteomics assays (RNA pulldown and ChIRP followed by mass spectrometry) to identify potential NRAD1 nuclear protein partners that may contribute to its oncogenic activity via nuclear/chromatin binding function[350,351] (Figures 6 and 8).

The RNA pulldown assay was accomplished using two different *in vitro* transcribed forms of the long NRAD1 transcript. Following cloning of the long NRAD1 transcript into the pCDNA3 expression vector, distinct NRAD1 transcripts were *in vitro* transcribed from either the T7 or Sp6 promoters and biotinylated. Transcripts incubated in MDA-MB-468 cell nuclear lysates were pulled down and subjected to mass spectrometry (MS) (Fig. 6A and B). MS analyses revealed 184 proteins commonly pulled down by T7 and Sp6-synthesized NRAD1 transcripts, which were of interest for further analysis (Figure 7) Appendix 1).

To provide further support for the results of the NRAD1 pulldown, I performed an additional RNA-directed proteomics experiment, chromatin isolation by RNA purification followed by mass spectrometry (ChIRP-MS, Fig. 8A). Proteins identified as potential NRAD1 interactors in the RNA pulldown and ChIRP-MS experiments would be of particular interest, given their identification in two separate NRAD1-directed proteomics experiments. As NRAD1 may regulate gene expression through chromatin interactions, I performed ChIRP-MS, a method based on ChIRP, to enhance the capture of proteins that may interact with chromatin. ChIRP is based on the capture of target lncRNA: chromatin complexes with tiling antisense probes, producing a landscape of genomic binding regions[356]. ChIRP-MS aims to identify the lncRNA-binding proteins associated with lncRNA: chromatin complexes (Fig. 8A).

NRAD1 ChIRP-MS was first performed with and without the addition of RNase A to demonstrate the ability of ChIRP-MS to retrieve a specific RNA (Fig. 8B). Specifically, expression of the NRAD1 transcript was >500-fold enriched over GAPDH and 18S rRNA in RNA samples (-RNase A) retrieved with both odd and even probes (Fig. 8B). NRAD1, 18S rRNA, and GAPDH were not detected in RNase A-treated samples, demonstrating that ChIRP-MS is dependent upon RNA retrieval (Fig. 8B). Thus, we performed NRAD1 ChIRP-MS with separate odd and even probesets, eluting proteins for subsequent TCA precipitation, SDS-PAGE with silver staining, and MS analyses. Once more, QPCR on odd or even probe-bound RNA fractions revealed that ChIRP-MS successfully enriched NRAD1 over input versus reference genes B2M and ARF, providing confidence that identified proteins were associated with NRAD1 (Fig 8C). Specifically, NRAD1 expression was >200-fold enriched over B2M and ARF in RNA samples retrieved with odd and even probesets (Fig. 8C). Following MS analyses of protein samples retrieved using odd or even NRAD1-targeted probes, 5 unique proteins were identified using even probes, 14 unique proteins were identified using odd probes, and 22 proteins were identified by both probesets (Fig. 8D).

To focus our downstream analyses on promising protein targets, with which NRAD1 may interact, we were specifically interested in proteins that were pulled down by both NRAD1 transcripts in our RNA pulldown experiments and were also identified through ChIRP-MS (either even or odd probeset-retrieved NRAD1-bound proteins). This analysis demonstrated that 22 unique proteins were identified by ChIRP-MS, 297 unique proteins were identified by the RNA pulldown, and 20 proteins were identified by both methods (Fig. 9A). Of the 20 proteins identified by ChIRP-MS and the RNA pulldown, 16 were likely contaminants (Fig. 9B), and thus, we focused our subsequent analyses on four key potential protein targets: HIST1H1C (histone H1.2), HIST1H1E (histone H1.4), HRNR (Hornerin) and S100A8 (S100 calcium-binding protein A8).
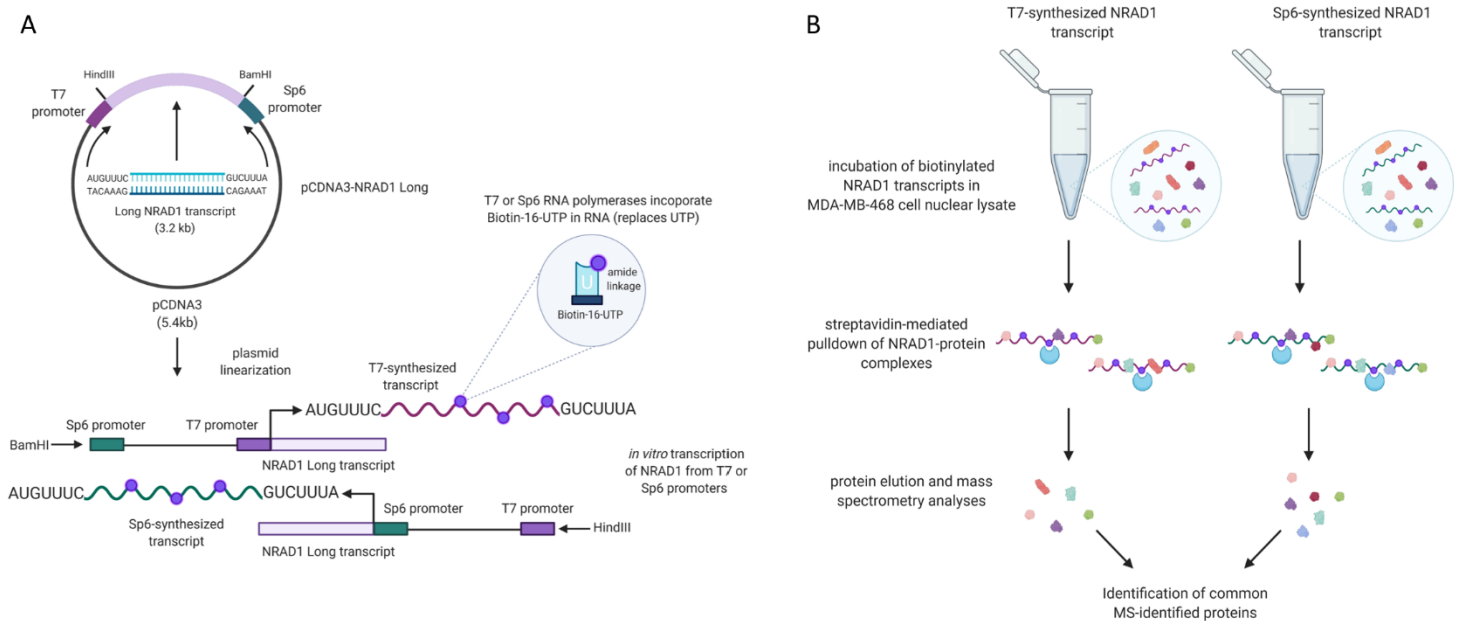
Figure 6. NRAD1 pulldown followed by mass spectrometry identifies potential NRAD1-bound proteins in MDA-MB-468 cells. (A) The DNA sequence for the long NRAD1 transcript variant was recombinantly inserted into the multiple cloning site of the pCDNA3.1 expression vector using HindIII and BamHI restriction sites. The recombinant plasmid was linearized for subsequent *in vitro* transcription with either BamHI (for transcription via the T7 promoter) or HindIII (for transcription via the Sp6 promoter) and biotinylation. T7 and Sp6 RNA polymerases can incorporate Biotin-16-UTP (biotin attached to UTP via an amide linkage) into the transcribed RNA, replacing UTP. **(B)** T7- or Sp6-synthesized biotinylated NRAD1 transcripts were separately incubated in MDA-MB-468 cell nuclear lysates, where NRAD1-protein complexes were pulled down using streptavidin-coated magnetic beads. Proteins from each transcript were eluted and subjected to mass spectrometry (MS). MS-identified proteins in common between both transcripts were of interest for further experiments. Figure was created using *Biorender*.

Proteins identified by NRAD1 pulldown in MDA-MB-468 cells

T7-transcribed NRAD1 transcript

Sp6-transcribed NRAD1 transcript

45   184   88
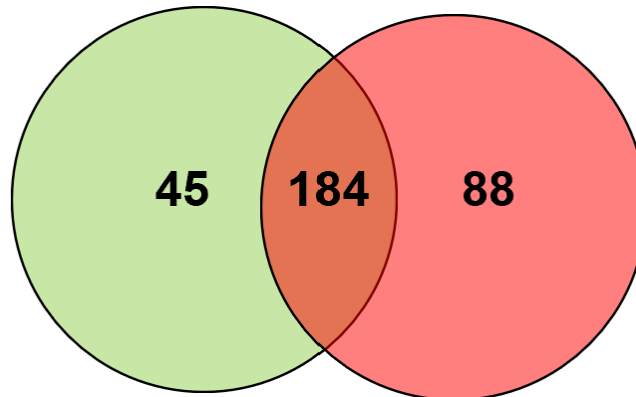
Figure 7. NRAD1 pulldown and MS analysis with transcripts in vitro synthesized from the T7 or Sp6 promoters identify common proteins. MS-identified proteins via NRAD1 pulldown with long NRAD1 transcripts *in vitro* transcribed from T7 or Sp6 promoters. 184 proteins were found to be common between differentially synthesized NRAD1 transcripts and were therefore of interest for downstream experimentation.
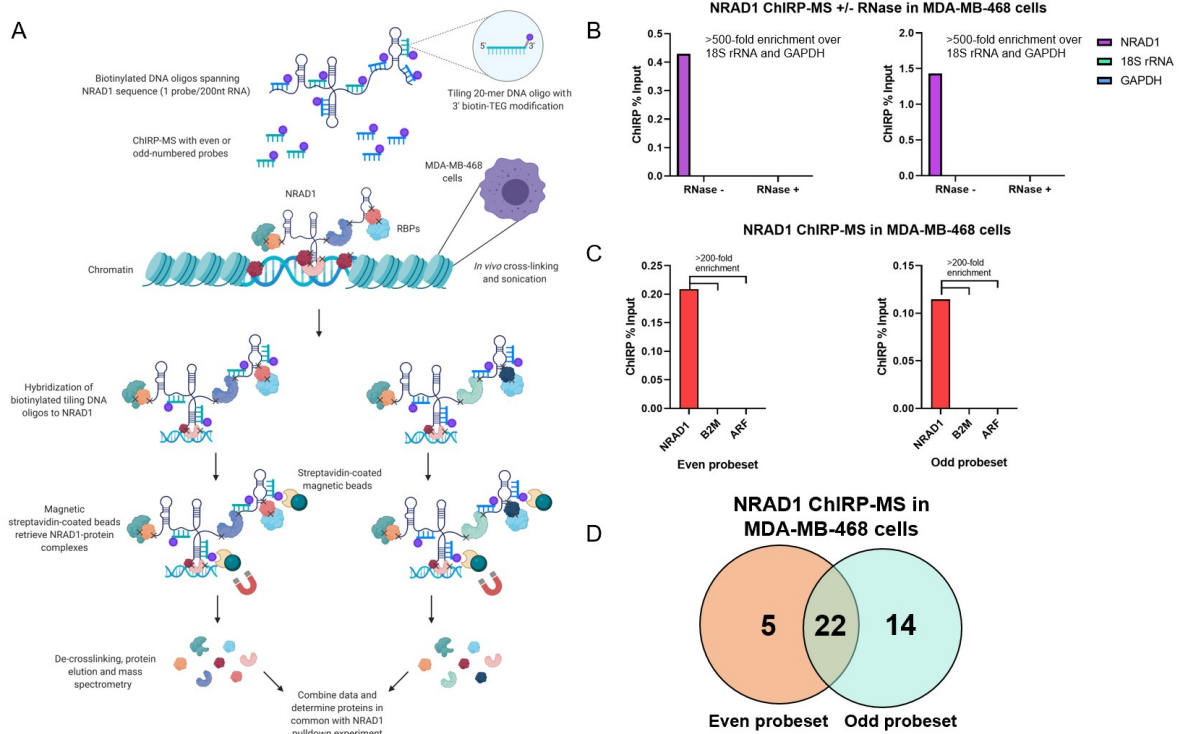
Figure 8. NRAD1 ChIRP-MS identifies potential chromatin-associated proteins bound to NRAD1 in MDA-MB-468 cells. **(A)** DNA oligonucleotides with 3' biotin-TEG modification were designed to span the sequence of NRAD1. ChIRP-MS was performed using either even or odd numbered probes in MDA-MB-468 cells. **(B)** NRAD1 ChIRP-MS with even or odd probes was performed with or without RNase treatment to illustrate successful retrieval of RNA. QPCR of odd and even probe-retrieved RNA (+ or – RNase A) was performed to demonstrate enrichment of NRAD1 over input versus reference genes GAPDH and 18S ribosomal RNA. **(C)** QPCR on odd or even probe-bound RNA fractions reveals that ChIRP-MS successfully enriches NRAD1 over input versus reference genes B2M and ARF. **(D)** Venn diagram showing the number MS-identified proteins from NRAD1 ChIRP-MS with odd or even probes. Figure was created using *Biorender*.

A

Proteins identified via NRAD1 pulldown or ChIRP-MS in MDA-MB-468 cells

ChIRP-MS   NRAD1 pulldown

22   20   297
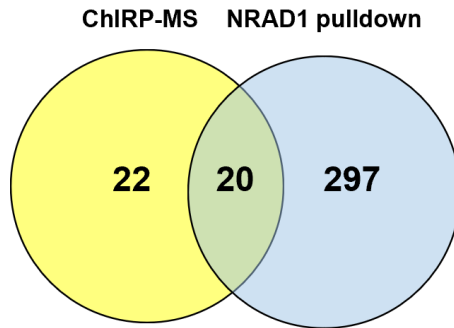
B

Gene Symbol	Gene Name
ACTB	Actin Beta
ALB	Albumin
DCD	Dermcidin
DSP	Desmoplakin
KRT1	Keratin 1
KRT2	Keratin 2
KRT5	Keratin 5
KRT6A	Keratin 6A
KRT6B	Keratin 6B
KRT9	Keratin 9
KRT10	Keratin 10
KRT14	Keratin 14
KRT16	Keratin 16
KRT17	Keratin 17
KRT74	Keratin 74
RNH1	Ribonuclease/Angiogenin Inhibitor 1
S100A8*	S100 Calcium-Binding Protein A8
HRNR*	Hornerin (S100A16/A18)
HIST1H1C*	Histone H1.2
HIST1H1E*	Histone H1.4

* Denotes protein of interest

Figure 9. NRAD1-directed ChIRP-MS and lncRNA pulldown-MS experiments identify common potential NRAD1-binding proteins. (A) Proteins identified via MS from NRAD1 pulldown or ChIRP-MS in MDA-MB-468 cells. **(B)** 20 proteins were identified in common between NRAD1 pulldown and ChIRP-MS experiments. Proteins of interest for their potential activity in NRAD1-mediated oncogenic functions are denoted with an asterisk (*) and all other proteins are likely contaminants.

3.3 S100A8 is a promising NRAD1-associated protein with a gene expression profile similar to NRAD1 in patient breast tumors.

To further prioritize the MS-identified hits for validation, we reasoned that proteins interacting with NRAD1 in a functional manner would regulate some of the same genes as NRAD1 in patient breast tumors. I therefore assessed the Spearman correlations of all genes co-expressed with NRAD1 and each of the MS-identified proteins in patient breast tumors (20,200 genes) using RNA-sequencing data accessible via cBioPortal (TCGA Breast Invasive Carcinoma, Cell 2015; n = 816, Fig. 10). This analysis revealed a high degree of overlap between the Spearman correlations of genes co-expressed with NRAD1 and those of genes co-expressed with S100A8 in patient breast tumors (Fig. 10). Specifically, heatmap clustering showed that gene co-expression patterns were the most similar between NRAD1 and S100A8. Conversely, the Spearman correlations of genes co-expressed with HRNR, HIST1H1C, or HIST1H1E did not show a high level of similarity with the Spearman correlations of the genes co-expressed with NRAD1. However, gene co-expression patterns with HRNR demonstrated some similarity to the co-expression patterns with NRAD1 and demonstrated the next highest degree of similarity to NRAD1 (Fig. 10). This similarity may be of interest given that like S100A8, HRNR is a member of the S100 protein family[357]. Together, Given the high level of similarity of the correlations of genes co-expressed with NRAD1 and those co-expressed with S100A8 in patient breast tumors, these results highlighted S100A8 as a key protein of interest for downstream experiments to confirm its possible functional interaction with NRAD1. This data may suggest partial common gene regulation between NRAD1 and S100A8 in breast cancer.

To further explore the overall portrait of genes that may be commonly regulated by NRAD1 and potential NRAD1-associated proteins identified by MS, I once again extracted breast tumor (invasive breast carcinoma) RNA-sequencing data from The TCGA (Cell 2015) dataset (n = 816) and plotted the Spearman correlations of protein coding genes (~20,200) co-expressed with each MS-identified protein (S100A8, HRNR, HIST1H1C, or HIST1H1E) against the Spearman correlations of genes co-expressed with NRAD1 (Fig. 11A, B, C, and D). The correlation coefficient (R) and p-values were calculated using

R-studio. Of the identified proteins, Spearman correlations between protein-coding genes co-expressed with S100A8 were the most highly correlated with those co-expressed with NRAD1 (R=0.76, p < 2.2E-16) (Fig. 11A), mirroring the results shown in Figure 9. Similarly, Spearman values of the genes co-expressed with HRNR also correlated positively with those of NRAD1 (R=0.42, p < 2.2E-16), although to a lesser extent (Fig. 11B). The Spearman correlations of genes co-expressed with HIST1H1C or HIST1H1E were negatively or weakly correlated with those of NRAD1 (Fig. 11C and D). Thus, given the strong relationship between genes correlated with S100A8 and those correlated with NRAD1 in patient breast tumors, S100A8 was selected as the single protein of interest for further investigation of its interactions with NRAD1.

To provide further evidence for the possible role of S100A8 in NRAD1-mediated gene regulation, we performed microarray analyses using RNA extracted from MDA-MB-468 cells with NRAD1 knockdown (control vs. GapmeR #4-treated, Fig. 12) and identified NRAD1-regulated genes that are also correlated with NRAD1 and S100A8 in patient breast tumors (Fig. 13). To this end, we identified genes that were significantly (p < 0.05) up- or down-regulated (≥ 1.2 or ≤ -1.2-fold expression change) upon NRAD1 knockdown and assessed their correlations with NRAD1 and S100A8 in patient breast tumors using RNA-sequencing data from the TCGA (Cell 2015 dataset). Of the 5852 genes from the microarray that met our cutoff criteria (p < 0.05, ≥ 1.2 or ≤ -1.2-fold expression change), we were able to match 3733 genes with RNA-seq co-expression data from patient tumors from cBioPortal. Of these, 1332 genes were identified to be upregulated by NRAD1 and positively correlated with NRAD1 and S100A8 or downregulated by NRAD1 and negatively correlated with NRAD1 and S100A8 in patient breast tumors (Fig. 13, Appendix 2). This data provides evidence that NRAD1 and S100A8 co-regulate at least some gene expression in breast cancer and provides support for the potential role of S100A8 in NRAD1-mediated gene regulation.

Spearman correlations of genes co-expressed with MS-identified proteins
in patient breast tumors (TCGA, Cell 2015)
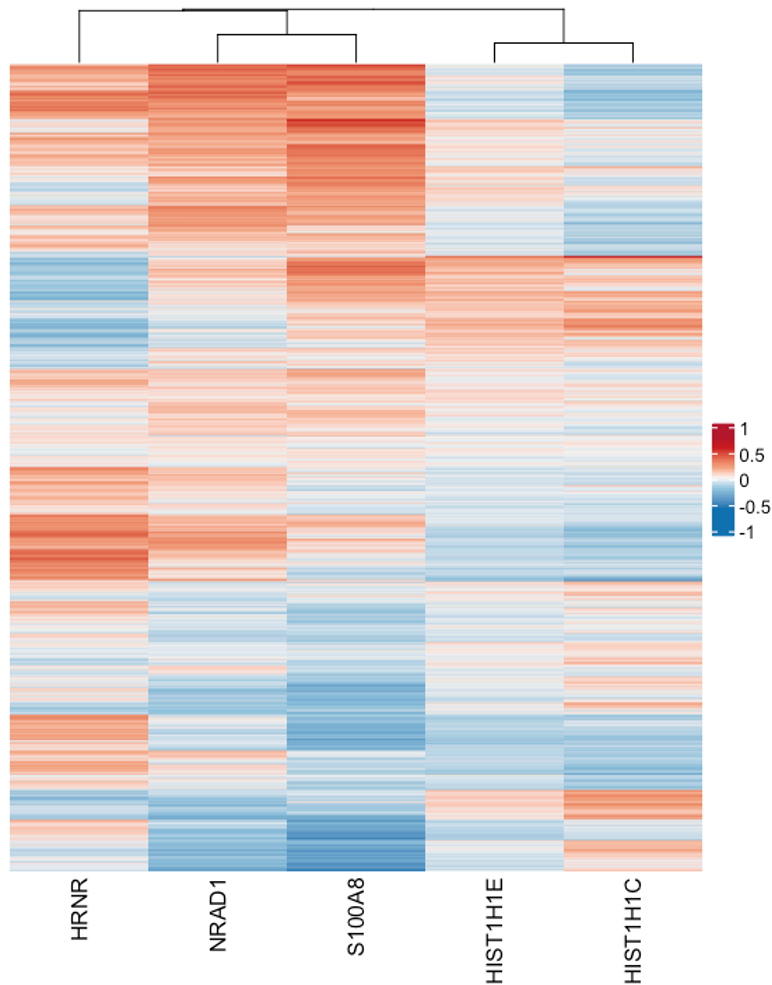
Figure 10. Correlations of genes co-expressed with S100A8 show the highest degree of similarity to those of genes co-expressed with NRAD1 in patient breast tumors. Using patient breast tumor RNA-sequencing data from the TCGA (Cell, 2015) dataset (breast invasive carcinoma; n=816) accessible via cBioPortal, Spearman correlations of all genes (~20,200) co-expressed with each MS-identified hit (S100A8, HRNR, HIST1H1C, and HIST1H1E) and NRAD1 were extracted and used to compare correlation profiles between each MS-identified hit and NRAD1 (shown in the heatmap).
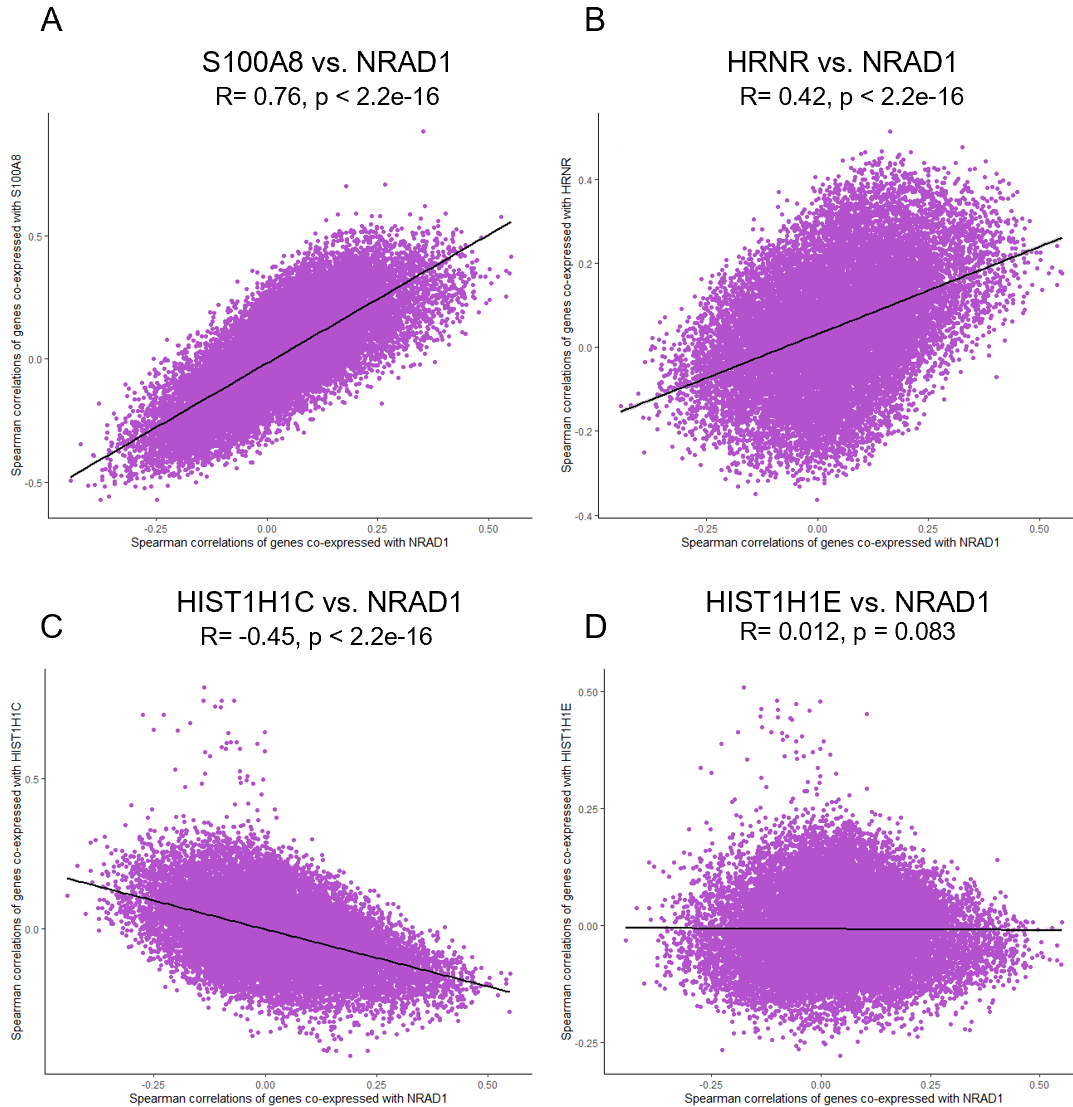
Figure 11. Spearman correlations of genes co-expressed with NRAD1 are strongly correlated with the spearman correlations of genes co-expressed with S100A8 in patient breast tumors. Using RNA-sequencing data from the TCGA (Cell, 2015) dataset (breast invasive carcinoma; n=816) accessible via cBioPortal, Spearman correlations between all protein-coding genes (~20,200) co-expressed with NRAD1 were plotted against the Spearman correlations of all protein-coding genes co-expressed with each protein of interest: S100A8 **(A)**, HRNR **(B)**, HIST1H1C **(C)**, and HIST1H1E **(D)**. The correlation coefficients and p-values were calculated using R.
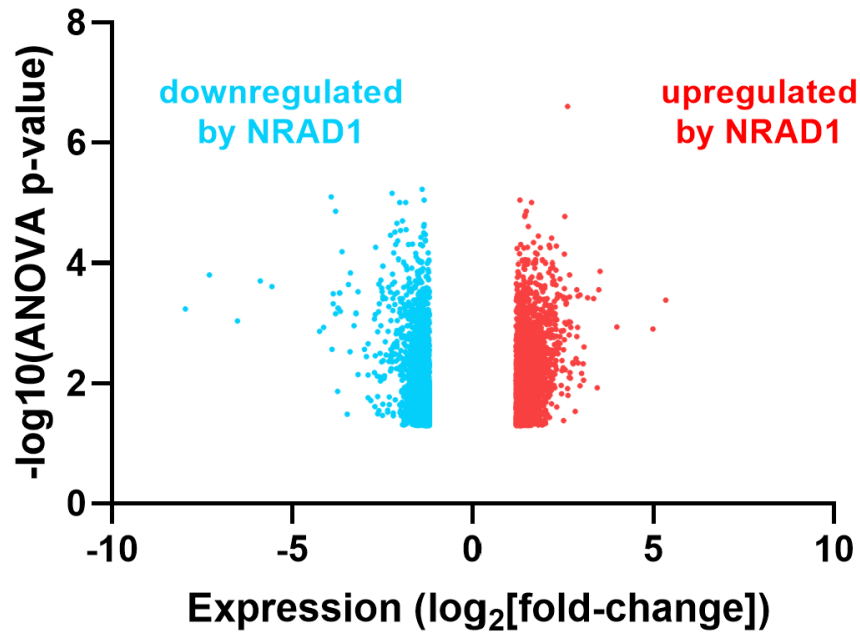
Figure 12. NRAD1-induced gene expression in MDA-MB-468 cells. Genome-wide gene expression changes induced by NRAD1 inhibition (control versus GapmeR#4-treated) is quantified in MDA-MB-468 cells using the Affymetrix Human Gene 2.0 ST microarray platform (n=3). The log2-fold change in expression is plotted versus the –log10(ANOVA p-value) of over 50,000 probes corresponding to 24,838 probesets covering 24,838 RefSeq (Entrez) genes. Only probes with a ≥1.20-fold expression change and a p-value of <0.05 are plotted. Upregulated genes are coloured red and downregulated genes are coloured blue.
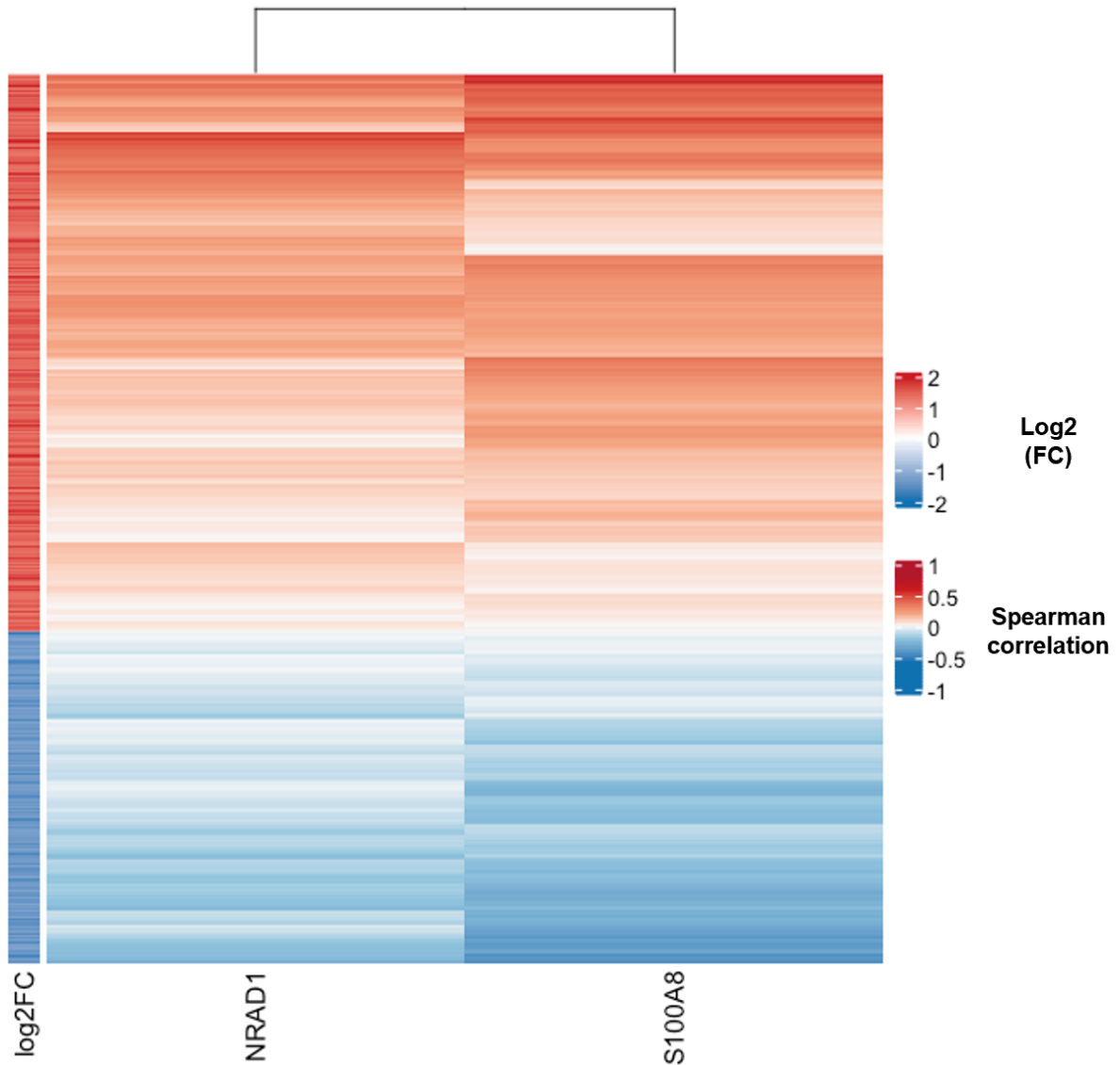
Figure 13. Numerous NRAD1-regulated genes are correlated with NRAD1 and S100A8 in patient breast tumors. NRAD1-regulated genes identified using the Affymetrix Human Gene 2.0 ST microarray platform (control GapmeR vs. GapmeR #4; n=3) in MDA-MB-468 cells ($\geq$ 1.2 or $\leq$ -1.2-fold expression change, p < 0.05) were assessed for their Spearman co-expression correlations with NRAD1 and S100A8 in patient breast tumors. To identify co-expression Spearman correlations, RNA-sequencing data (RNA-seq RSEM V2) of patient breast tumors (invasive breast carcinoma) was extracted from the TCGA, Cell 2015 dataset (n = 816), accessible via cBioPortal. The Spearman correlations of NRAD1-upregulated genes positively correlated, and NRAD1-downregulated genes negatively correlated with NRAD1 and S100A8 in patient breast tumors are shown in the heatmap. The Log$_2$ fold change in expression from the microarray is shown on the left of the heatmap.

3.4 RNA Immunoprecipitation validates the interaction of S100A8 with NRAD1 in MDA-MB-468 cells.

        Given our identification of S100A8 in two separate lncRNA-based proteomics experiments (Fig. 6 and 8), the high degree of similarity and strong correlation between the Spearman correlations of genes co-expressed with NRAD1 and those of the genes expressed with S100A8 in patient breast tumors, and the evidence for S100A8/NRAD1 co-regulation of gene expression (Fig. 10, 11A, and 13), I explored the potential interactions between S100A8 and NRAD1 in further experiments. To this end, I first performed nuclear RNA immunoprecipitation (RIP) with a S100A8-specific antibody followed by QPCR to confirm the interaction of S100A8 with NRAD1 *in vitro* (Fig. 14).

        Nuclear RIP permits the isolation and identification of RNA species associated with chromatin[358]. We performed nuclear RIP with formaldehyde cross-linking, which preserves protein: DNA, protein: protein and protein: RNA interactions, to validate the interaction of S100A8 with NRAD1. Nuclear RIP was accomplished using chromatin prepared from MDA-MB-468 cells treated with formaldehyde as a cross-linking agent. I incorporated a positive control in the assay (pulldown of lncRNA NEAT1 by anti-EZH2 antibody since NEAT1 is known to interact with EZH2[359]. As expected, NEAT1 expression was >1000-fold enriched using EZH2-specific antibody over the mouse IgG negative control (Fig. 15A). Similarly, NRAD1 expression was >100-fold enriched using S100A8-specific antibody over the rabbit IgG negative control, indicating that NRAD1 binds endogenous S100A8 (Fig. 15B). These results suggest that NRAD1 binds S100A8 in MDA-MB-468 cells and confirms our ChIRP-MS and RNA pulldown results.
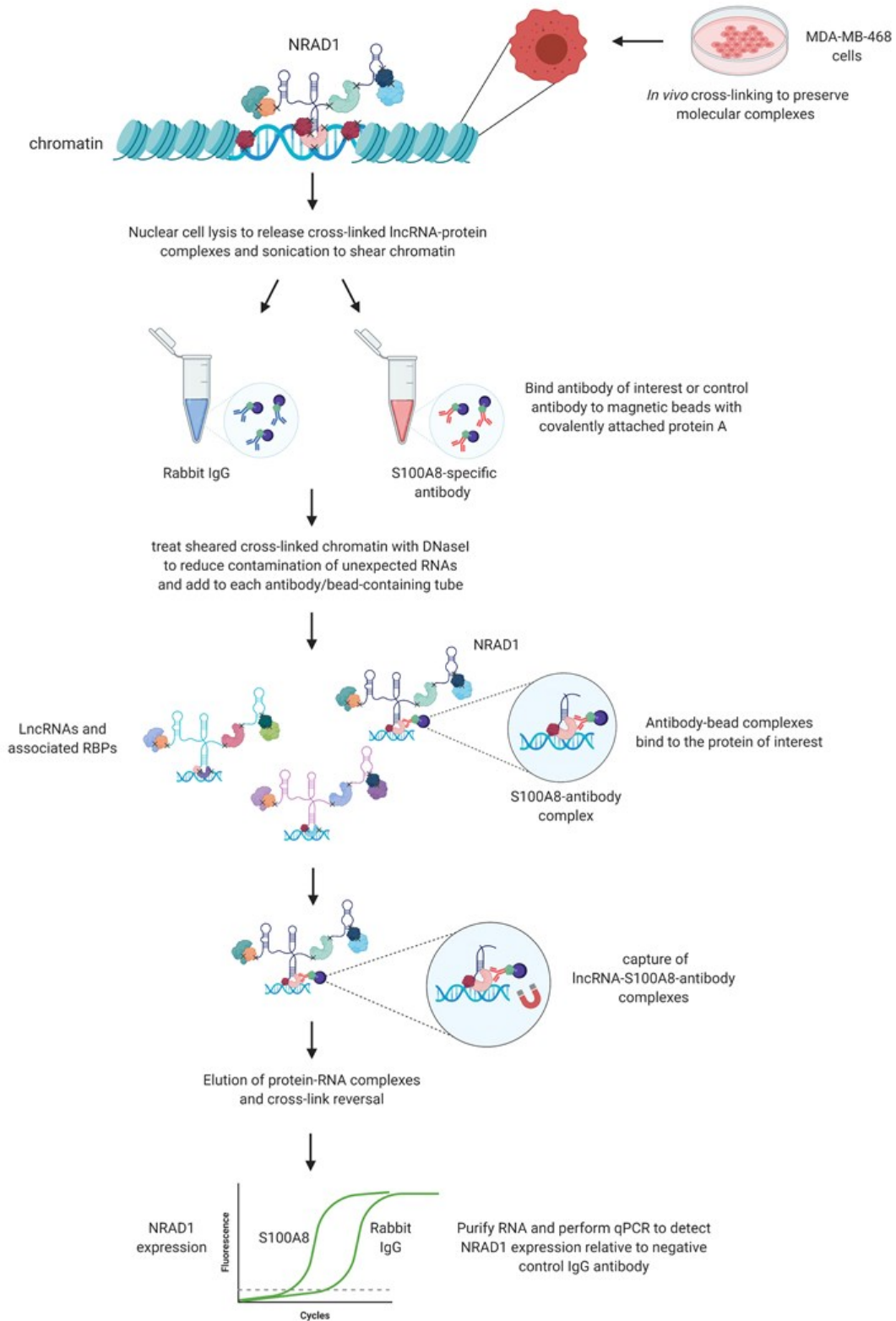
Figure 14. Nuclear RNA immunoprecipitation (RIP) validates the interaction between S100A8 and NRAD1. Graphical representation of nuclear RIP using Rabbit IgG or anti-S100A8 antibody (Biorender).

A NEAT1 IP in MDA-MB-468 cells

B NRAD1 IP in MDA-MB-468 cells
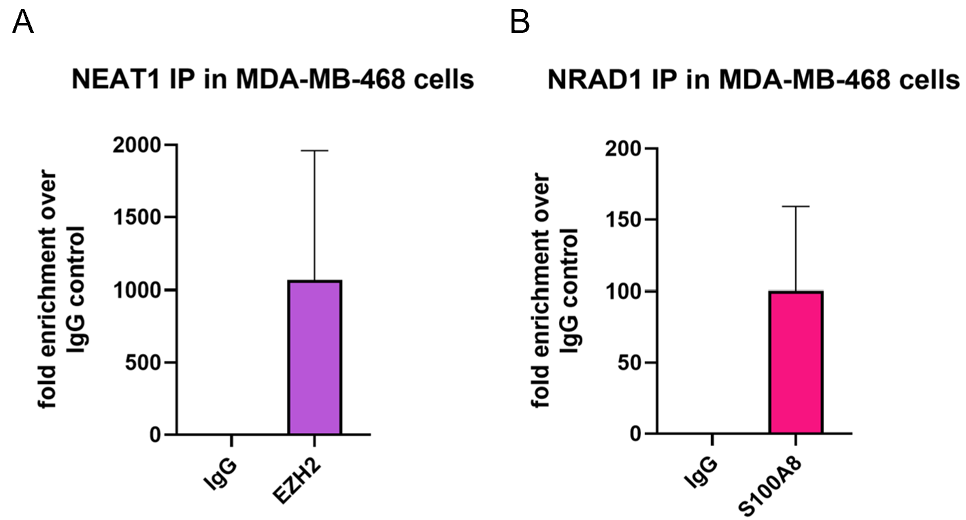
Figure 15. Nuclear RIP confirms binding of NRAD1 to S100A8. Nuclear RNA immunoprecipitation was performed using antibodies specific to S100A8 and EZH2 (positive control) or IgG negative controls (mouse IgG for EZH2 or Rabbit IgG for S100A8). QPCR was used to detect NEAT1 bound to EZH2 **(A)** or NRAD1 bound to S100A8 **(B)** relative to the respective IgG controls. Fold enrichment of NEAT1 captured by EZH2 or NRAD1 captured by S100A8 relative to respective IgG controls is shown as the mean of three technical replicates. Error bars represent standard deviation.

3.5 S100A8 contributes to NRAD1 and ALDH1A3-mediated gene regulation in MDA-MB-468 cells.

Vidovic and colleagues (2020) recently showed via microarray (control vs. GapmeR-mediated NRAD1 knockdown or control vs. ALDH1A3 knockdown) that NRAD1 and ALDH1A3 regulate some of the same genes in MDA-MB-468 cells[302] (Fig. 16A and B). Using the Affymetrix Human Gene 2.0 ST microarray platform, genome-wide gene expression changes induced by NRAD1 inhibition (control versus GapmeR#4-treated) (Fig. 16B) or ALDH1A3 knockdown (control versus shALDH1A3) (Fig. 16A) were quantified in MDA-MB-468 cells. Of the identified genes, those exhibiting a ≥1.70-fold expression change and an ANOVA p-value of <0.05 were used as a cutoff. This analysis revealed 383 genes upregulated and 232 genes downregulated by NRAD1 (Fig. 16B), and 75 upregulated and 54 downregulated by ALDH1A3 (Fig. 16A). Importantly, 7 of the 75 ALDH1A3-upregulated genes (~10%) and 13 of the 54 ALDH1A3-downregulated genes (~25%) were also upregulated or downregulated by NRAD1, respectively (Fig. 16C). This result demonstrated the role of NRAD1 in partial regulation of ALDH1A3-mediated gene expression. We reasoned that if S100A8 contributes to NRAD1 and ALDH1A3-mediated gene regulation, it may regulate the same genes as NRAD1 and ALDH1A3. We therefore knocked down S100A8 using shRNAs in MDA-MB-468 cells (Fig. 17A and B) and performed QPCR on NRAD1/ALDH1A3-regulated genes in the S100A8 knockdown cells (n=5) (Fig. 18A and B). S100A8 knockdowns were confirmed using QPCR (n=5) and Western blotting (Fig. 17A and B). Given that only one of the shRNAs (shRNA #3) was significantly knocked down at the gene level and at the protein level, the gene expression experiment was performed only with shRNA #3. Of the genes that were downregulated upon NRAD1 and ALDH1A3 knockdown, 6 of 7 were significantly downregulated and 1 was modestly downregulated (RARRES3) upon S100A8 knockdown (Fig. 18A). Of the 13 genes that were upregulated upon NRAD1 and ALDH1A3 knockdown, I assessed the expression of 11 via QPCR in S100A8 knockdown cells, where 9 of 11 were significantly upregulated, 1 was modestly upregulated (FGF2) and one remained unchanged (SLITRK6) (Fig. 18B). Together, this data supports the role of S100A8 in NRAD1/ALDH1A3-mediated gene regulation.
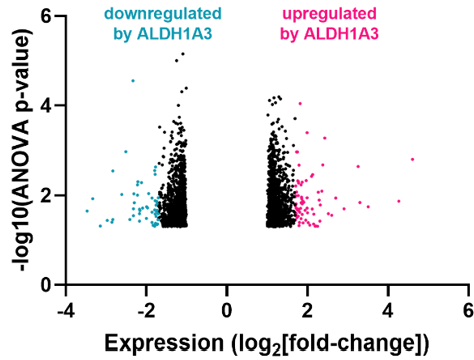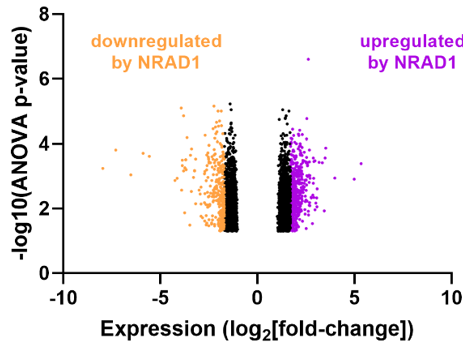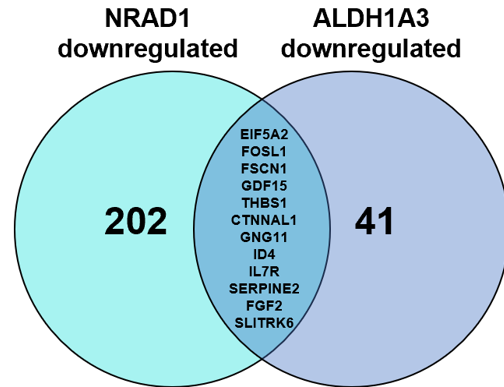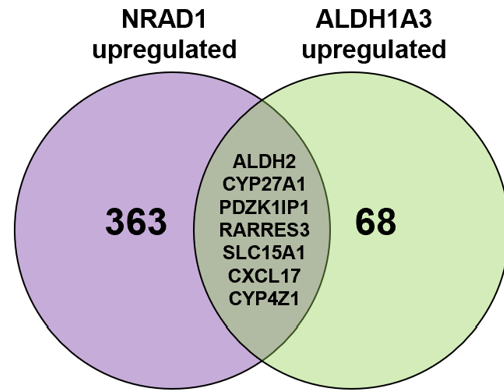
Figure 16. NRAD1 and ALDH1A3 co-regulate gene expression in MDA-MB-468 cells. Genome-wide gene expression changes induced by NRAD1 or ALDH1A3 knockdown (control vs. knockdown) were previously quantified by Vidovic and colleagues (2020) in MDA-MB-468 cells using the Affymetrix Human Gene 2.0 ST microarray platform (n=3). **(A)** Genome-wide gene expression changes induced by NRAD1 inhibition (control versus GapmeR#4-treated) quantified in MDA-MB-468 cells using the Affymetrix Human Gene 2.0 ST microarray platform (n=3). The log2-fold change in expression is plotted versus the –log10(ANOVA p-value) of over 50,000 probes corresponding to 24,838 probesets covering 24,838 RefSeq (Entrez) genes. Only probes with an ANOVA p-value of < 0.05 are plotted. Probes with a ≥1.70-fold expression change (p < 0.05) are coloured. **(B)** Genome-wide gene expression changes induced by ALDH1A3 knockdown (control vs. ALDH1A3 knockdown) quantified in MDA-MB-468 cells using the Affymetrix Human Gene 2.0 ST microarray platform (n=3). The log2-fold change in expression is plotted against the -log10(ANOVA p-value). Only probes with an ANOVA p-value of <0.05 are plotted. Probes with a ≥1.70-fold expression change (p < 0.05) are coloured. **(C)** The number of genes up- or down-regulated by both NRAD1 and ALDH1A3 are shown with Venn diagrams, where NRAD1/ALDH1A3 co-regulated genes are found within the centres.

A

S100A8 expression in MDA-MB-468 cells with S100A8 knockdown
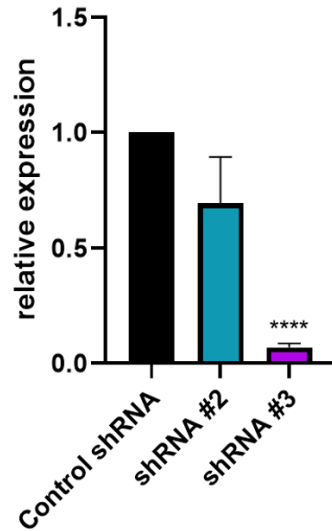
B

S100A8 Western Blot in MDA-MB-468 cells

WB: S100A8   11kDa

Total protein
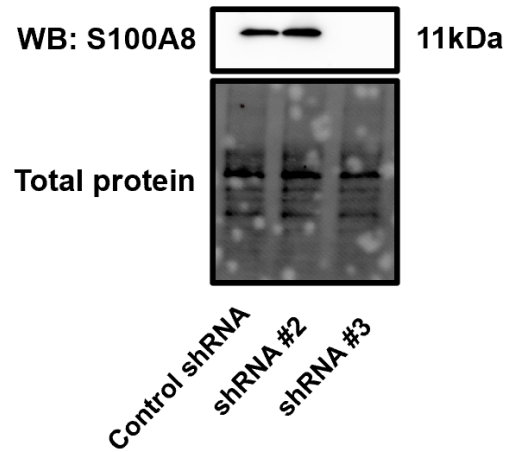
Figure 17. Confirmation of S100A8 knockdown in MDA-MB-468 cells using Western blotting and QPCR. **(A)** S100A8 knockdown in MDA-MB-468 cells was confirmed using QPCR (n=5). Expression relative to the control shRNA is shown, using ARF and PUM as reference genes. Significance was determined using a one-way ANOVA (**** = $p<0.0001$). Error bars represent standard deviation. **(B)** Western blotting was used to confirm a decrease in S100A8 at the protein level, using total protein as a loading control.

A

QPCR gene expression in MDA-MB-468 cells with S100A8 knockdown
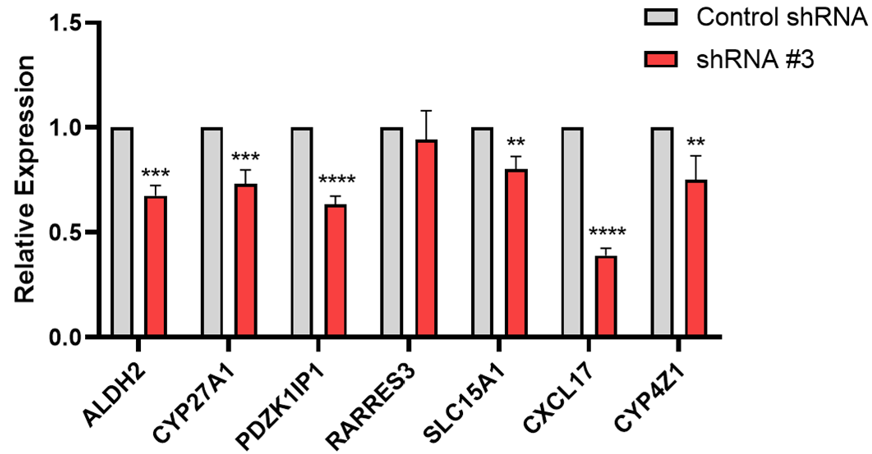
Downregulated upon NRAD1/ALDH1A3 knockdown in gene chip

B
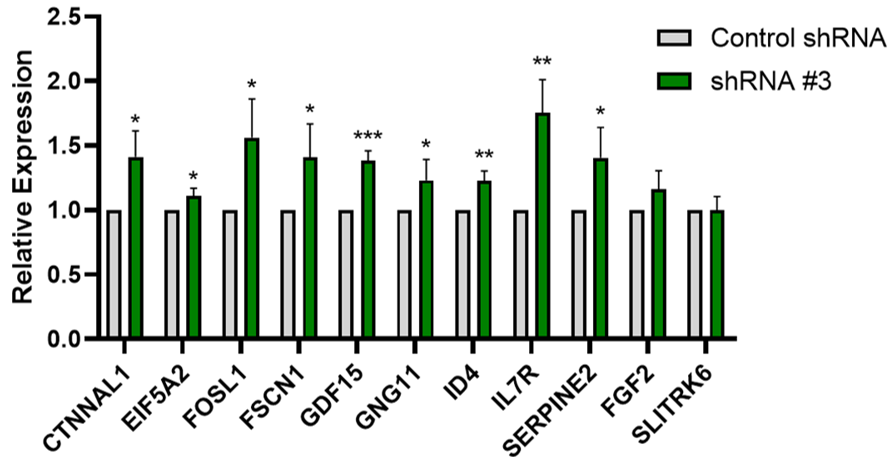
Upregulated upon NRAD1/ALDH1A3 knockdown in gene chip

Figure 18. Genes up- or downregulated upon NRAD1 and ALDH1A3 knockdown in MDA-MB-468 cells are also up- or downregulated upon S100A8 knockdown in MDA-MB-468 cells. The expression of genes previously shown to be regulated by NRAD1 and ALDH1A3 in MDA-MB-468 cells were investigated using QPCR in the context of S100A8 knockdown in MDA-MB-468 cells (n=5). **(A)** Expression of NRAD1/ALDH1A3 upregulated genes following S100A8 knockdown in MDA-MB-468 cells. **(B)**. Expression of NRAD1/ALDH1A3 downregulated genes following S100A8 knockdown in MDA-MB-468 cells. Expression relative to the control shRNA is shown, using ARF and PUM as reference genes. Significance was determined using a paired two-tail t-test (* = $p<0.05$, ** = $p<0.01$, *** = $p<0.001$, **** = $p<0.0001$). Error bars represent standard deviation.

3.6 Investigating the contributions of S100A9 and ALDH1A3 to S100A8/NRAD1-mediated gene regulation

To assess the degree of importance of the potential common gene regulation between NRAD1 and S100A8, I performed the Spearman correlation analysis between ALDH1A3 and NRAD1 (induced by ALDH1A3), which have been validated to regulate some of the same genes in TNBC cells[302]. This analysis revealed a correlation coefficient of 0.81 (p < 2.2E-16) between the spearman correlations of genes co-expressed with NRAD1 and those of ALDH1A3 in patient breast tumors in the TCGA Cell, 2015 dataset (Fig. 19A). Given that the same analyses with NRAD1 and S100A8 yielded a correlation coefficient of 0.76 (Fig. 11A), comparable to the correlation coefficients between NRAD1 and ALDH1A3, it is reasonable to suspect a similar level of common gene regulation between NRAD1 and S100A8. Due to the finding that ALDH1A3 induces NRAD1 expression in TNBC, where NRAD1 contributes to over 20% of ALDH1A3-mediated gene regulation[302], if S100A8 indeed contributes to NRAD1-mediated gene regulation, S100A8 and ALDH1A3 must commonly regulate at least some genes. Thus, I performed the Spearman correlation analysis between ALDH1A3 and S100A8 in patient breast tumors. This analysis revealed a correlation coefficient of 0.61, indicating possible common gene regulation between S100A8 and ALDH1A3 in patient breast tumors (Fig. 19B).

S100A8 often forms a heterodimeric complex with S100A9 called calprotectin which plays critical roles in inflammation and cancers[360,361]. Notably, S100A9 was identified via MS in the NRAD1-pulldown experiment using both T7 and Sp6-synthesized NRAD1 transcripts, pointing to a possible association with NRAD1 (Appendix 1). Given this information, it is possible that S100A8 and S100A9 co-operatively interact with NRAD1 to regulate gene expression. S100A8 and S100A9 expression is highly correlated in patient breast tumors, evidenced by the TCGA (*Cell, 2015*) dataset (R=0.92). Thus, we would expect a high correlation between the Spearman correlations of genes co-expressed with S100A8 and those of S100A9. Indeed, the correlation of the spearman correlations of genes co-expressed with S100A8 and the spearman correlations of genes co-expressed with S100A9 in patient breast tumors was 0.99 (p < 2.2E-16), illustrating the high level of common gene regulation between these proteins and

validating our other Spearman correlation analyses (Fig. 20A). I therefore performed the previous

Spearman correlation analyses with S100A9. Like those with S100A8, the Spearman correlations of

genes co-expressed with S100A9 were highly correlated with those co-expressed with NRAD1 (R = 0.80,

p < 2.2E-16) (Fig. 20B) and ALDH1A3 (R = 0.65, p < 2.2E-16) (Fig. 20C) in patient breast tumors,

supporting the potential for the role of the S100A8/A9 complex in NRAD1-mediated gene regulation.

To further demonstrate the potential common gene regulation between ALDH1A3, NRAD1,

S100A8, and S100A9, I generated a heatmap showing the Spearman correlations of all genes co-

expressed with each of ALDH1A3, NRAD1, S100A8, and S100A9 in patient breast tumors (Fig. 21). As

expected, this analysis revealed a striking degree of similarity with respect to the genes correlated with

each of ALDH1A3, NRAD1, S100A8, and S100A9 (Fig. 21). Supporting the previous finding that

NRAD1 is induced by ALDH1A3 and contributes to ALDH1A3-mediated gene regulation[302], and our

Spearman correlation analyses (Fig. 19A), genes correlated with ALDH1A3 showed the highest degree of

similarity with NRAD1. Further, the almost identical pattern between S100A8 and S100A9 also supports

our Spearman correlation analysis (Fig. 20A) and demonstrates the possibility for the role of the

S100A8/A9 complex in NRAD1-mediated gene regulation in breast cancer.

To further investigate the potential co-regulatory roles of ALDH1A3, NRAD1, S100A8, and

S100A9 in breast cancer, I performed the same analysis described in figure 13 with ALDH1A3, NRAD1,

S100A8, and S100A9 (Fig. 22). Of the 5852 genes from the microarray that met our cutoff criteria (p <

0.05, ≥ 1.2 or ≤ -1.2-fold expression change), we were able to match 3733 genes with RNA-seq co-

expression data from patient tumors from cBioPortal. Of these, 1080 genes were identified to be

upregulated by NRAD1 and positively correlated with ALDH1A3, NRAD1, S100A8, and S100A9, or

downregulated by NRAD1 and negatively correlated with ALDH1A3, NRAD1, S100A8, and S100A9 in

patient breast tumors (Fig. 22, Appendix 3). This data provides further evidence that ALDH1A3,

NRAD1, S100A8, and S100A9 may co-regulate at least some gene expression in breast cancer, providing

additional support for the potential roles of S100A8/A9 in NRAD1-mediated gene regulation.

Given the strong correlation between genes co-expressed with S100A8 and S100A9 and those

expressed with NRAD1, we further investigated S100A8 and S100A9 expression in TNBC and basal-like breast cancer using patient tumor RNA-seq data (Fig. 24A, B, C, and D). First, we assessed the expression of NRAD1 and ALDH1A3 in basal-like relative to non-basal-like, and in TNBC relative to non-TNBC patient tumors using the TCGA, (*Cell, 2015*) dataset (Fig. 23A, B, C, and D). The expression of both NRAD1 and ALDH1A3 was significantly higher in basal-like and TNBC patient tumors relative to non-basal and non-TNBC tumors, aligning with their CSC-associated roles[117,302] (Fig. 23A, B, C, and D). Similarly, the expression of both S100A8 and S100A9 was significantly higher in Basal-like patient breast tumors relative to non-basal tumors (Fig. 24B and D). S100A8 and S100A9 expression was elevated in TNBC tumors relative to non-TNBC tumors, although the difference in expression was not significant (Fig. 24A and C). Thus, given that S100A8 and S100A9 are more highly expressed in basal-like and TNBC patient tumors, mirroring the expression of ALDH1A3 and NRAD1, their potential roles in NRAD1-mediated gene regulation are further supported.

As NRAD1, S100A8 and S100A9 expression is significantly higher in basal-like patient tumors relative to non-basal patient tumors (Fig. 23B, Fig. 24B and D), we performed the co-expression Spearman correlation analysis to assess the correlation of Spearman correlations of genes co-expressed with NRAD1 vs. those of genes co-expressed with S100A8 or S100A9 in basal-like patient tumors (Fig. 25A and B). This analysis revealed positive correlations between NRAD1 vs. S100A8 (R = 0.36, p < 2.2e-16) (Fig. 25A) and NRAD1 vs. S100A9 (R = 0.37, p < 2.2e-16) (Fig. 25B), demonstrating that the positive correlations observed previously in patient tumors of all subtypes (Fig. 11A, Fig. 20B) were not due entirely to non-basal subtypes. This data also supports our findings showing the regulatory association of S100A8 with NRAD1 in MDA-MB-468 cells and supports the likely role of S100A8/A9 in NRAD1-mediated gene regulation in TNBC/Basal-like breast cancers.

Spearman correlations of genes co-expressed with ALDH1A3 vs. Spearman correlations of genes co-expressed with NRAD1 or S100A8 in patient breast tumors (TCGA Cell, 2015)

A

NRAD1 vs. ALDH1A3
R= 0.81, p < 2.2e-16

B

S100A8 vs. ALDH1A3
R= 0.61, p < 2.2e-16
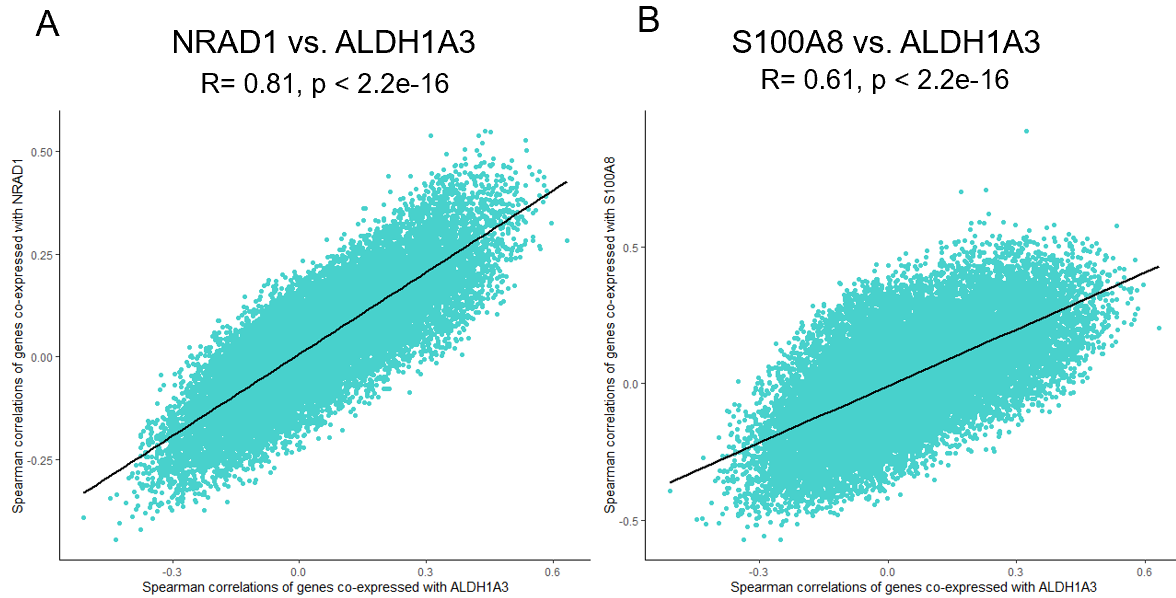
Figure 19. Spearman correlations of genes co-expressed with ALDH1A3 are strongly correlated with the spearman correlations of genes co-expressed with NRAD1 or S100A8 in patient breast tumors. Using RNA-sequencing data from the TCGA (Cell, 2015) dataset (breast invasive carcinoma; n=816) accessible via cBioPortal, Spearman correlations between all protein-coding genes (~20,200) co-expressed with ALDH1A3 were plotted against the Spearman correlations of genes co-expressed with NRAD1 **(A)** or S100A8 **(B).** The correlation coefficients and p-values were calculated using R.

Spearman correlations of genes co-expressed with S100A8, NRAD1, or ALDH1A3 vs. Spearman correlations of genes co-expressed with S100A9 in patient breast tumors (TCGA Cell, 2015)
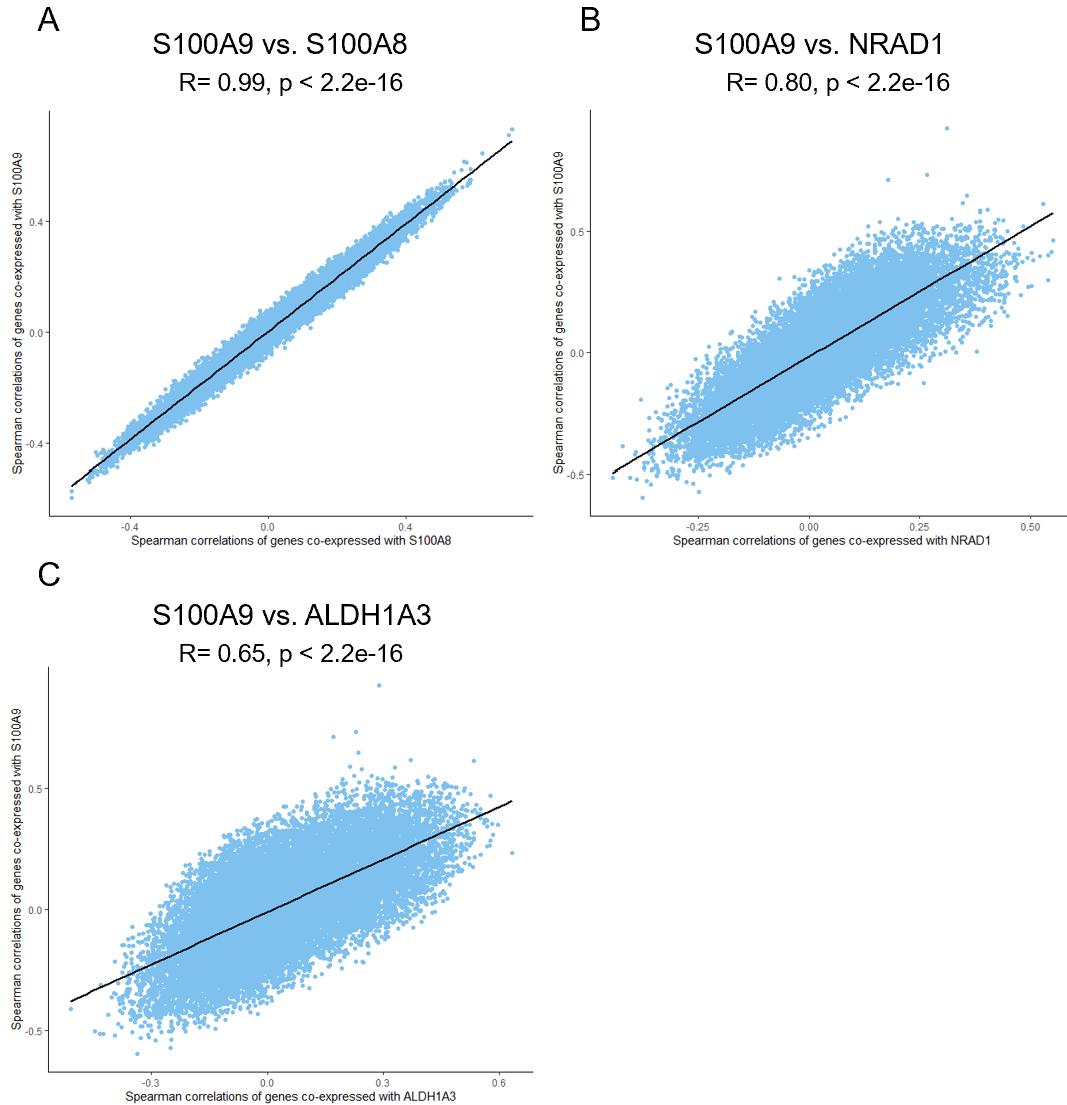
Figure 20. Spearman correlations of genes co-expressed with S100A9 are strongly correlated with the spearman correlations of genes co-expressed with S100A8, NRAD1, or ALDH1A3 in patient breast tumors. Using patient breast tumor RNA-sequencing data from the TCGA (Cell, 2015) dataset (breast invasive carcinoma; n=816) accessible via cBioPortal, Spearman correlations between all protein-coding genes (~20,200) co-expressed with S100A9 were plotted against the Spearman correlations of all protein-coding genes co-expressed with S100A8 **(A)**, NRAD1 **(B)**, or S100A9 **(C)**. The correlation coefficients and p-values were calculated using R.

Spearman correlations of genes co-expressed with ALDH1A3, NRAD1, S100A8, and S100A9 in patient breast tumors (TCGA, Cell 2015)
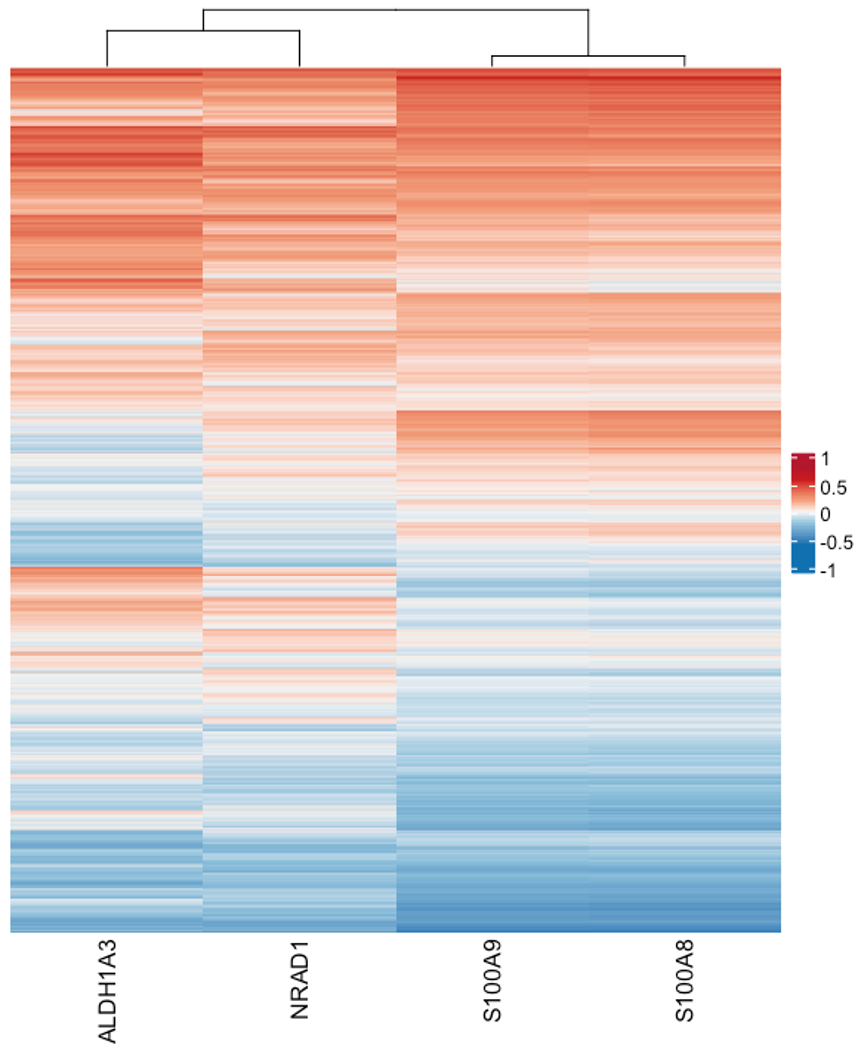
Figure 21. A high degree of similarity exists among genes correlated with ALDH1A3, NRAD1, S100A8, and S100A9 in patient breast tumors. Using patient breast tumor RNA-sequencing data from the TCGA (Cell, 2015) dataset (breast invasive carcinoma; n=816) accessible via cBioPortal, Spearman correlations of all genes (~20,200) co-expressed with each of ALDH1A3, NRAD1, S100A8, and S100A9 were extracted and used to compare correlation profiles (shown in the heatmap).

Spearman correlations of NRAD1-regulated genes co-expressed with ALDH1A3, NRAD1, S100A8, and S100A9 in patient breast tumors (TCGA, Cell 2015)
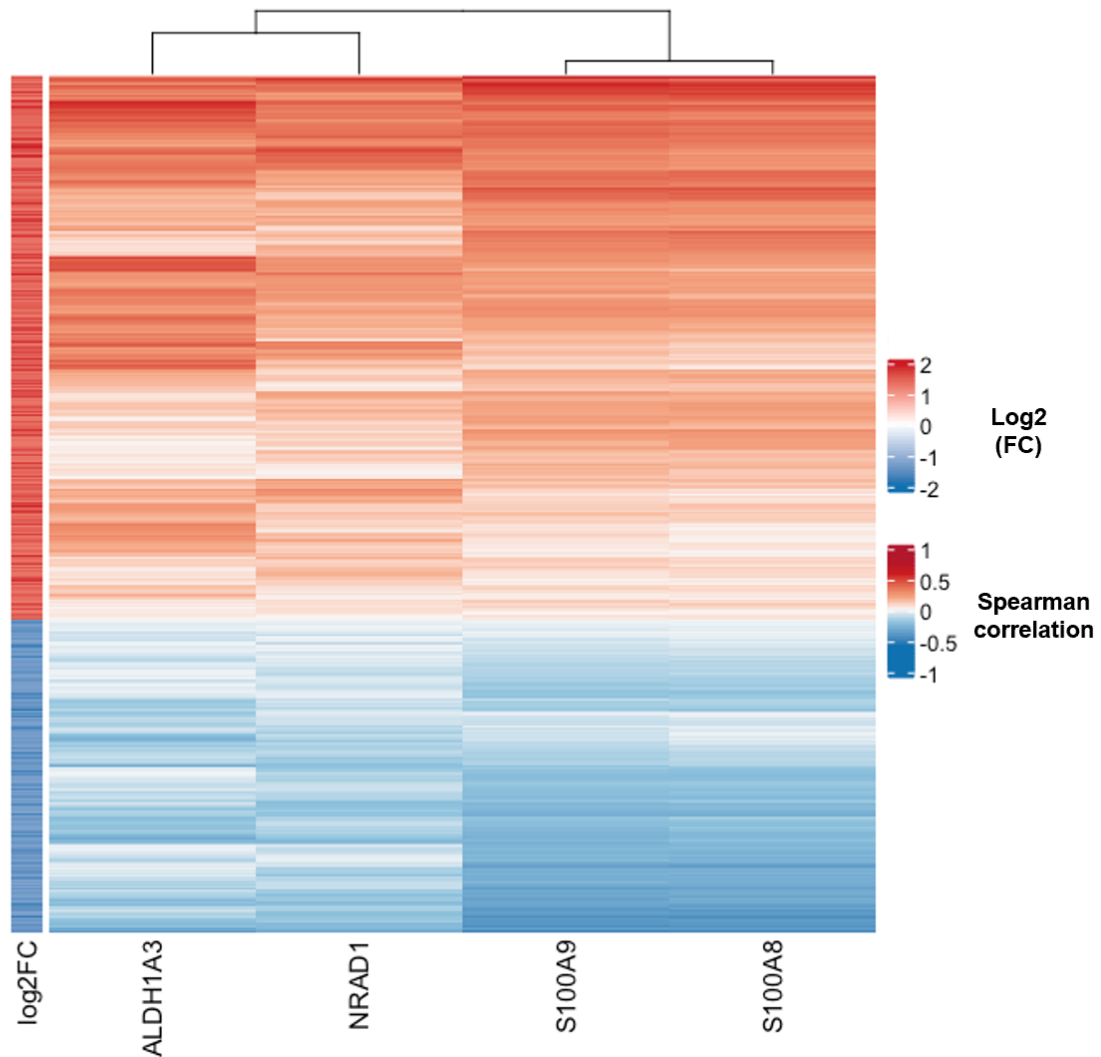
Figure 22. Numerous NRAD1-regulated genes are correlated with ALDH1A3, NRAD1, S100A8, and S100A9 in patient breast tumors. NRAD1-regulated genes identified using the Affymetrix Human Gene 2.0 ST microarray platform (control vs. GapmeR #4; n=3) in MDA-MB-468 cells ($\geq$ 1.2 or $\leq$ -1.2-fold expression change, $p < 0.05$) were assessed for their Spearman co-expression correlations with NRAD1, ALDH1A3, S100A8, and S100A9 in patient breast tumors. To identify co-expression Spearman correlations, RNA-sequencing data (RNA-seq RSEM V2) of patient breast tumors (invasive breast carcinoma) was extracted from the TCGA, Cell 2015 dataset (n = 816), accessible via cBioPortal. The Spearman correlations of NRAD1-upregulated genes positively correlated, and NRAD1-downregulated genes negatively correlated with NRAD1, ALDH1A3, S100A8, and S100A9 in patient breast tumors are shown in the heatmap. The Log_2 fold-change in expression from the microarray is shown on the left.

NRAD1 and ALDH1A3 expression in patient breast tumors (TCGA, Cell 2015)
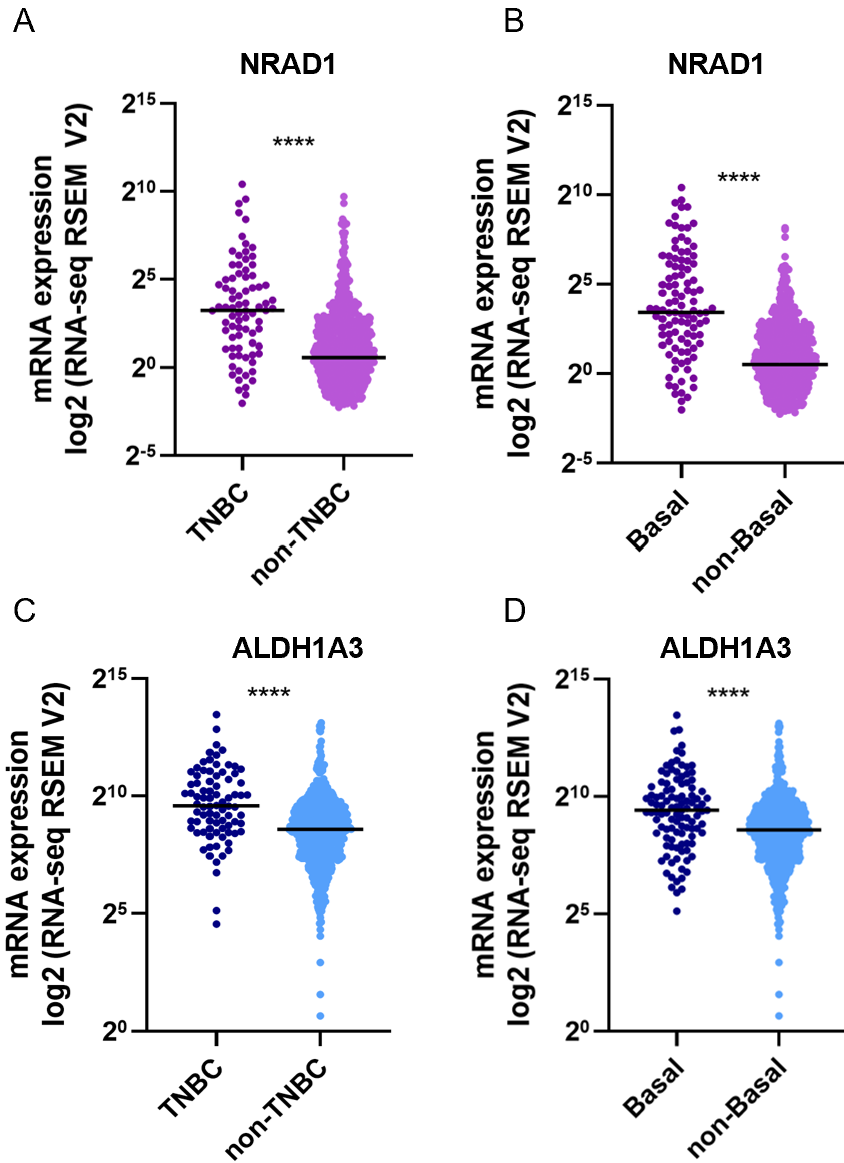
Figure 23. NRAD1 and ALDH1A3 expression is significantly higher in TNBC and Basal-like patient tumors. mRNA expression data (RNA-seq RSEM V2) from patient breast tumors (breast invasive carcinoma, n = 107 basal-like tumors and 97 TNBC tumors) was extracted from the TCGA, Cell 2015 dataset via cBioPortal. mRNA expression of NRAD1 or ALDH1A3 in TNBC or Basal-like tumors vs. non-TNBC or non-Basal tumors is plotted as $\log_2$ (RNA-seq RSEM V2). Significance comparing NRAD1 and ALDH1A3 expression in TNBC vs. non-TNBC patient tumors (**A** and **C**, respectively) and in Basal vs. non-Basal patient tumors (**B** and **D**, respectively) was determined using an unpaired two-tailed t-test (**** = $p<0.0001$). Bars represent mean mRNA expression.
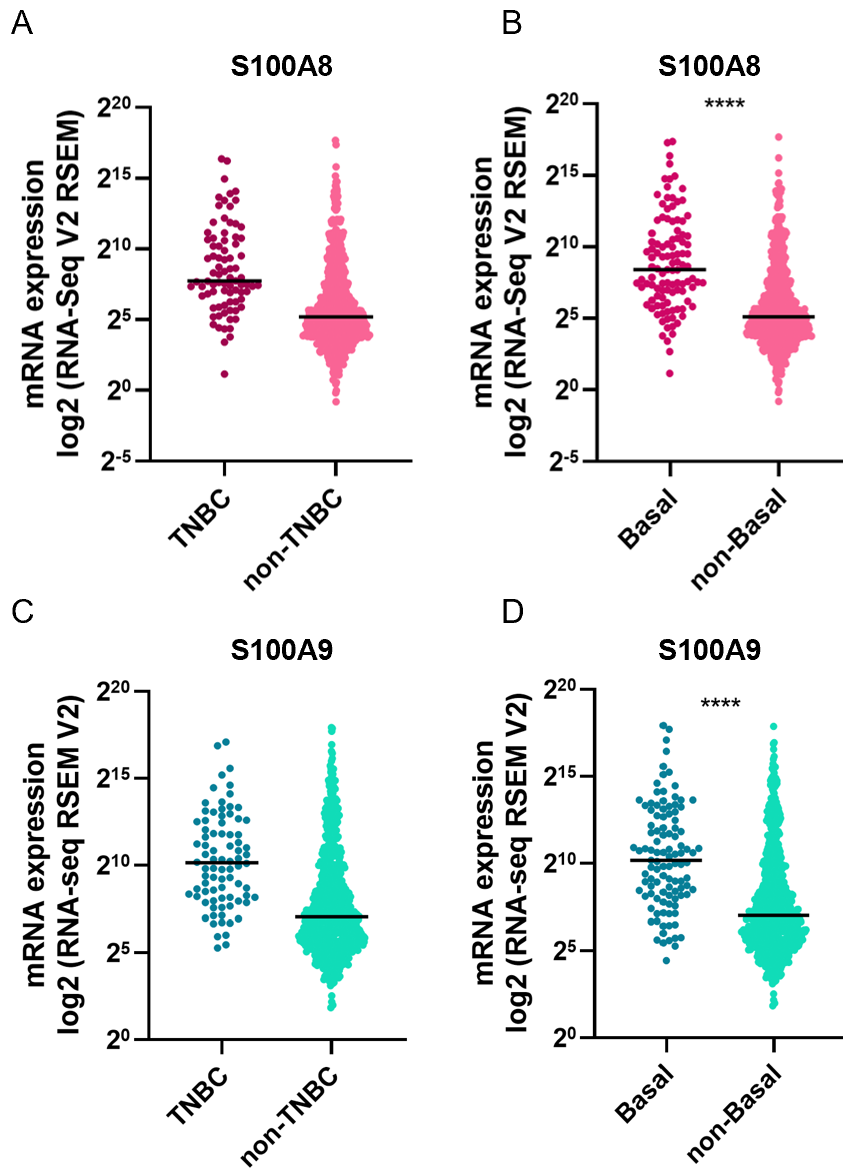
Figure 24. S100A8 and S100A9 expression is significantly higher in Basal-like patient tumors and is elevated in TNBC tumors. mRNA expression data (RNA-seq RSEM V2) from patient breast tumors (breast invasive carcinoma, n = 107 basal-like tumors and 97 TNBC tumors) was extracted from the TCGA, Cell 2015 dataset via cBioPortal. mRNA expression of S100A8 or S100A9 in TNBC or Basal-like tumors vs. non-TNBC or non-Basal tumors is plotted as $\log_2$ (RNA-seq RSEM V2). Significance comparing S100A8 and S100A9 expression in TNBC vs. non-TNBC patient tumors (**A** and **C**, respectively) and in Basal vs. non-Basal patient tumors (**B** and **D**, respectively) was determined using an unpaired two-tailed t-test (**** = p<0.0001). Bars represent mean mRNA expression.

Spearman correlations of genes co-expressed with NRAD1 vs. Spearman correlations of genes co-expressed with S100A8 or S100A9 in Basal-like patient breast tumors (TCGA Cell, 2015)

NRAD1 vs. S100A8

R= 0. 36, p < 2.2e-16

A

NRAD1 vs. S100A9

R= 0. 37, p < 2.2e-16

B
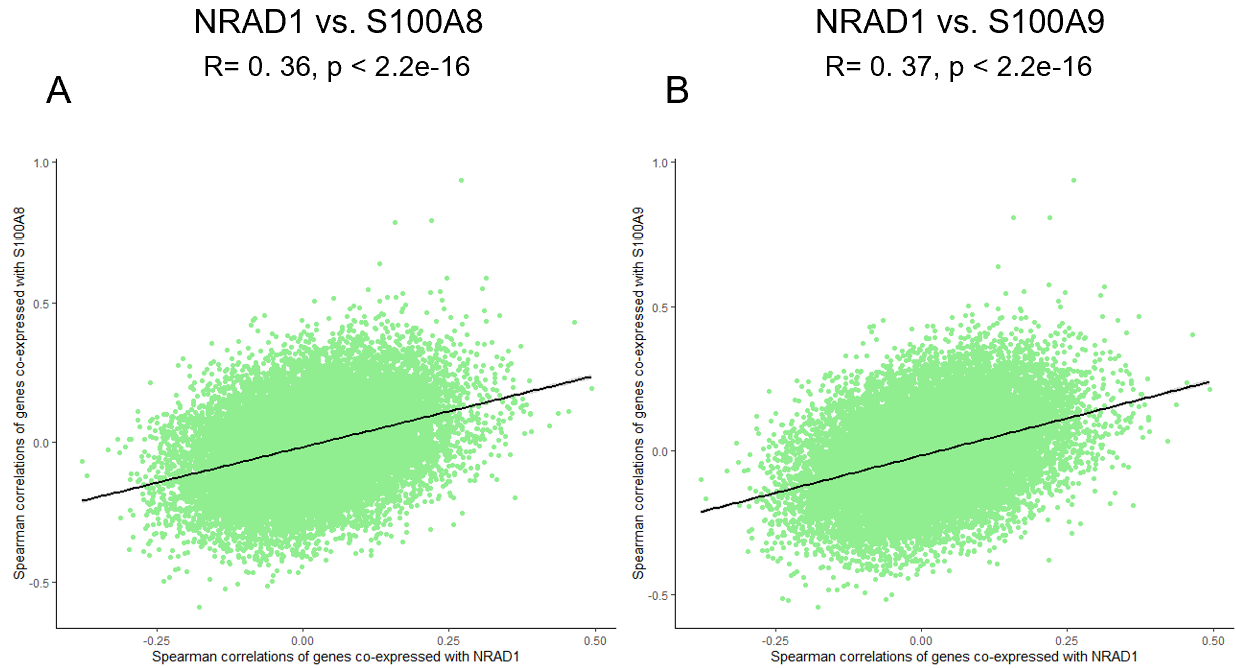
Figure 25. Spearman correlations of genes co-expressed with NRAD1 are strongly correlated with the spearman correlations of genes co-expressed with S100A8 or S100A9 in Basal-like patient breast tumors. Using RNA-sequencing data from the TCGA (Cell, 2015) dataset (breast invasive carcinoma; Basal subtype: n=107) accessible via cBioPortal, Spearman correlations between all protein-coding genes (~20,200) co-expressed with NRAD1 were plotted against the Spearman correlations of genes co-expressed with S100A8 **(A)** or S100A9 **(B)**. The correlation coefficients and p-values were calculated using R.

3.7 Investigating the potential roles of S100A8/A9 and NRAD1 in ovarian cancer

In addition to its oncogenic role in breast cancer, NRAD1 has also been implicated in ovarian cancer progression[339,340]. NRAD1 has been shown to be upregulated in ovarian carcinoma tissues relative to normal tissues and to promote angiogenesis during ovarian cancer development via recruitment of NF-κB1 and downregulation of MEST (Mesoderm Specific Transcript Homolog)[339]. NRAD1-mediated inhibition of MEST was proposed to occur via NRAD1-mediated localization of NF-κB1 to the MEST promoter, initiating ovarian cancer cell invasion, migration, and angiogenesis via MMP-2, MMP-9, VEGF, and CD31 upregulation. Importantly, S100A8 is also highly expressed in ovarian cancer tissues relative to non-cancerous tissues[362], supporting the potential interaction of NRAD1 and S100A8/A9 in ovarian cancer. To show this, I analyzed microarray data from normal and ovarian tumor tissues accessible from NCBI's gene expression omnibus (GEO) datasets platform (GSE38666, Fig. 26A and B). Here, Lili and colleagues utilized laser capture microdissection to isolate cancerous cells from the tumors of 18 serous ovarian cancer patients[352]. Surface ovarian epithelial cells isolated from the normal ovaries of an additional 12 individuals were used as controls. These cells were used to perform microarray analysis (Affymetrix Human Genome U133 Plus 2.0 Array) to quantify changes in gene expression between normal and ovarian tumor tissues. Our analysis of this data revealed that NRAD1 expression was significantly higher in ovarian tumor tissue relative to normal tissue, supporting the results of previous studies (Fig. 26A and B).

To further explore the potential interaction between NRAD1 and S100A8/A9 in ovarian cancer, I performed the Spearman correlation analysis as previous, plotting the Spearman correlations of the genes co-expressed with S100A8 or S100A9 against those of the genes co-expressed with NRAD1 in ovarian cancer patient tumors (ovarian serous cystadenocarcinoma) using RNA-seq data extracted from the TCGA PanCancer Atlas accessible via cBioPortal (n=585) (Fig. 27A and B). Strikingly, this analysis revealed strong and significant positive correlations between the Spearman correlations of genes co-expressed with NRAD1 and those of the genes co-expressed with S100A8 ($R = 0.58$, $p < 2.2e-16$) (Fig.

27A) and S100A9 (R = 0.63, p < 2.2e-16) (Fig. 27B), providing support for the potential role of

S100A8/A9 in NRAD1-mediated gene regulation in ovarian cancer.

NRAD1 expression in normal vs. ovarian tumor tissue (GSE38666)
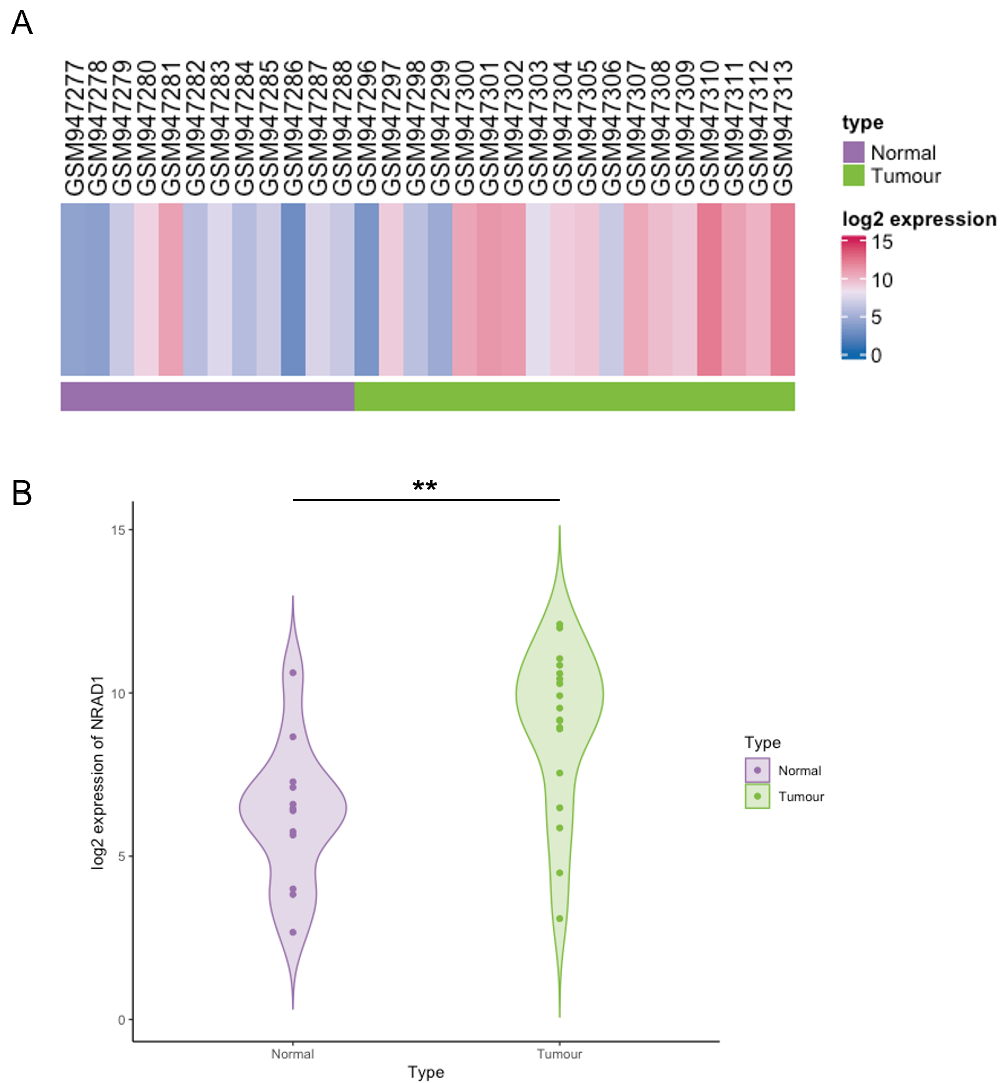
Figure 26. NRAD1 expression is significantly elevated in ovarian tumor tissue relative to normal ovarian tissue. Data from gene expression profiling via microarray (Affymetrix Human Genome U133 Plus 2.0 Array) to assess genome-wide changes in gene expression between normal and ovarian tumor tissues (serous ovarian cancer) was extracted from the NCBI Gene Expression Omnibus (GEO) (GSE38666) and used to assess NRAD1 expression in normal and ovarian tumor samples. (A) Heatmap showing log2 expression of NRAD1 in surface ovarian epithelial cells isolated from the normal ovaries of 12 individuals (purple) and in cancerous cells from the tumors of 18 serous ovarian cancer patients (green). Patient IDs (above) are included with each sample. (B) Log2 NRAD1 expression in normal surface ovarian epithelial cells (purple) vs. serous ovarian tumor tissue (green). Significance was determined using an unpaired two-tailed t-test (** = $p < 0.01$).

Spearman correlations of genes co-expressed with NRAD1 vs. Spearman correlations of genes co-expressed with S100A8 or S100A9 in ovarian cancer patient tumors (TCGA PanCancer Atlas)

A S100A8 vs. NRAD1
R= 0.58, p < 2.2e-16

B S100A9 vs. NRAD1
R= 0.63, p < 2.2e-16
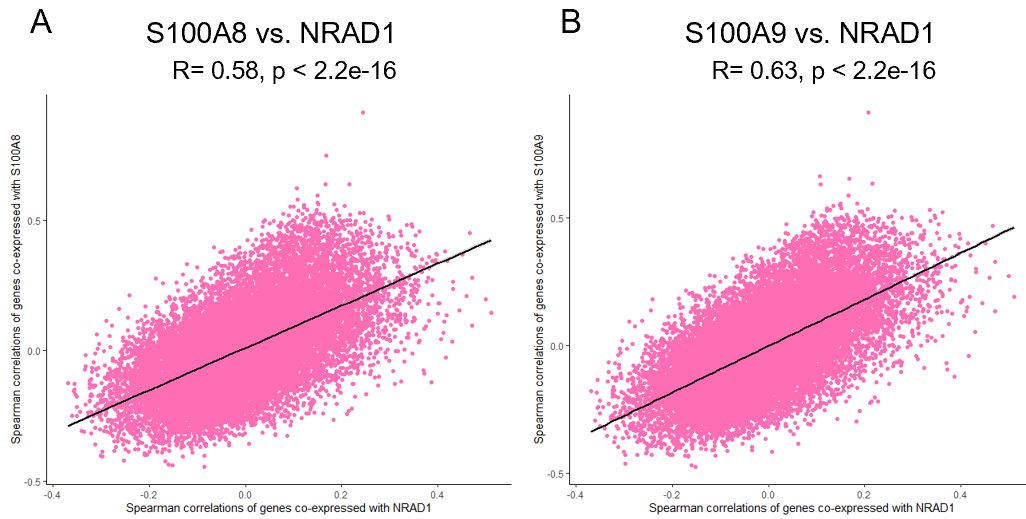
Figure 27. Spearman correlations of genes co-expressed with S100A8 or S100A9 are strongly correlated with the spearman correlations of genes co-expressed with NRAD1 in ovarian cancer patient tumors. Using RNA-sequencing data from the TCGA (PanCancer Atlas) dataset (serous ovarian cystadenocarcinoma; n=585) accessible via cBioPortal, Spearman correlations between all protein-coding genes (~20,200) co-expressed with NRAD1 were plotted against the Spearman correlations of all protein-coding genes co-expressed with S100A8 **(A)** or S100A9 **(B)**. The correlation coefficients and p-values were calculated using R.

CHAPTER 4 – RESULTS DATA CHAPTER 2

4.1 NRAD1 regulates miRNA expression in MDA-MB-468 and SUM149 cells.

Given that most lncRNAs with miRNA-sequestering activity exhibit cytoplasmic residency and our detection of both NRAD1 transcripts in the MDA-MB-468 cell cytoplasm (Fig. 4B, C, and D), NRAD1 may interact with cytoplasmic miRNAs to regulate gene expression in breast cancer. Intriguingly, in support of this, there are now reports of NRAD1 acting as a miRNA sponge, contributing to the progression of other human cancers[340,343,347]. Furthermore, while NRAD1 was previously shown to regulate gene expression via chromatin binding, almost 40% of NRAD1-regulated genes are not bound by NRAD1 at genic regions[302]. This finding suggests that NRAD1 regulates the expression of genes not bound by NRAD1 through mechanisms other than chromatin binding.

In this data chapter, we explore the putative interactions between NRAD1 and miRNAs to identify potential novel mechanisms through which NRAD1 may regulate gene expression in the absence of chromatin binding. I hypothesize that NRAD1 modulates miRNA activity to regulate gene expression in breast cancer. To explore this, I treated MDA-MB-468 and SUM149 cells with two separate anti-NRAD1 LNA GapmeRs (GapmeR #3 or GapmeR# 4) or the negative control GapmeR and extracted total RNA enriched for miRNAs using the mirVana miRNA isolation kit (Thermo Fisher). miRNA-enriched RNA samples were subjected to microarray analyses (Affymetrix GeneChip miRNA 4.0 Array) to elucidate changes in the cellular miRNA landscape in response to NRAD1 knockdown.

Following NRAD1 knockdown in MDA-MB-468 and SUM149 cells and miRNA microarray analyses, we selected a fold-miRNA expression change cutoff of ≥ 1.3 or ≤ -1.3 and a p-value cutoff of $p < 0.05$. In both cell lines, a series of miRNAs meeting our cutoff criteria were found to be upregulated or downregulated by NRAD1 inhibition (Fig. 28A, B, C, and D). Accordingly, treatment of MDA-MB-468 cells with anti-NRAD1 GapmeR #3 resulted in the upregulation of 49 miRNAs and downregulation of 63 miRNAs Fig. 28C, Table 6). Similarly, treatment with GapmeR #4 induced upregulation of 64 and downregulation of 66 miRNAs (Fig. 28D, Table 6). Treatment of SUM149 cells with GapmeR #3 resulted

in upregulation of 130 and downregulation of 111 miRNAs (Fig. 28A, Table 6). SUM149 cell treatment with GapmeR #4 led to upregulation and downregulation of 127 miRNAs each (Fig. 28B, Table 6). To demonstrate that treatment of either cell line with two different GapmeRs yields similar miRNA expression changes, we plotted the fold-expression change of all miRNAs induced by NRAD1 knockdown with GapmeR #3 and GapmeR #4 in MDA-MB-468 or SUM149 cells using a heatmap (Fig. 29). This analysis revealed that NRAD1 knockdown with GapmeRs #3 and #4 results in similar miRNA expression changes within and across both cell lines (Fig. 29). Together, this data demonstrates that NRAD1 regulates the expression of miRNAs in two different TNBC cell lines and provides preliminary support for the hypothesis that NRAD1 regulates gene expression via miRNA interactions.

Importantly, we observed miRNAs upregulated (Fig. 30, Table 7 and 8) or downregulated (Fig. 31, Table 9 and 10) by both GapmeR #3 and GapmeR #4 in a cell line specific manner. MiRNAs upregulated upon NRAD1 knockdown point to the role of NRAD1 as a miRNA sponge in the absence of NRAD1 inhibition. Given that lncRNA sponging of miRNAs is a well characterized regulatory mechanism and that NRAD1 has been previously shown to sequester miRNAs in cancers, we focused our downstream analyses on miRNAs upregulated upon NRAD1 knockdown. Thus, to narrow the set of miRNAs that may interact with NRAD1 to regulate gene expression via NRAD1-mediated miRNA sponging, we identified miRNAs meeting our cutoff criteria that were upregulated upon NRAD1-inhibition with both GapmeRs in both cell lines. To this end, we identified two miRNAs, hsa-miR-4485-3p and hsa-miR-595 that were upregulated following NRAD1 knockdown by GapmeRs 3 and 4 in both MDA-MB-468 and SUM149 cells and were therefore further investigated for their potential roles in NRAD1 mediated gene regulation (Fig. 30). Of note, we also identified two miRNAs, hsa-miR-1226-3p and hsa-miR-4458 that were downregulated following NRAD1 knockdown with both GapmeRs in both cell lines. These miRNAs represent attractive targets for the study of their roles in NRAD1-mediated gene regulation through mechanisms other than miRNA sponging (Fig. 31).

NRAD1-induced miRNA expression changes in MDA-MB-468 and SUM149 cells
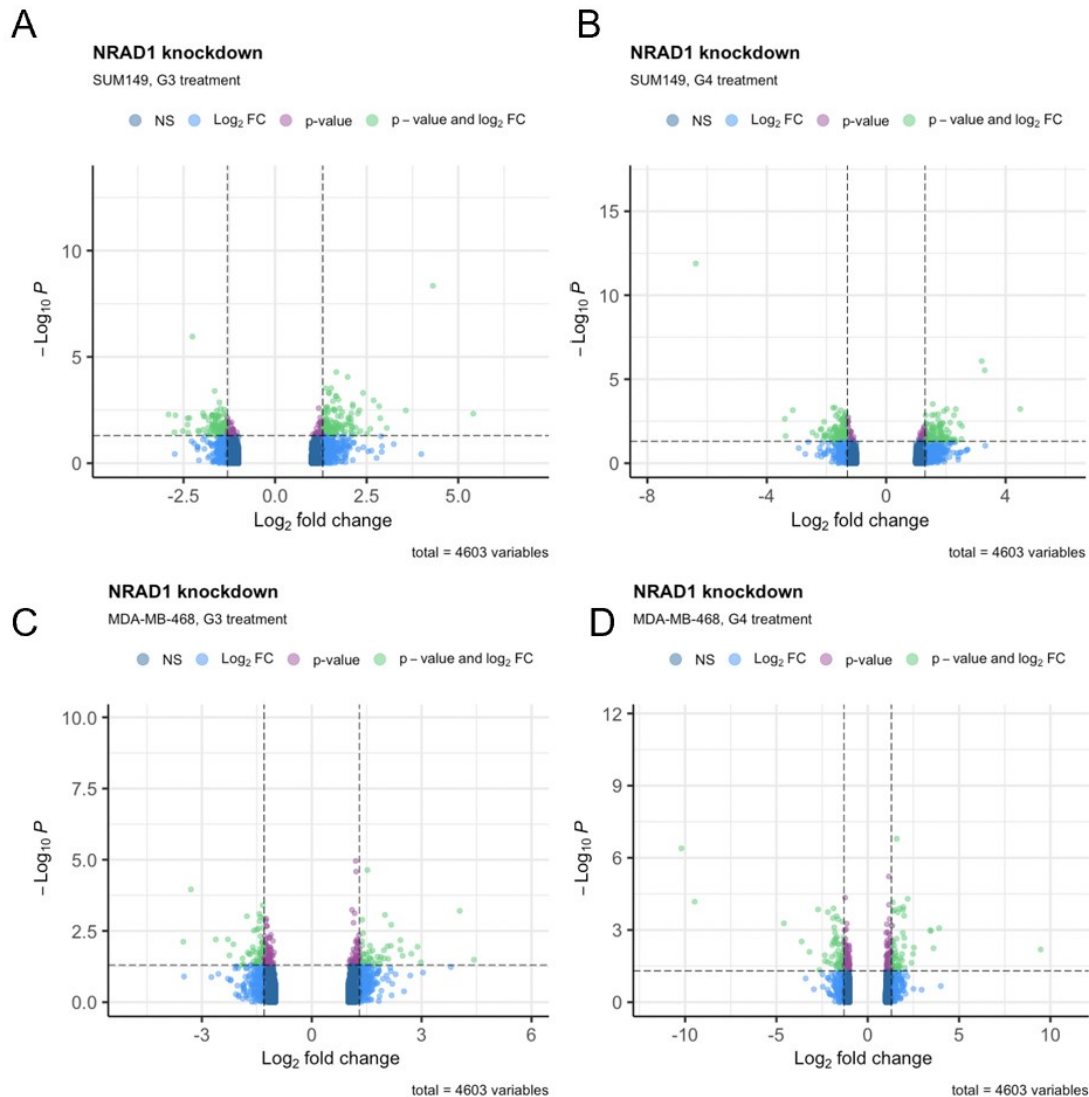
Figure 28. NRAD1 induces miRNA expression changes in MDA-MB-468 and SUM149 cells. Genome-wide miRNA expression changes induced by GapmeR-mediated NRAD1 inhibition (control vs. GapmeR #3 or GapmeR #4-treated) was quantified in MDA-MB-468 and SUM149 cells using the Affymetrix GeneChip miRNA 4.0 Array (n=3). The Log_2 fold-change in miRNA expression is plotted versus the –log10(ANOVA p-value) of probes corresponding to 2578 mature human miRNAs and 2025 human pre-miRNAs (4603 total). Dots coloured light green show miRNAs with a fold-expression change of ≥ 1.3 or ≤ -1.3 and a p-value < 0.05. NRAD1 knockdown with GapmeR #3 or GapmeR #4 in SUM149 cells is shown in panels **A** and **B** respectively. NRAD1 knockdown with GapmeR #3 or GapmeR #4 in MDA-MB-468 cells is shown in panels **C** and **D** respectively.

Table 6. Changes in gene expression in response to NRAD1 inhibition in MDA-MB-468 and SUM149 cells. The number of miRNAs upregulated (FC ≥ 1.3, p < 0.05) or downregulated (FC ≤ -1.3, p < 0.05) by NRAD1 knockdown with GapmeR #3 or GapmeR #4 in MDA-MB-468 or SUM149 cells are shown. Gene expression changes were quantified using the Affymetrix GeneChip miRNA 4.0 Array (n=3).

Cell line	GapmeR	Upregulated miRNAs	Downregulated miRNAs
MDA-MB-468	GapmeR #3	49	63
	GapmeR #4	64	66
SUM149	GapmeR #3	130	111
	GapmeR #4	127	127
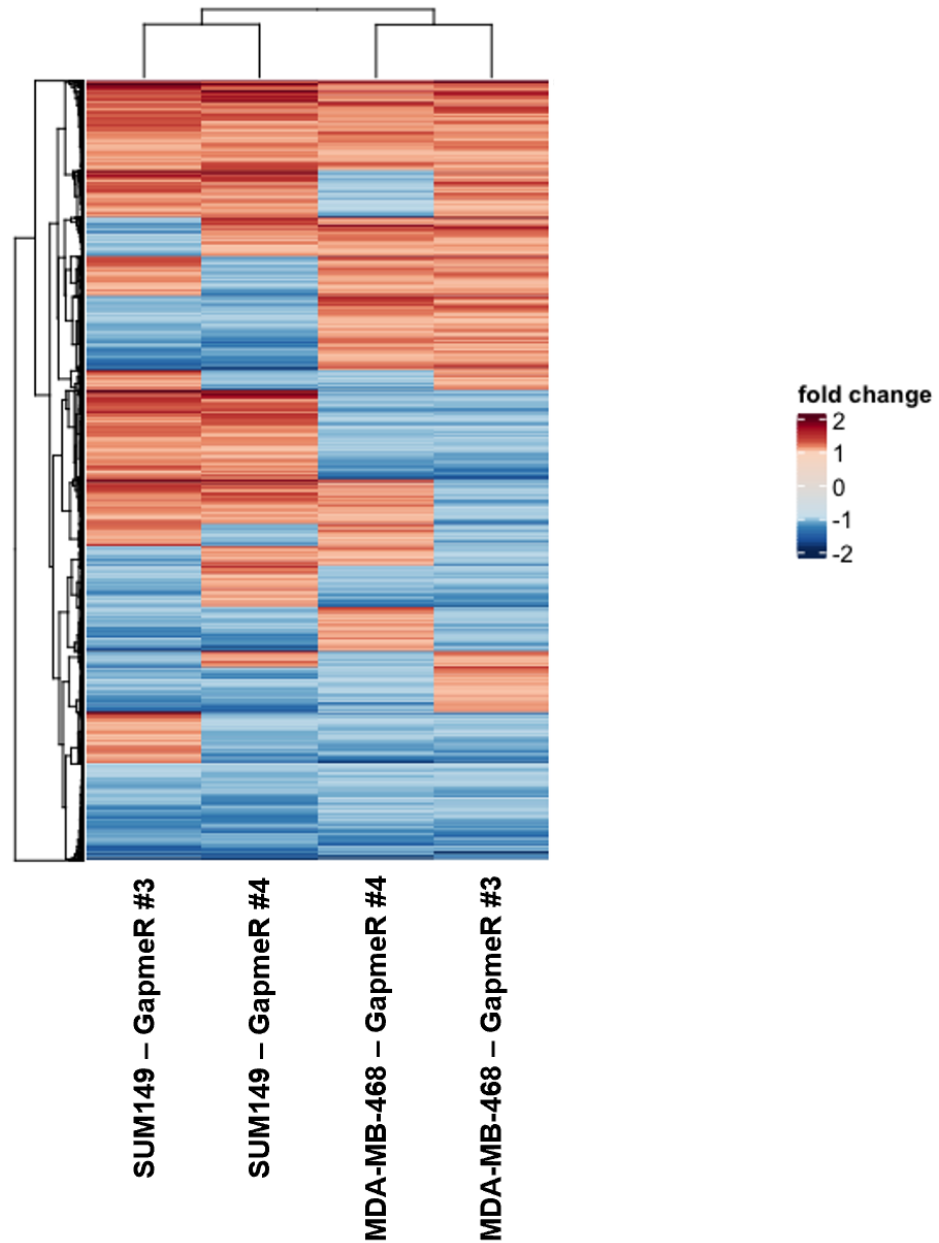
Figure 29. NRAD1 inhibition with GapmeR #3 or GapmeR #4 induces similar gene expression changes. Fold-change in expression (Control vs. GapmeR-treated) of 4603 miRNAs quantified using the Affymetrix GeneChip miRNA 4.0 array following NRAD1 knockdown with GapmeR #3 or GapmeR #4 in MDA-MB-468 or SUM149 cells (n=3).

miRNAs upregulated upon GapmeR-mediated NRAD1 knockdown in MDA-MB-468 and SUM149 cells
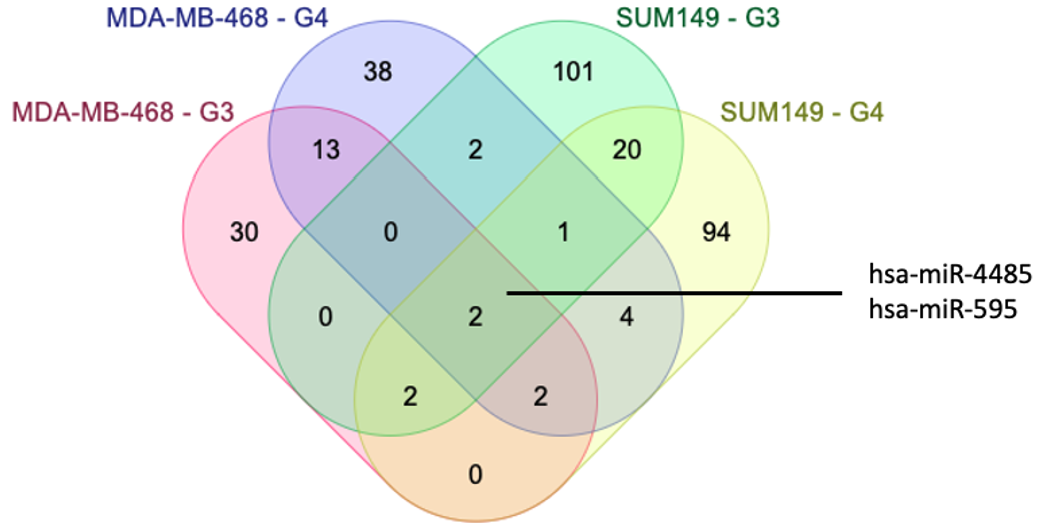
	MDA-MB-468 cells				SUM149 cells			
	GapmeR #3		GapmeR #4		GapmeR #3		GapmeR #4	
miRNA	FC	p-value	FC	p-value	FC	p-value	FC	p-value
miR-595	1.42	0.019966	1.88	0.010027	2.69	0.0011	1.82	0.0125
miR-4485-3p	4.04	0.000621	1.95	0.000111	2.12	0.0017	3.3	2.95E-06

Figure 30. miR-4485-3p and miR-595 are upregulated following NRAD1 knockdown with GapmeRs 3 and 4 in MDA-MB-468 and SUM149 cells. Changes in miRNA expression were quantified using the Affymetrix GeneChip miRNA 4.0 Array following NRAD1 knockdown with GapmeR #3 or GapmeR #4 in MDA-MB-468 and SUM149 cells (n=3). The miRNAs upregulated (FC ≥ 1.3, p < 0.05) following NRAD1 knockdown with GapmeR #3 or GapmeR #4 in MDA-MB-468 or SUM149 cells are shown. Commonly upregulated miRNAs are shown by overlapping regions in the Venn diagram. miR-4485-3p and miR-595 are annotated given their upregulation in GapmeR #3 and #4-treated MDA-MB-468 and SUM149 cells. Fold-expression change (Control vs. GapmeR-treated) and p-values are shown for miR-4485-3p and miR-595 in GapmeR #3 or GapmeR #4-treated MDA-MB-468 and SUM149 cells.

Table 7. Fold-expression change values and p-values of miRNAs upregulated (FC ≥ 1.3, p < 0.05) following NRAD1 inhibition with GapmeR #3 and GapmeR #4 in MDA-MB-468 cells.

miRNA (hsa)	GapmeR #3 (FC)	GapmeR #3 (p-value)	GapmeR #4 (FC)	GapmeR #4 (p-value)
miR-29b-1-5p	1.79	0.025435	3.6	0.00577
miR-139-5p	2.98	0.039981	2.4	0.023137
miR-212-3p	2.89	0.011398	2.53	0.005355
miR-194-3p	2.17	0.00192	9.47	0.006474
miR-584-3p	1.45	0.040424	1.77	0.029331
miR-595	1.42	0.019966	1.88	0.010027
miR-1247-3p	2.48	0.023543	1.88	0.018212
miR-1973	1.73	0.048452	1.46	0.048598
miR-3622b-5p	2.72	0.019928	1.54	0.01151
miR-4485	4.04	0.000621	1.95	0.000111
miR-2392	2.07	0.010926	2	0.004711
miR-4649-5p	2.42	0.006695	1.58	0.002306
miR-6779-5p	1.4	0.003796	2.19	5.10E-05
miR-6795-5p	2.16	0.010547	2.48	0.009829
miR-6851-5p	4.43	0.032424	1.56	0.0382
miR-6870-5p	2	0.000868	1.92	0.001322
mir-4485	1.89	0.025513	1.31	0.015867

Table 8. Fold-expression change values and p-values of miRNAs upregulated (FC ≥ 1.3, p < 0.05) following NRAD1 inhibition with GapmeR #3 and GapmeR #4 in SUM149 cells.

miRNA (hsa)	GapmeR #3 (FC)	GapmeR #3 (p-value)	GapmeR #4 (FC)	GapmeR #4 (p-value)
miR-96-3p	2.12	0.0021	1.61	0.0087
miR-192-5p	1.36	0.017	1.53	0.0014
miR-500a-3p	1.67	5.18E-05	1.56	3.00E-04
miR-595	2.69	0.0011	1.82	0.0125
miR-4461	1.58	0.0012	1.62	0.0013
miR-3135b	1.98	8.80E-05	3.2	8.34E-07
miR-4485	2.12	0.0017	3.3	2.95E-06
miR-4533	1.34	0.0129	1.51	0.0283
miR-4734	1.4	0.0131	1.3	0.0267
miR-4793-3p	1.58	0.0371	1.77	0.0406
miR-5690	1.53	3.00E-04	1.32	0.0473
miR-6075	2.2	0.0212	1.81	0.0394
miR-6723-5p	1.67	0.0327	1.81	0.0119
miR-6782-5p	1.7	0.0209	1.82	0.0071
miR-6807-5p	1.36	0.0487	2.33	6.00E-04
miR-6825-3p	1.43	0.0149	1.4	0.0457
miR-6864-3p	1.61	0.0029	1.54	0.0234
miR-7850-5p	1.83	0.0039	1.53	0.049
mir-523	1.39	0.0173	1.39	0.0326
mir-596	1.45	0.038	1.53	0.0081
mir-663a	1.73	0.001	1.55	0.0048
mir-3648	1.95	0.0159	1.38	0.0226
mir-4634	2.01	0.007	2.27	0.0014
mir-5699	1.46	0.0361	1.81	0.005
mir-6804	1.4	0.0333	1.78	0.0273

miRNAs downregulated upon GapmeR-mediated NRAD1 knockdown in MDA-MB-468 and SUM149 cells
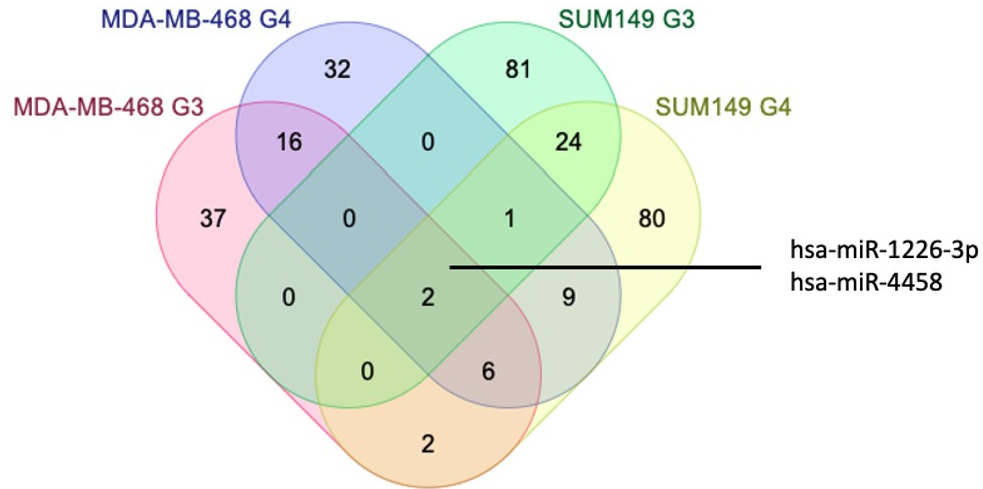
	MDA-MB-468 cells				SUM149 cells			
	GapmeR #3		GapmeR #4		GapmeR #3		GapmeR #4	
miRNA	FC	p-value	FC	p-value	FC	p-value	FC	p-value
miR-1226-3p	-1.41	0.013069	-1.38	0.012007	-1.52	0.0243	-1.88	0.0097
miR-4458	-3.51	0.007582	-9.48	6.70E-05	-2.39	0.0077	-3.13	7.00E-04

Figure 31. miR-1226-3p and miR-4458 are downregulated following NRAD1 knockdown with GapmeRs 3 and 4 in MDA-MB-468 and SUM149 cells. Changes in miRNA expression were quantified using the Affymetrix GeneChip miRNA 4.0 Array following NRAD1 knockdown with GapmeR #3 or GapmeR #4 in MDA-MB-468 and SUM149 cells (n=3). The miRNAs downregulated (FC ≤ -1.3, p < 0.05) following NRAD1 knockdown with GapmeR #3 or GapmeR #4 in MDA-MB-468 or SUM149 cells are shown. Commonly upregulated miRNAs are shown by overlapping regions in the Venn diagram. miR-1226-3p and miR-4458 are annotated given their upregulation in GapmeR #3 and #4-treated MDA-MB-468 and SUM149 cells. Fold-expression change (Control vs. GapmeR-treated) and p-values are shown for miR-1226-3p and miR-4458 in GapmeR #3 or GapmeR #4-treated MDA-MB-468 and SUM149 cells.

Table 9. Fold-expression change values and p-values of miRNAs downregulated (FC ≤ -1.3, p < 0.05) following NRAD1 inhibition with GapmeR #3 and GapmeR #4 in MDA-MB-468 cells.

miRNA (hsa)	GapmeR #3 (FC)	GapmeR #3 (p-value)	GapmeR #4 (FC)	GapmeR #4 (p-value)
miR-28-5p	-1.77	0.000965	-1.87	0.000127
miR-28-3p	-1.89	0.02921	-1.7	0.025946
miR-99a-5p	-1.58	0.003048	-1.33	0.002045
miR-199a-5p	-1.54	0.017845	-1.32	0.023914
miR-181a-2-3p	-1.61	0.011719	-1.53	0.011154
miR-151a-5p	-1.4	0.004854	-1.71	0.000268
miR-151a-3p	-1.46	0.002601	-1.9	0.000322
miR-345-5p	-1.41	0.020877	-1.31	0.02564
miR-519d-5p	-1.41	0.033234	-1.5	0.018136
miR-615-3p	-2.14	0.009303	-1.31	0.005816
miR-421	-1.35	0.0079	-1.63	0.000355
miR-767-5p	-1.39	0.028786	-2.72	0.000141
miR-1185-1-3p	-1.31	0.031671	-1.48	0.031286
miR-1226-3p	-1.41	0.013069	-1.38	0.012007
miR-1303	-2.29	0.006248	-1.68	0.001965
miR-1914-3p	-1.86	0.046571	-3.19	0.008125
miR-3136-5p	-1.74	0.018043	-1.96	0.01296
miR-3158-3p	-1.34	0.000396	-1.39	0.000451
miR-3187-3p	-1.59	0.015458	-1.34	0.008636
miR-4306	-1.61	0.048636	-1.76	0.010289
miR-3615	-1.5	0.002836	-1.34	0.003696
miR-4458	-3.51	0.007582	-9.48	6.70E-05
mir-101-1	-1.35	0.007524	-1.44	0.011727
mir-151a	-1.42	0.001445	-1.62	0.007278

Table 10. Fold-expression change values and p-values of miRNAs downregulated (FC ≤ -1.3, p < 0.05) following NRAD1 inhibition with GapmeR #3 and GapmeR #4 in SUM149 cells.

miRNA (hsa)	GapmeR #3 (FC)	GapmeR #3 (p-value)	GapmeR #4 (FC)	GapmeR #4 (p-value)
miR-484	-1.62	0.0059	-1.41	0.0019
miR-504-5p	-1.71	0.0168	-2.08	0.0034
miR-548a-5p	-1.71	0.0159	-1.53	0.0408
miR-936	-1.63	0.0245	-1.55	0.0273
miR-1226-3p	-1.52	0.0243	-1.88	0.0097
miR-1265	-1.65	4.00E-04	-1.41	0.0047
miR-3620-3p	-2.72	0.0056	-2.06	0.0255
miR-4458	-2.39	0.0077	-3.13	7.00E-04
miR-4521	-2.26	1.10E-06	-6.38	1.29E-12
miR-4709-5p	-1.55	0.0475	-1.74	0.0079
miR-4772-3p	-1.38	0.0142	-1.52	0.0164
miR-5571-5p	-1.97	0.0264	-2.17	0.009
miR-6509-3p	1.91	0.0065	-1.74	0.01
miR-7155-5p	-1.49	0.0027	-1.47	0.0055
miR-8075	-1.59	0.0034	-1.45	0.0265
mir-29b-2	-1.39	0.0148	-1.37	0.0393
mir-125b-1	-1.59	0.0155	-1.47	0.0354
mir-125b-1	-1.43	0.0166	-1.37	0.0286
mir-621	-1.34	0.0138	-1.38	0.01
mir-548i-4	-1.52	0.0014	-1.5	0.0026
mir-548i-4	-1.53	0.0209	-1.53	0.025
mir-513b	-1.51	0.0474	-1.45	0.0171
mir-4521	-1.59	0.0061	-1.78	5.00E-04
mir-4537	-1.41	0.0047	-1.35	0.0357
mir-4763	-1.71	0.0181	-1.59	0.0292
mir-5009	-1.38	0.0123	-1.49	0.0158
mir-6080	-2.34	0.0074	-2.08	0.0218
mir-6722	-1.62	0.0084	-1.34	0.0258
mir-6890	-1.52	0.0064	-1.58	0.0022

4.2 Investigating the interactions of miR-595 and miR-4485-3p with NRAD1

To investigate the potential for NRAD1 binding to miRNAs miR-4485-3p and miR-595, we determined whether the sequences of NRAD1 transcripts contained binding sites for each miRNA. Through this analysis we revealed potential binding sites for both miR-4485-3p (Fig. 32B) and miR-595 (Fig. 32A) on both the long and short NRAD1 transcripts. Intriguingly, miR-595 binding to NRAD1 appeared to be stronger, evidenced by its longer stretches of sequence complementarity with NRAD1 at the 3' end of each NRAD1 transcript (Fig. 32A). However, the sequence of miR-4485-3p showed sequence complementarity with the identical NRAD1 sequence region in both the long and short NRAD1 transcripts, marking it as a promising miRNA with which NRAD1 may interact (Fig. 32B). Together, the finding that both of our strongest miRNA microarray hits have the potential to bind NRAD1 helps support the reliability of our microarray data and prompts the further exploration of these miRNAs in the context of their roles in NRAD1-mediated gene regulation.

To investigate the possible roles of miR-4485-3p and miR-595 in NRAD1-mediated gene regulation, we used TargetScan Human to predict the mRNA targets of these miRNAs in human cells (Fig. 33A and B). This *in silico* tool predicts the mRNA targets of miRNAs through searching and identifying conserved 8-,7-, or 6mer sites matching specific miRNA seed regions[355]. Following the identification of the predicted mRNA targets of each miRNA, we identified genes that were downregulated following NRAD1 inhibition (upregulated by NRAD1, Control vs. GapmeR #4) in MDA-MB-468 cells using the microarray data shown in Figure 11. We then identified the list of mRNAs that were found to be targets of miR-4485-3p or miR-595 that were also downregulated by NRAD1 knockdown in our microarray. The rationale for this is that we propose that NRAD1 sponges miRNAs that would otherwise antagonize mRNAs encoding oncogenic proteins. Thus, mRNAs that are targets of a miRNA and are upregulated by NRAD1, points to the NRAD1-mediated sponging of these miRNAs, allowing upregulation of target mRNAs. Accordingly, TargetScan revealed 526 mRNA targets for miR-4485-3p (Fig. 33A) and 3535 targets for miR-595 (Fig. 33B). From the NRAD1 knockdown microarray, 2598 genes were found to be downregulated following NRAD1 inhibition. Together, we identified 53 genes that were predicted mRNA targets for miR-4485-3p

and downregulated by NRAD1 knockdown (Fig. 33A) and 409 genes that were predicted mRNA targets for miR-595 and downregulated by NRAD1 knockdown (Fig. 33B).

To narrow the list of mRNAs upregulated by NRAD1 that are predicted targets of miR-595 or miR-4485-3p, we plotted the TargetScan binding score for the miRNA with its predicted mRNA target versus the fold-expression change of the mRNA, identified via microarray in MDA-MB-468 cells with NRAD1 knockdown (fold-change ≥ 1.2, p < 0.05). The TargetScan binding score indicates the predicted efficacy of mRNA targeting by a miRNA, where higher binding scores correspond to a higher likelihood of targeting[355]. We used this analysis to identify mRNAs with both a high fold-expression change following NRAD1 knockdown and a TargetScan binding score of >0.2. Specifically, we identified 44 and 39 mRNAs that were upregulated by NRAD1 (fold-change ≥ 1.2, p < 0.05) and were predicted targets of miR-595 and miR-4485-3p, respectively (TargetScan binding score >0.2) (Figures 34 and 35). This data therefore provides evidence that miR-4485-3p and perhaps to a larger degree, miR-595, may be sponged by NRAD1 in breast cancer, supporting the hypothesis that NRAD1 regulates gene expression by acting as a miRNA sponge.

Potential binding sites of miR-595 and miR-4485-3p on NRAD1 transcripts

A

```
1896  5' AGACAAACUAUGGCUGCUUCUUCAAGCAAACCACCUCACCACACU 3' 1941        ————————    NRAD1 long transcript (3208 bp)
         |||||                          ||||.|..||||||
      3' UCUGU------------------------GUGGUGCCGUGUGAAG 5'          ————————    miR-595 sequence


                713 5' AAAC-CACCTCACCACACTATC 3' 733                ————————    NRAD1 short transcript (2521 bp)
                       |.|| ||||.|..|||||| ||
                    3' UCUGUGUGGUGCCGUGUGA-AG 5'                    ————————    miR-595 sequence
```

B

```
        2686 5' AGGGUACACACGCAGAGAAAAGGCCAU 3' 2712                 ————————    NRAD1 long transcript (3208 bp)
                |||||||   |||        ||||.|
             3' AAUCCCAUG---GCG-------CCGGCAAU 5'                   ————————    miR-4485-3p sequence


        1475 5' AGGGUACACACGCAGAGAAAAGGCCAU 3' 1501                 ————————    NRAD1 short transcript (2521 bp)
                |||||||   |||        ||||.|
             3' AAUCCCAUG---GCG-------CCGGCAAU 5'                   ————————    miR-4485-3p sequence
```

Figure 32. NRAD1-downregulated miRNAs, miR-595 and miR-4485-3p may bind to short and long NRAD1 transcripts. The putative binding sites of miR-595 and miR-4485-3p to short and long NRAD1 transcripts are shown. **(A)** Putative miR-595 binding sites on short (2521 bp) and long (3208 bp) NRAD1 transcripts. **(B)** Putative miR-4485-3p binding sites on short and long NRAD1 transcripts. The sequences of long (ENST00000423211.1) and short (ENST00000439707.6) NRAD1 transcripts were derived from Ensembl, and miR-595 (MIMAT0003263) and miR-4485-3p (MIMAT0019019) sequences were derived from miRBase. Complimentary regions between the NRAD1 and miRNA sequences are denoted with a vertical line, while differences between the two sequences within the binding site are shown with a dot. Dotted lines represent regions of non-complementarity.

mRNAs upregulated by NRAD1 in MDA-MB-468 cells and are predicted mRNA targets of miR-4485-3p or miR-595
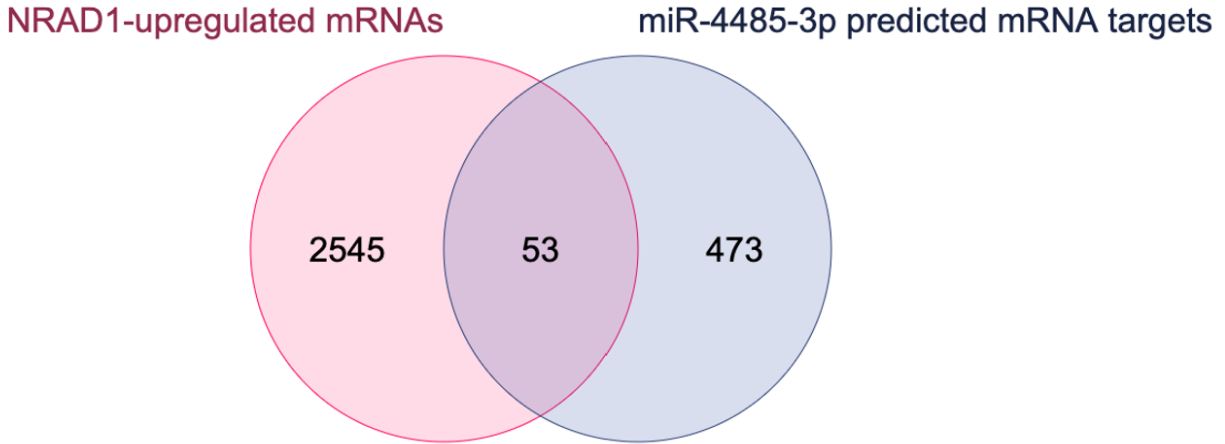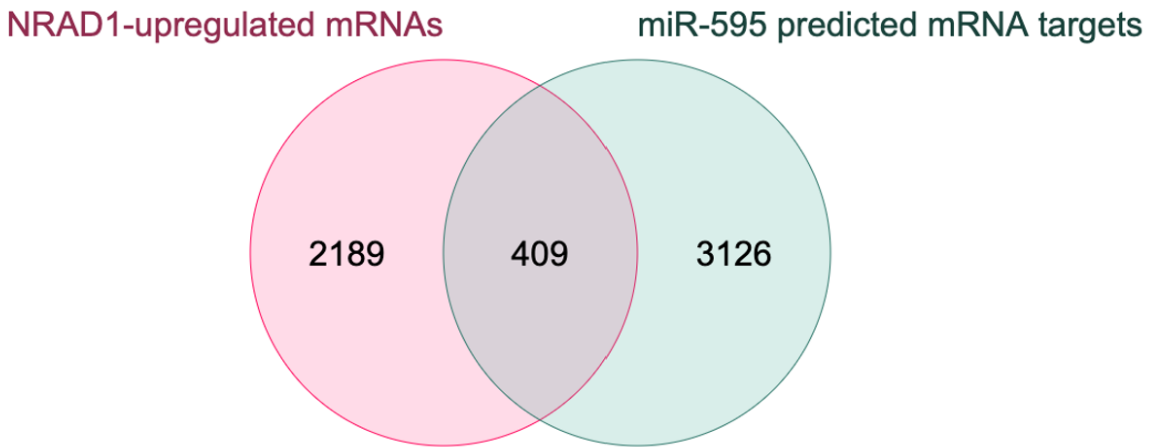
Figure 33. NRAD1 may interact with miR-4485-3p or miR-595 to regulate gene expression. Genome-wide gene expression changes induced by NRAD1 knockdown (Control vs. GapmeR #4) were quantified in MDA-MB-468 cells using the Affymetrix Human Gene 2.0 ST microarray platform (n=3). Genes upregulated by NRAD1 (FC $\geq$ 1.2, p < 0.05) were identified and compared with predicted mRNA targets of miR-4485-3p or miR-595 identified using TargetScan human. **(A)** mRNAs upregulated by NRAD1 in MDA-MB-468 cells (downregulated upon GapmeR #4-mediated knockdown) that are also predicted targets of miR-4485-3p. **(B)** mRNAs upregulated by NRAD1 in MDA-MB-468 cells (downregulated upon GapmeR #4-mediated knockdown) that are also predicted targets of miR-595.

Fold-change in expression versus binding score of predicted miR-595 mRNA targets upregulated by NRAD1
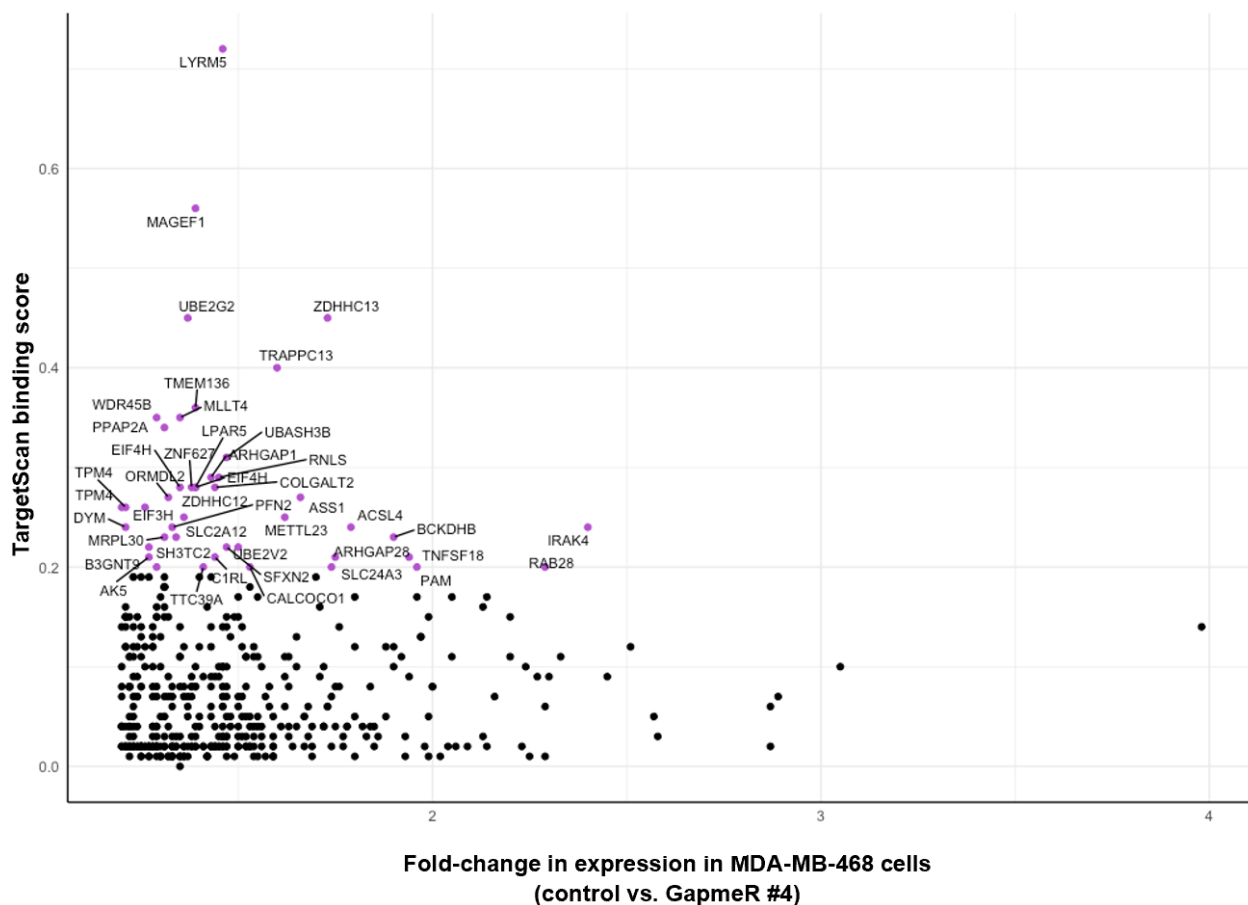
Figure 34. Analysis of predicted miR-595 mRNA targets also upregulated by NRAD1 identifies key mRNAs for downstream experiments. Changes in gene expression in response to NRAD1 knockdown (control vs. GapmeR #4) in MDA-MB-468 cells were quantified using the Affymetrix Human Gene 2.0 ST Array platform. The fold-expression change of genes upregulated by NRAD1 ($\geq$ 1.2-fold change, $p < 0.05$) that are also predicted targets of miR-595 (TargetScan) are plotted versus the predicted TargetScan binding score of miR-595 binding to each mRNA. mRNAs with a TargetScan binding score > 0.2 are coloured purple.

Fold-change in expression versus binding score of predicted miR-4485-3p mRNA targets upregulated by NRAD1
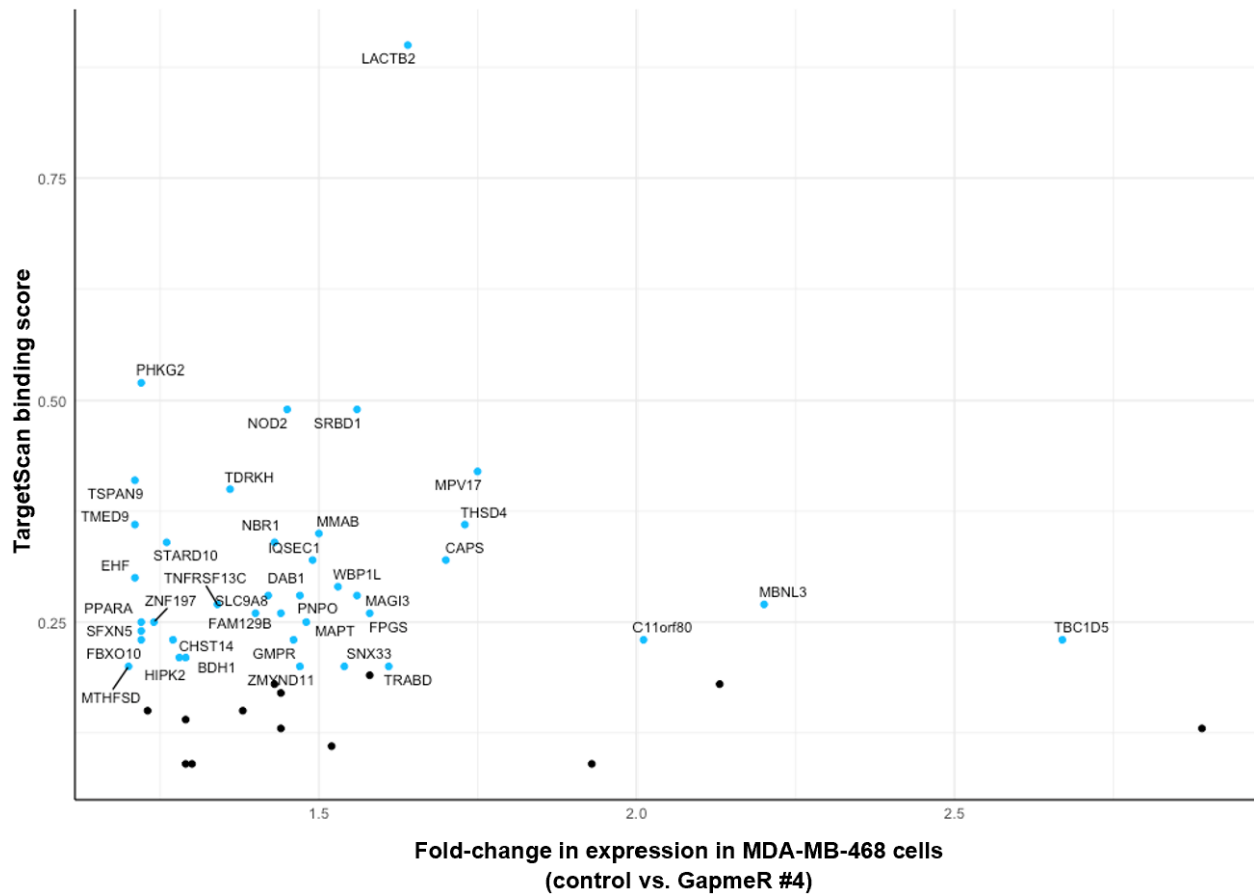
Figure 35. Analysis of predicted miR-4485-3p mRNA targets also upregulated by NRAD1 identifies key mRNAs for downstream experiments. Changes in gene expression in response to NRAD1 knockdown (control vs. GapmeR #4) in MDA-MB-468 cells were quantified using the Affymetrix Human Gene 2.0 ST Array platform. The fold-expression change of genes upregulated by NRAD1 ($\geq$ 1.2-fold change, $p < 0.05$) that are also predicted targets of miR-4485-3p (TargetScan) are plotted versus the predicted TargetScan binding score of miR-4485-3p binding to each mRNA. mRNAs with a TargetScan binding score > 0.2 are coloured blue.

4.3 ALDH1A3 regulates miRNA expression in MDA-MB-468 and SUM149 cells.

As I described earlier, NRAD1 contributes to ALDH1A3 gene expression[302]. We therefore wondered if the gene expression changes induced by ALDH1A3/NRAD1 could be explained by common changes in miRNAs. I therefore performed the same miRNA microarray experiments using two separate ALDH1A3 shRNA knockdowns (61 and 63) in MDA-MB-468 and SUM149 cells to identify changes in miRNA expression commonly induced by ALDH1A3 and NRAD1 knockdown. Like NRAD1, ALDH1A3 knockdown promoted changes in miRNA expression, where knockdown of ALDH1A3 using shRNA-61 or shRNA-63 induced miRNA expression changes meeting our cutoff criteria in both MDA-MB-468 and SUM149 cells (Fig. 36A, B, C, and D). Specifically, ALDH1A3 knockdown with shRNA-61 resulted in upregulation of 91 and downregulation of 114 miRNAs in MDA-MB-468 cells (Fig. 36C, Table 11) and up- or downregulation of 111 and 108 miRNAs respectively in SUM149 cells (Fig. 36A, Table 11). Similarly, knockdown with shRNA-63 led to up- or downregulation of 124 and 109 miRNAs in MDA-MB-468 cells (Fig. 36D, Table 11) and 106 and 87 miRNAs in SUM149 cells respectively (Fig. 36B, Table 11). Further, ALDH1A3 knockdown with different shRNAs resulted in similar miRNA expression changes within both cell lines with respect to ALDH1A3-induced miRNA expression changes of all miRNAs in the microarray (Fig. 37). Thus, ALDH1A3 regulates miRNA expression in TNBC cell lines. Importantly, no miRNAs were found to be up or downregulated by both shRNA 61 and 63 in both MDA-MB-468 and SUM149 cells (Fig. 38A and B). However, significant overlap in miRNA expression changes induced by different ALDH1A3 knockdowns within each cell line was observed. To this end, 26 miRNAs were upregulated by ALDH1A3 knockdown with both shRNAs in MDA-MB-468 cells (Fig. 38A, Table 12) and 21 miRNAs were upregulated by knockdown with both shRNAs in SUM149 cells (Fig. 38A, Table 12). Similarly, 43 miRNAs were downregulated following ALDH1A3 knockdown with both shRNAs in MDA-MB-468 cells (Fig. 38B, Table 14) and 34 miRNAs were downregulated by knockdown with both shRNAs in SUM149 cells (Fig. 38B, Table 15).

ALDH1A3-induced miRNA expression changes in MDA-MB-468 and SUM149 cells
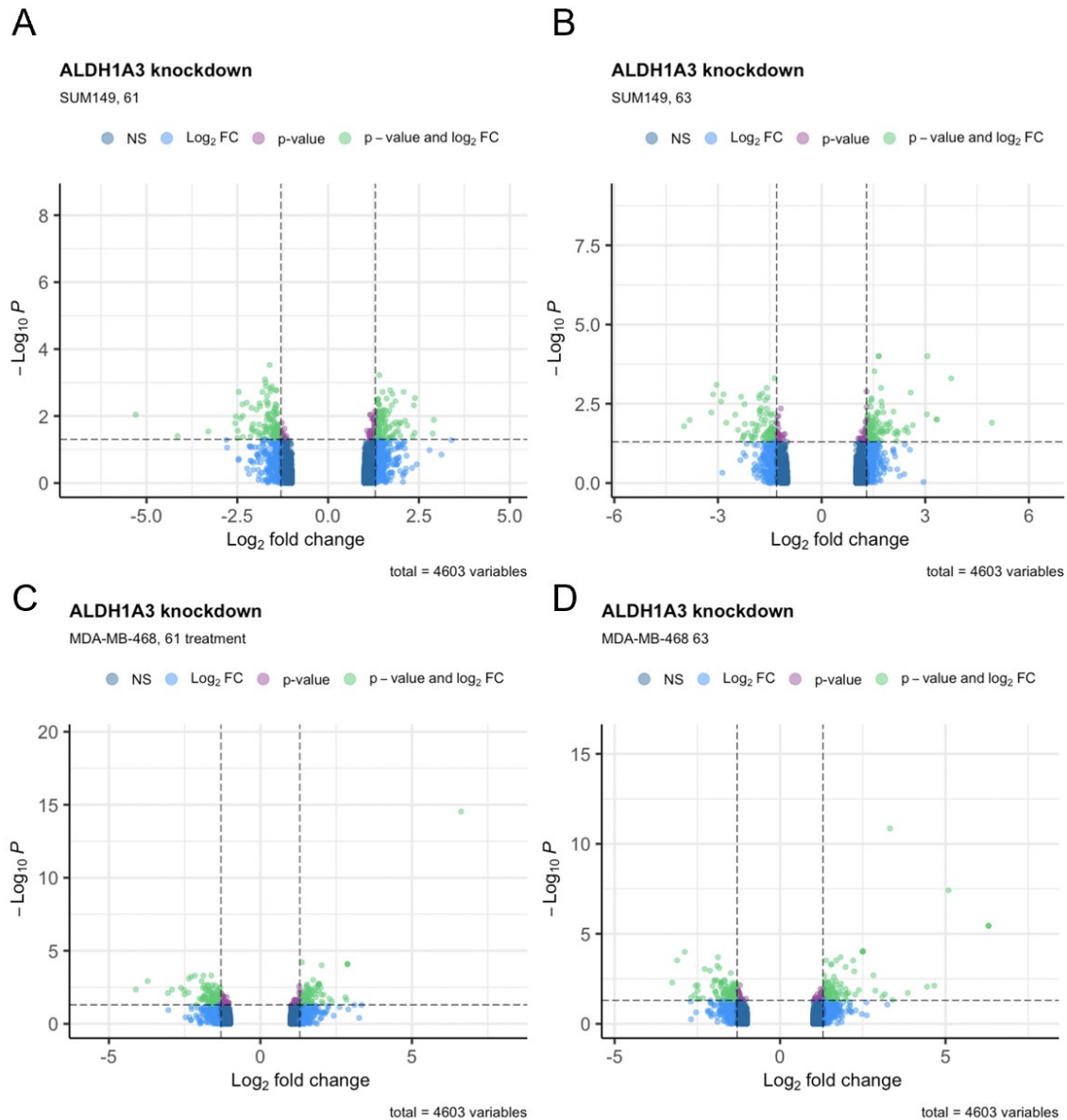
Figure 36. ALDH1A3 induces miRNA expression changes in MDA-MB-468 and SUM149 cells. Genome-wide miRNA expression changes induced by shRNA-mediated ALDH1A3 knockdown (control vs. shRNA-61 or shRNA-63) was quantified in MDA-MB-468 and SUM149 cells using the Affymetrix GeneChip miRNA 4.0 Array (n=3). The Log_2 fold-change in miRNA expression is plotted versus the –log10(ANOVA p-value) of probes corresponding to 2578 mature human miRNAs and 2025 human pre-miRNAs (4603 total). Dots coloured light green show miRNAs with a fold-expression change of ≥ 1.3 or ≤ -1.3 and a p-value < 0.05. ALDH1A3 knockdown with shRNA-61 or 63 in SUM149 cells is shown in panels **A** and **B** respectively. ALDH1A3 knockdown with shRNA-61 or 63 in MDA-MB-468 cells is shown in panels **C** and **D** respectively.

Table 11. Changes in gene expression in response to ALDH1A3 inhibition in MDA-MB-468 and SUM149 cells. The number of miRNAs upregulated (FC ≥ 1.3, p < 0.05) or downregulated (FC ≤ -1.3, p < 0.05) by ALDH1A3 knockdown (shRNA-61 or shRNA-63) in MDA-MB-468 or SUM149 cells are shown. Gene expression changes were quantified using the Affymetrix GeneChip miRNA 4.0 Array (n=3).

Cell line	shRNA	Upregulated miRNAs	Downregulated miRNAs
MDA-MB-468	61	91	114
	63	124	109
SUM149	61	111	108
	63	106	87
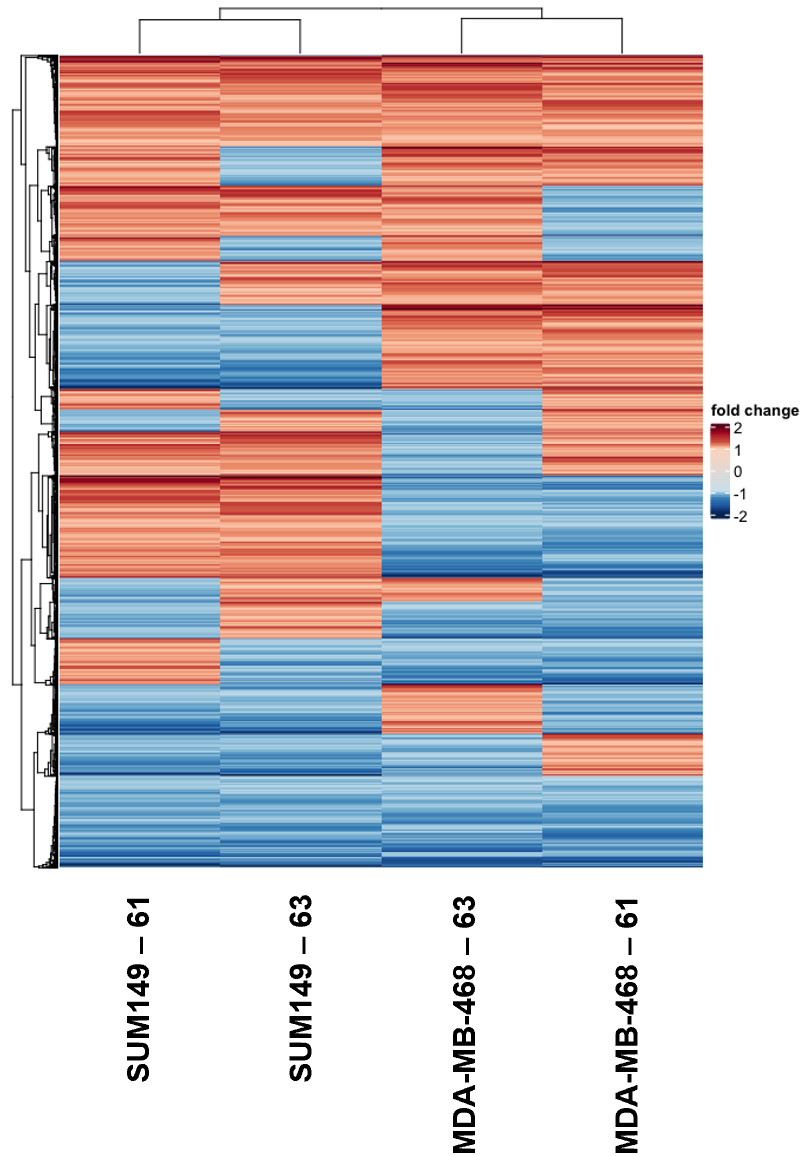
Figure 37. ALDH1A3 inhibition with shRNA-61 or shRNA-63 induces similar gene expression changes. Fold-change in expression (Control vs. shALDH1A3) of 4603 miRNAs quantified using the Affymetrix GeneChip miRNA 4.0 array following ALDH1A3 knockdown with shRNA-61 or shRNA-63 in MDA-MB-468 or SUM149 cells (n=3).

ALDH1A3-regulated miRNAs in MDA-MB-468 and SUM149 cells

A   **miRNAs upregulated upon ALDH1A3 knockdown**
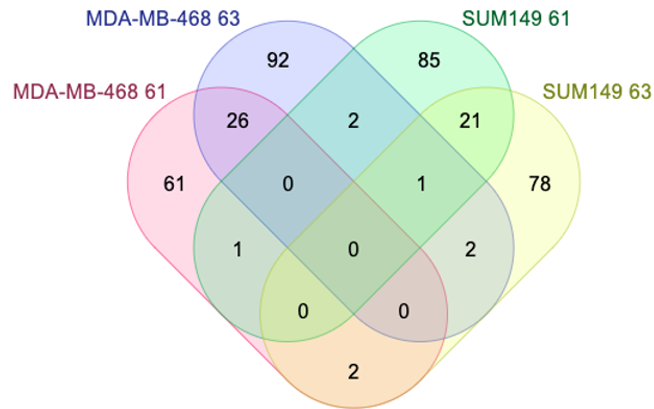
B   **miRNAs downregulated upon ALDH1A3 knockdown**
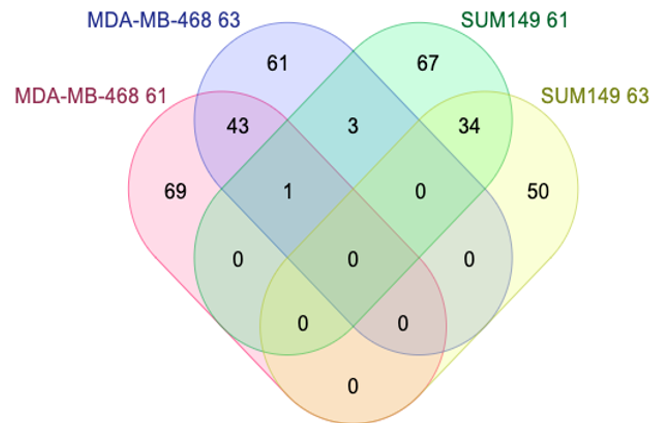
Figure 38. miRNAs regulated by ALDH1A3 knockdown with different shRNAs show similarities within cell lines. Changes in miRNA expression were quantified using the Affymetrix GeneChip miRNA 4.0 Array following ALDH1A3 knockdown with shRNA-61 or shRNA-63 in MDA-MB-468 and SUM149 cells (n=3). The miRNAs upregulated (FC ≥ 1.3, p < 0.05) **(A)** or downregulated (FC ≤ -1.3, p < 0.05) **(B)** following ALDH1A3 knockdown with shRNA-61 or 63 in MDA-MB-468 or SUM149 cells are shown. Commonly upregulated or downregulated miRNAs are shown by overlapping regions in the Venn diagram.

Table 12. Fold-expression change values and p-values of miRNAs upregulated (FC ≥ 1.3, p < 0.05) following ALDH1A3 knockdown with shRNA-61 and shRNA-63 in MDA-MB-468 cells.

miRNA	shRNA-61 (FC)	shRNA-61 (p-value)	shRNA-63 (FC)	shRNA-63 (p-value)
miR-29a-3p	1.48	0.0262	1.35	0.036
miR-100-5p	6.62	2.93E-15	3.32	1.41E-11
miR-125b-5p	1.37	6.12E-05	1.39	3.00E-04
miR-519c-5p	1.93	0.0019	2.5	9.53E-05
miR-519b-5p	1.93	0.0019	2.5	9.53E-05
miR-523-5p	1.93	0.0019	2.5	9.53E-05
miR-526a	2.87	8.15E-05	6.3	3.59E-06
miR-520c-5p	2.87	8.15E-05	6.3	3.59E-06
miR-518e-5p	1.93	0.0019	2.5	9.53E-05
miR-518d-5p	2.87	8.15E-05	6.3	3.59E-06
miR-522-5p	1.93	0.0019	2.5	9.53E-05
miR-519a-5p	1.93	0.0019	2.5	9.53E-05
miR-92b-5p	1.65	0.0043	1.51	0.0406
miR-542-5p	1.83	0.0154	2.88	0.0144
miR-1909-3p	2.84	0.0223	3.09	0.0305
miR-3195	2.8	0.0152	2.38	0.0134
miR-4756-5p	1.52	0.0026	1.63	0.0024
miR-6804-3p	2.28	0.0191	1.96	0.0398
mir-125b-2	1.58	0.0053	1.51	0.0335
mir-193b	1.54	0.002	1.7	3.00E-04
mir-519b	1.91	0.0053	5.09	3.82E-08
mir-3918	1.74	0.001	1.37	0.0061
mir-4721	1.51	0.0188	1.68	0.0071
mir-5095	1.8	0.0053	1.98	7.00E-04
mir-6743	1.82	0.0091	1.68	0.0174
mir-7158	1.43	0.0119	1.3	0.0427

Table 13. Fold-expression change values and p-values of miRNAs downregulated (FC ≤ -1.3, p < 0.05) following ALDH1A3 knockdown with shRNA-61 and shRNA-63 in MDA-MB-468 cells.

miRNA	shRNA-61 (FC)	shRNA-61 (p-value)	shRNA-63 (FC)	shRNA-63 (p-value)
let-7f-1-3p	-1.74	0.0182	-2.04	0.0078
miR-192-3p	-1.88	5.00E-04	-1.9	6.00E-04
miR-204-5p	-1.4	0.0355	-1.42	0.0171
miR-181a-3p	-2.37	0.0073	-2.49	0.0073
miR-144-5p	-1.43	0.006	-1.35	0.0115
miR-200a-3p	-1.31	0.0079	-1.43	0.0074
miR-363-3p	-1.68	0.0049	-1.91	0.0038
miR-373-5p	-1.46	0.0454	-1.45	0.0148
miR-504-3p	-2.28	6.00E-04	-1.76	0.0112
miR-505-5p	-1.62	0.0118	-2.28	9.00E-04
miR-553	-1.73	0.0089	-1.58	0.0309
miR-592	-1.36	0.0091	-1.37	0.0159
miR-593-3p	-1.86	0.0024	-1.88	2.00E-04
miR-641	-1.94	0.021	-1.45	0.0204
miR-543	-1.71	0.0025	-1.66	0.0121
miR-1228-3p	-1.93	0.0225	-2.88	1.00E-04
miR-3154	-3.72	0.0012	-2.58	0.024
miR-6499-5p	-1.35	0.0442	-1.46	0.0141
miR-6746-3p	-1.78	0.0168	-2.12	0.0011
miR-6793-5p	-1.6	0.0186	-1.98	0.004
miR-6827-5p	-1.56	0.0464	-1.95	0.0166
miR-7109-3p	-3.05	0.0079	-2.44	0.0413
mir-29c	-1.49	0.0392	-1.33	0.0454
mir-490	-1.62	0.0071	-1.51	0.0185
mir-605	-1.34	0.03	-1.41	0.0199
mir-619	-1.58	0.0022	-1.42	0.0143
mir-943	-1.59	0.0202	-1.66	0.0043
mir-1538	-2.36	8.00E-04	-1.88	0.036
mir-3181	-1.61	0.0105	-1.45	0.0128
mir-3192	-1.43	0.009	-1.46	0.0328
mir-4292	-1.42	0.0396	-1.59	0.0086
mir-4638	-1.53	0.0091	-1.65	0.0078
mir-4757	-1.48	0.0024	-1.4	0.0149
mir-4802	-1.38	0.0245	-1.66	0.0146
mir-6511a-1	-1.87	0.0159	-1.78	0.0051
mir-6511b-1	-1.87	0.0159	-1.78	0.0051
mir-6848	-1.51	0.042	-1.59	0.0148
mir-6876	-1.57	0.0396	-1.55	0.0449
mir-6511b-2	-1.87	0.0159	-1.78	0.0051
mir-6511a-2	-1.87	0.0159	-1.78	0.0051
mir-6511a-3	-1.87	0.0159	-1.78	0.0051
mir-6511a-4	-1.87	0.0159	-1.78	0.0051
mir-7850	-1.35	0.0037	-1.34	0.0085
mir-8061	-1.62	5.00E-04	-1.31	0.0059

Table 14. Fold-expression change values and p-values of miRNAs upregulated (FC ≥ 1.3, p < 0.05) following ALDH1A3 knockdown with shRNA-61 and shRNA-63 in SUM149 cells.

miRNA	shRNA-61 (FC)	shRNA-61 (p-value)	shRNA-63 (FC)	shRNA-63 (p-value)
miR-29b-3p	2.38	0.0317	2.5	0.0252
miR-181a-2-3p	2.36	0.0048	2.58	0.0014
miR-181b-5p	1.35	0.0274	1.41	0.0111
miR-212-3p	1.61	0.003	1.31	0.0135
miR-152-3p	1.53	0.0211	1.75	0.0198
miR-146a-5p	2.08	0.0019	3.06	1.00E-04
miR-195-5p	1.37	0.0115	1.32	0.0179
miR-301a-3p	1.93	0.0188	1.98	0.0122
miR-99b-3p	1.36	0.0366	1.36	0.0193
miR-346	1.85	0.0036	1.5	0.0325
miR-181d-5p	1.36	0.0032	1.3	0.0175
miR-1301-3p	1.69	0.0069	1.75	0.0038
miR-889-5p	1.32	0.0258	1.35	0.0069
miR-940	1.52	0.0096	3.34	0.0101
miR-1228-3p	2.12	0.0139	1.91	0.0229
miR-3147	2.88	0.0333	2.17	0.0474
miR-3157-3p	1.38	0.0072	1.46	0.011
miR-4472	2.91	0.013	3.75	5.00E-04
miR-4731-5p	1.75	0.0048	1.31	0.0352
miR-6511a-3p	1.77	0.0179	1.58	0.0151
miR-8085	2.39	0.0029	2.08	0.0056
mir-1286	1.64	0.0087	1.43	0.0406

Table 15. Fold-expression change values and p-values of miRNAs downregulated (FC ≤ -1.3, p < 0.05) following ALDH1A3 knockdown with shRNA-61 and shRNA-63 in SUM149 cells.

miRNA	shRNA-61 (FC)	shRNA-61 (p-value)	shRNA-63 (FC)	shRNA-63 (p-value)
miR-182-3p	-1.7	0.0036	-1.61	0.0066
miR-187-3p	-1.81	0.0368	-2.84	0.0016
miR-372-3p	-1.63	0.0084	-1.62	0.0101
miR-564	-1.73	0.0302	-2.22	0.0138
miR-629-5p	-1.3	0.049	-1.35	0.0451
miR-2682-3p	-1.45	0.0236	-1.5	0.0441
miR-466	-2.24	0.004	-3.14	0.0016
miR-3149	-1.57	0.0218	-1.75	0.0043
miR-3173-3p	-1.35	0.0492	-1.49	0.0371
miR-3175	-2.04	0.0288	-2.91	0.0027
miR-4458	-1.72	0.0196	-2.5	0.007
miR-4489	-1.5	0.0307	-1.66	0.0138
miR-4657	-1.5	0.0132	-1.59	0.0045
miR-3529-3p	-1.61	0.0119	-1.8	0.0021
miR-5684	-2.1	0.0256	-2.25	0.0306
miR-6077	-1.45	0.0459	-1.54	0.0448
miR-6801-5p	-2.07	0.0175	-1.9	0.0145
miR-6884-5p	-1.39	0.0079	-1.41	0.028
miR-6887-5p	-1.66	0.0087	-1.45	0.0127
let-7e	-1.59	0.0015	-1.77	0.0054
mir-92a-1	-1.4	0.0179	-1.8	0.0017
mir-149	-1.81	0.0019	-1.55	0.0242
mir-634	-1.46	0.009	-1.59	0.001
mir-320b-1	-1.47	0.0159	-1.5	0.0234
mir-921	-1.97	0.0064	-1.64	0.0289
mir-1284	-1.63	0.0215	-1.78	0.0068
mir-1324	-1.47	0.0455	-1.34	0.0379
mir-4716	-1.61	3.00E-04	-1.56	0.002
mir-4716	-1.55	0.0013	-1.33	0.0361
mir-4758	-1.51	0.0272	-1.7	0.0376
mir-4758	-2.36	0.0045	-2.23	0.0033
mir-1295b	-1.74	8.00E-04	-1.63	0.0024
mir-6793	-1.44	0.0017	-1.36	5.00E-04
mir-6799	-1.53	0.0115	-1.74	0.0099
mir-8071-1	-1.4	0.0189	-1.45	0.0176
mir-8071-2	-1.4	0.0189	-1.45	0.0176

4.4 Investigating NRAD1 and ALDH1A3 co-regulated miRNA expression in MDA-MB-468 and SUM149 cells.

To identify miRNA targets whose expression is commonly regulated by ALDH1A3 and NRAD1, we plotted the fold-expression change (control vs. knockdown) of all 4603 miRNAs in the microarray for each NRAD1 and ALDH1A3 knockdown (GapmeRs 3 and 4 and shRNAs 61 and 63) in both MDA-MB-468 and SUM149 cells in a heatmap (Fig. 39). This analysis showed that changes in miRNA expression were most similar between the two different knockdowns of a single gene within a single cell line (i.e. GapmeR #3 and GapmeR #4-mediated NRAD1 inhibition in MDA-MB-468 cells). However, miRNA expression changes induced by ALDH1A3 knockdown were more closely related to changes in NRAD1-mediated miRNA expression changes in the same cell line than to ALDH1A3-induced miRNA expression changes from the other cell line. To illustrate this, miRNA expression changes induced by ALDH1A3 knockdown with shRNAs 61 and 63 in SUM149 cells were most closely related to the miRNA expression changes induced by NRAD1 knockdown with GapmeRs 3 and 4 in SUM149 cells. The same result was demonstrated in MDA-MB-468 cells. Thus, ALDH1A3 may contribute to NRAD1-mediated regulation of miRNA expression in a cell line-specific manner. Importantly, four miRNAs (mir-4690, miR-6780b-5p, mir-711, and miR-4303) were identified as upregulated across all knockdowns and cell lines when a cutoff of ≥1.1 fold-expression change (control vs. knockdown) was applied. Similarly, miR-5187-5p, miR-95-5p, miR-6779-3p, and miR-6848-5p were identified as downregulated (FC ≤ -1.1) in all conditions and cell lines.

To find miRNAs with greater likelihood of functional relevance that would serve as more reliable targets for *in vitro* analyses, we applied the cutoff of FC ≥ 1.3 or ≤ -1.3, p < 0.05. Of the miRNAs that met this cutoff, none were commonly upregulated by NRAD1 inhibition with both GapmeR #3 and #4 and ALDH1A3 knockdown with both shRNA-61 and shRNA-63 within a cell line (Fig. 40A). We showed the same results with the downregulated miRNAs apart from one miRNA, miR-4458, that was shown to be downregulated by NRAD1 inhibition with both GapmeRs and ALDH1A3 knockdown with both shRNAs in SUM149 cells (Fig. 40B). This miRNA may therefore represent a miRNA regulated by NRAD1 and

ALDH1A3 in this cell line, warranting further investigation.

Together, these results suggest that ALDH1A3 does not regulate miRNA expression via its induction of NRAD1, at least not at a level comparable to ALDH1A3-mediated induction of NRAD1 and its regulation of nuclear gene expression. However, given our evidence of ALDH1A3-mediated miRNA regulation in different cell lines, ALDH1A3 may induce the expression of different lncRNAs that may interact with miRNAs to regulate gene expression in breast cancer.
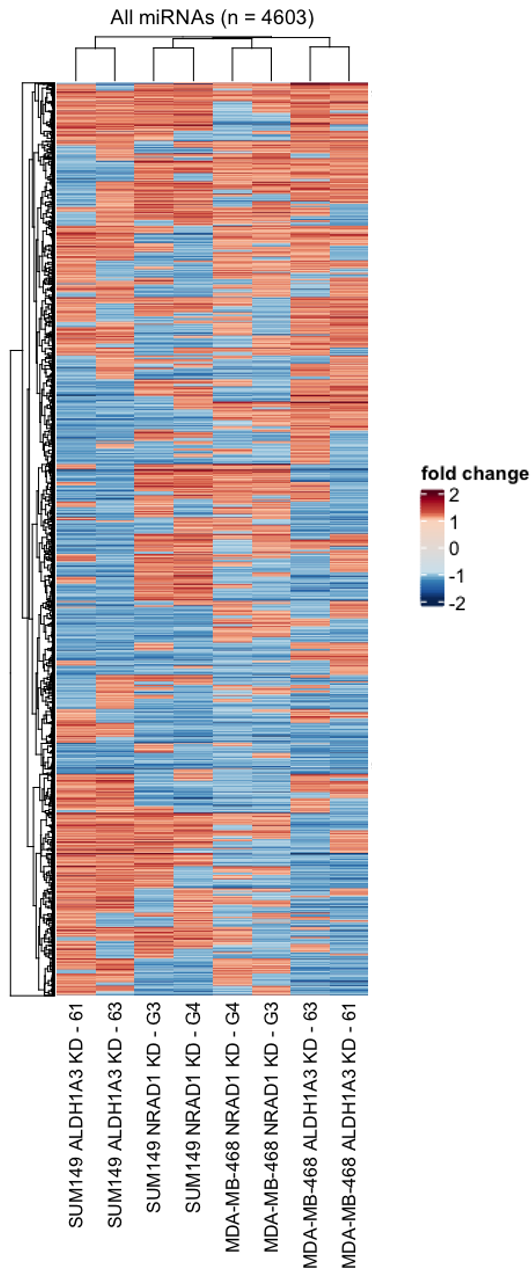
Figure 39. Changes in miRNA expression induced by NRAD1 knockdown show partial overlap with ALDH1A3 knockdown-induced miRNA expression changes in a cell line-specific manner. Changes in miRNA expression in the context of NRAD1 (GapmeR #3 or GapmeR #4) or ALDH1A3 (shRNA-61 or shRNA-63) knockdown in MDA-MB-468 and SUM149 cells were quantified using the Affymetrix GeneChip miRNA 4.0 Array (n=3). The fold-miRNA expression change (control vs. knockdown) of miRNAs up- or downregulated by NRAD1 inhibition with GapmeR #3 and GapmeR #4, and ALDH1A3 knockdown with shRNA-61 and shRNA-63 in MDA-MB-468 and SUM149 cells are plotted in the heatmap (n = 4603 miRNAs).

miRNAs Regulated by NRAD1 or ALDH1A3 in
MDA-MB-468 and SUM149 cells

A   **miRNAs upregulated upon NRAD1 or ALDH1A3 knockdown**
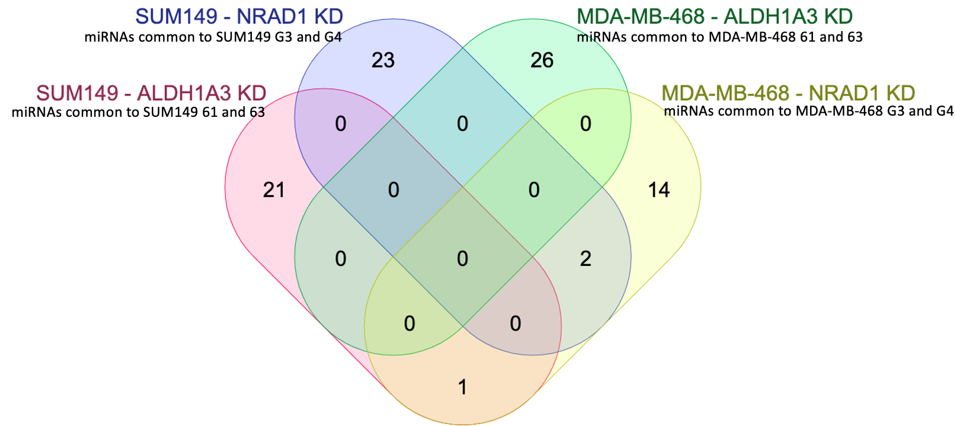
B   **miRNAs downregulated upon NRAD1 or ALDH1A3 knockdown**
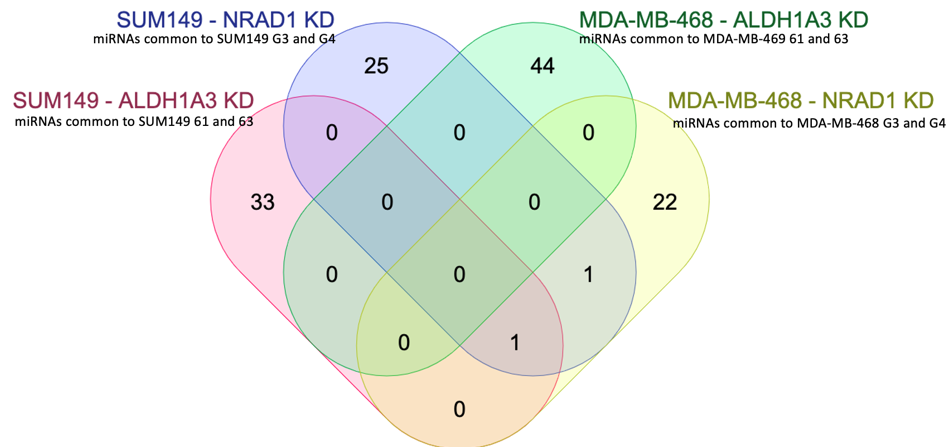
Figure 40. NRAD1 and ALDH1A3 do not induce common miRNA expression changes within MDA-MB-468 or SUM149 cells. Changes in miRNA expression following NRAD1 (GapmeR #3 or GapmeR #4) or ALDH1A3 (shRNA-61 or shRNA-63) knockdown in MDA-MB-468 and SUM149 cells were quantified using the Affymetrix GeneChip miRNA 4.0 Array (n=3). miRNAs upregulated **(A)** or downregulated **(B)** by NRAD1 knockdown with both GapmeR #3 and GapmeR #4, or ALDH1A3 knockdown with shRNA-61 and shRNA-63 (FC ≥ 1.3 or ≤ -1.3, p < 0.05) are shown in the Venn diagrams.

CHAPTER 5 – DISCUSSION

5.1 Preamble

The limited number of targeted treatment options for TNBCs remains a hindrance to the effective widespread treatment of TNBC patients. Specifically lacking are treatments that target TNBC-enriched breast CSCs which contribute to disease progression and recurrence through their resistance to treatments and their self-renewing ability[69,72,73]. The discovery of novel anti-CSC therapies, therefore, remains a clinically relevant undertaking. Much of the research focused on CSC-directed targeted therapies has focused on CSC-enriched targets such as ALDH1A3[117], members of the Wnt, Notch, and Hedgehog developmental signaling pathways[154], and players in immune checkpoints such as PD-L1[200]. Numerous studies have recently implicated lncRNAs as key players in CSC dynamics and thus, potential targets for anti-CSC therapies[268].

To mediate CSC activity, lncRNAs often interact with regulatory protein or miRNA partners, which may represent promising targets for abrogation of oncogenic lncRNAs associated with CSC functions. One such lncRNA enriched in TNBCs and breast CSC populations is NRAD1, an ALDH1A3/retinoic acid-induced mediator of TNBC cell survival, tumor growth, and acquisition of CSC characteristics[302]. NRAD1 was shown to regulate gene expression, partially contributing to ALDH1A3-mediated gene regulation in TNBC, providing a foundation for how NRAD1 promotes the gain of oncogenic attributes in TNBC. However, the mechanism through which NRAD1 modulates changes in gene expression remains to be determined.

Intriguingly, NRAD1 has multiple transcript variants with somewhat distinct subcellular localization profiles, which may indicate the presence of different mechanisms of gene expression regulation. Here, we explore the interactions between NRAD1 and protein or miRNA partners to identify putative novel mediators of NRAD1 functions in TNBC. Our use of lncRNA-directed proteomics methods, RNA immunoprecipitation, and shRNA knockdown followed by QPCR have identified the protein S100A8 as a putative NRAD1-binding protein with possible functions in ALDH1A3 and

NRAD1-mediated regulation of gene expression in TNBC. Further, using miRNA microarray analyses, we have identified changes in the miRNA landscape in response to NRAD1 or ALDH1A3 knockdown in different TNBC cell lines and identified miRNAs predicted to interact with NRAD1. Together, we have provided a framework for the possible mechanisms through which NRAD1 regulates gene expression in TNBC.

5.2 NRAD1 is a primarily nuclear lncRNA with multiple transcript variants and distinct subcellular localization

Many lncRNAs have multiple transcript variants that exhibit differential tissue expression patterns and therefore, biological functions. For example, the lncRNA SOX2OT (SOX2 Overlapping Transcript) exerts oncogenic functions in several human cancers and produces eight transcript variants which demonstrate differential expression patterns in different human cancer cell or tissue types[363,364]. Chang and colleagues investigated the expression profiles of different SOX2OT transcript variants in cervical cancer cells and showed that one transcript variant, SOX2OT-7, was associated with the stemness marker SOX2 and progression of specific cervical cancer types and therefore, represented a potential biomarker and therapeutic target for cervical cancer[364]. Thus, understanding the roles of specific lncRNA transcript variants will critically inform targeted efforts to abrogate the functions of oncogenic lncRNAs.

It is well-understood that lncRNA function depends on subcellular location, where nuclear lncRNAs modulate changes in gene expression through regulation of transcriptional activity, and cytoplasmic lncRNAs influence protein synthesis by interacting specific miRNAs. These location-specific lncRNA functions are particularly relevant in the context of cancers and can promote cancer progression or suppression.

Regulation of gene expression and cancer progression via nuclear lncRNAs is often achieved through interactions with chromatin. For example, the nuclear lncRNA NEAT1 functions as a chromatin remodeler, contributing to colorectal cancer stemness and resistance to 5-fluorouracil[365]. Specifically, ATAC-seq (Assay for Transposase Accessible Chromatin by sequencing) and chromatin

immunoprecipitation assays demonstrated that NEAT1 remodels chromatin, increasing H3K27 acetylation which prompted an increase in acetylation of ALDH1 and c-MYC promoter regions, increasing ALDH1 and c-MYC expression and enhancing colorectal cancer cell stemness. In addition, NEAT1 was shown to regulate the expression of key stemness genes SOX2, NANOG, OCT4, and c-MYC, contributing to CSC properties among colorectal cancer cells.

Cytoplasmic lncRNAs are known to regulate mRNA stability, often serving as ceRNAs that sequester miRNAs, preventing miRNA-mediated degradation of their target mRNAs[366]. Oncogenic cytoplasmic lncRNAs often take advantage of this, sponging miRNAs that would otherwise antagonize mRNAs encoding proteins contributing to cancer progression. For example, lncRNA H19 promotes breast cancer progression and is enriched in the ALDH[high] breast CSC populations of TNBCs[300]. H19 sponges miRNA tumor suppressor, let-7, facilitating a concomitant increase in the breast CSC-enriched pluripotency factor LIN28[303]. The sponging of let-7 by H19 also leads to the increased expression of the glycolytic enzyme pyruvate dehydrogenase kinase 1 (PDK1), which promotes the metabolic reprogramming of breast CSCs[299].

We have found that NRAD1 has multiple transcript variants (Fig. 3) that demonstrate different expression levels in the cytoplasm and nucleus of MDA-MB-468 cells (Fig. 4A, B, C, and D). Subcellular fractionation experiments followed by QPCR showed that both NRAD1 transcripts are expressed primarily in the nucleus (Fig. 4A, C, and D). Importantly, significantly reduced levels of the long NRAD1 transcript were detected in the cytoplasm of MDA-MB-468 cells (Fig. 4C). While expression of the short transcript was significantly higher in nuclear relative to cytoplasmic fractions, it demonstrated a mean expression level in the cytoplasm that was approximately half of its mean expression level in the nucleus (Fig. 4D). In addition, expression of the short transcript was significantly higher than the long transcript in cytoplasmic fractions (Fig. 4B). Thus, these results suggest that if the lncRNA does function in the cytoplasm, it may be predominately driven by the short transcript.

The Marcato lab has previously characterized the functions of the long NRAD1 transcript in the nucleus, where nuclear NRAD1-mediated gene expression changes are associated with chromatin

interactions[302]. However, these NRAD1-mediated changes in gene expression have not yet been linked directly to breast cancer progression, nor does it necessarily mean that the gene expression changes are mediated by the lncRNA's interactions with chromatin. Other studies have also demonstrated the role of NRAD1 in gene regulation. Accordingly, NRAD1 contributes to angiogenesis during ovarian cancer formation via recruitment of NF-κB1 to the MEST promoter and downregulation of MEST, evidenced by *in vitro* and *in vivo* studies[339]. In addition to the nuclear roles of NRAD1, some studies have shown that NRAD1 also sponges miRNAs, supporting its cytoplasmic activity[340,343,347]. Yan and colleagues (2021) most recently performed subcellular fractionation experiments using oral squamous cell carcinoma cells (OSCC) to demonstrate that NRAD1 was expressed primarily in the cytoplasm[347]. Consistently, NRAD1 was shown to function as a ceRNA, sponging miR-211-3p in OSCC cells, leading to upregulation of the MAFG transcription factor and promoting malignant characteristics of OSCC cells. The potential regulatory interactions of NRAD1 with miRNAs in breast cancer have not yet been explored.

Despite the evidence for NRAD1-mediated gene regulation in TNBC, the mechanism through which NRAD1 acts to confer gene expression changes has yet to be determined. The first aim of my work was therefore to construct a foundation upon which the framework for the mechanism of NRAD1-imparted gene regulation could be built. To this end, I sought to identify an NRAD1-binding protein that may be linked to the regulatory activity of the long NRAD1 transcript in the nucleus of TNBC cells. This work will comprise the majority of my thesis. The second aim of my work was to elucidate potential interactions between NRAD1 and miRNAs in the cytoplasm of these cells.

Given that NRAD1 has been shown to contribute to cancer progression through cytoplasmic functions and that NRAD1 is expressed in the cytoplasm of TNBC cells (Fig. 4B, C, and D), interrogation of the potential regulatory interactions of NRAD1 with miRNAs in TNBC is a relevant task that may provide novel insights into NRAD1-mediated gene regulation. To identify these regulatory interactions, NRAD1 was transiently inhibited in TNBC cells using LNA GapmeRs, after which RNA enriched for miRNAs was purified and subjected to miRNA microarray analysis to identify genome-wide miRNA expression changes in response to NRAD1 knockdown. Importantly, treatment of MDA-MB-468 and

SUM149 cells with anti-NRAD1 LNA GapmeRs significantly inhibits the expression of both the long (evidenced by Vidovic et al.[302]) and short NRAD1 transcripts (Fig. 5A, B, C, and D). Thus, it is unclear whether changes in NRAD1-mediated regulation of gene expression are due primarily to the short or long transcript.

5.3 lncRNA directed proteomics methods identify putative NRAD1-binding proteins

Several studies have previously shown that some oncogenic lncRNAs promote cancer progression through interactions with protein binding partners. For example, lncRNAs HOTAIR, BCAR4 and BORG interact with PRC2, PNUTS/SNIP1 and TRIM28 respectively, to promote breast cancer metastasis[272,291,292,334]. To begin to understand the mechanism through which NRAD1 regulates gene expression changes in the nucleus of TNBC cells, I sought to explore putative NRAD1-binding proteins to identify those that may contribute to NRAD1 activity.

I hypothesized that NRAD1-mediated gene regulation was achieved at least in part via NRAD1 interactions with regulatory binding proteins. Thus, to identify NRAD1-binding proteins, I performed two separate lncRNA-directed proteomics assays, RNA pulldown-MS[350] (Fig. 6) and ChIRP-MS[351] (Fig. 8). These two methods are similar in that they are both fundamentally designed to retrieve lncRNAs and their associated proteins, which are eluted from the lncRNA and subjected to MS, allowing the investigation of the lncRNA proteome. However, these methods differ in key areas and thus, the identification of lncRNA-associated proteins from both methods instills confidence in the validity of the lncRNA-protein interaction. The primary difference in these experiments is as follows. The lncRNA pulldown-MS experiment involved the *in vitro* synthesis and biotinylation of the long NRAD1 transcript from two different promoters (T7 or Sp6), which generated two different biotinylated NRAD1 transcripts of the same sequence. These transcripts were separately incubated in MDA-MB-468 cell nuclear lysates and retrieved with streptavidin for subsequent MS analyses (Fig. 6A and B). For ChIRP-MS, MDA-MB-468 cells were treated with formaldehyde in culture to cross-link molecular complexes that may involve lncRNAs, proteins, and chromatin. After sonication and cell lysis, antisense DNA oligo probes with

biotin modification designed to span the long NRAD1 transcript were hybridized to the native NRAD1 transcript, which was pulled down with streptavidin (Fig. 8A). Of note, ChIRP-MS was performed using two different probesets (either odd or even numbered probes).

Following MS analyses, proteins identified from both the T7 and Sp6-transcribed NRAD1 transcripts (RNA pulldown-MS) and those identified from both odd and even probes (ChIRP-MS) would be of interest for further analyses. The NRAD1 pulldown experiment revealed 184 proteins in common between both T7 and Sp6-synthesized transcripts (Fig. 7, Appendix 1), while ChIRP-MS revealed only 22 proteins in common between odd and even probe-retrieved NRAD1 (Fig. 8D). Importantly, we are confident in the ability of ChIRP-MS to effectively retrieve NRAD1 based on our experiments demonstrating that ChIRP-MS specifically retrieves RNA and that NRAD1 was highly enriched over housekeeping genes after ChIRP-MS with both odd and even probesets (Fig. 8B and C).

While the use of two separate transcripts in each experiment may help reduce potential non-specific protein-lncRNA interactions, the identification of a protein by both methods could be more important given the uniqueness of the two experiments. A protein identified as a putative NRAD1 binding partner using *in vitro* synthesized NRAD1 and native NRAD1 transcripts is of interest for downstream analyses due to its identification in two separate experiments with unique experimental conditions (synthesized vs. native transcript), ultimately supporting the legitimacy of the interaction. Further, foundational differences in the two experiments indicate that proteins identified by only one ChIRP-MS probeset are still important targets for further analysis. To this end, the number and/or identities of proteins identified by MS depend on the abundance of the NRAD1 transcript. As NRAD1 transcripts were synthesized *in vitro* using synthetic expression vectors and RNA polymerases, it is probable that the abundance of synthesized NRAD1 transcripts greatly outnumbered the native transcripts present in cross-linked MDA-MB-468 cell lysates used in ChIRP-MS. Further, *in vitro* synthesized NRAD1 transcripts were incubated in concentrated MDA-MB-468 cell lysates, providing enhanced ability for NRAD1-protein interactions. LncRNAs are expressed at low levels in human cancer cells[264], thus complicating efforts to capture native lncRNA-protein complexes at a level that will retrieve sufficient protein content

for adequate MS quantification. Further, the probes used to retrieve NRAD1 in ChIRP-MS may add an additional layer of complexity. As ChIRP-MS was performed with either odd or even numbered probes (based on their position along the NRAD1 sequence), it is possible that differences in the probes' binding location along the NRAD1 sequence affects the proteins identified. Further, the odd and even probesets likely differ in their ability to effectively retrieve NRAD1 and associated molecular complexes. Accordingly, QPCR analysis of odd and even probe-bound RNA revealed that NRAD1 enrichment over reference genes B2M and ARF was higher when ChIRP-MS was performed using the odd probeset relative to the even probeset (Fig. 8C). Thus, proteins identified using only one probeset were still considered as possible proteins of interest.

Of note, ChIRP-MS was used primarily as a tool to identify a shortlist of proteins that may contribute to NRAD1 function, which would be analyzed in downstream experiments. Given that the primary goal of this work was not to characterize the entire NRAD1 proteome, low numbers of proteins identified by ChIRP-MS can be tolerated if these proteins are also detected via lncRNA pulldown. For researchers seeking to investigate a more complete portrait of the chromatin-associated proteins bound to a lncRNA using ChIRP-MS, the experiment may need to be completed multiple times and eluted proteins pooled for MS analyses. In addition, the number of cells harvested, crosslinked, lysed, and sonicated for subsequent ChIRP analyses will also need to be optimized for the lncRNA being studied in order to retrieve adequate protein content. Finally, pooling of probes (instead of separating probes) for each ChIRP-MS experiment is suggested in the protocol[351] and may eliminate differences in NRAD1 retrieval and protein detection between odd and even probes.

Given the information discussed, we selected proteins that were identified using both T7- and Sp6-transcribed NRAD1 from the RNA pulldown that were also identified in ChIRP-MS (odd or even probeset) (Fig. 9A and B). We were interested in proteins that were identified with both T7- and Sp6-synthesized transcripts because of the large number of proteins identified and the likelihood that proteins identified using both transcripts were more reliable NRAD1 partners based on the experimental conditions (differentially synthesized sequence-identical NRAD1 transcripts incubated in protein-

enriched cell lysates). Together, there were 20 proteins that fit these criteria (Fig. 9A). Of the 20 proteins

identified, 16 were probable contaminants (i.e., keratins and other skin-related proteins) (Fig. 9B). Thus,

four proteins were highlighted as potential targets: HIST1H1C, HIST1H1E, HRNR, and S100A8 and

were further investigated for their potential roles with NRAD1.

5.4 S100A8 is a promising NRAD1-associated protein with gene expression profiles similar to NRAD1 in patient breast tumors.

Co-expression models investigating lncRNA and mRNA expression using RNA-sequencing data

are often employed to assess lncRNA function in human cancers. For example, Guo and colleagues

(2013) developed a lncRNA-mRNA co-expression network to predict lncRNA functions on a large-scale

using gene expression and protein interaction data[367]. More recently, Guo et al. (2016) investigated the

expression patterns of lncRNAs and mRNAs and generated a lncRNA-mRNA regulatory network using

gene expression profiles of ovarian cancer patients within the TCGA[368]. They were specifically interested

in developing a lncRNA-mRNA co-expression network associated with the progression of malignant

ovarian cancer. Their analysis of tumors of different stages revealed differentially expressed lncRNAs and

protein-coding genes associated with malignant ovarian cancer progression, whose co-expression

relationship was used to develop a co-expression network which can be used in the identification of

prognostic signatures[368,369]. For my work, I sought to reveal similarities in mRNA co-expression profiles

among mRNAs co-expressed with NRAD1 mRNA and mRNAs co-expressed with the mRNAs encoding

each protein of interest in patient breast tumors.

To prioritize MS-identified proteins with possible roles in NRAD1 function for further analyses,

we first interrogated RNA-sequencing data from patient breast tumors available from cBioPortal.

Specifically, we extracted mRNA expression data (RNA-seq) from the TCGA (Cell, 2015) breast

invasive carcinoma dataset to investigate the extent to which genes whose mRNA expression is positively

or negatively correlated with NRAD1 mRNA expression were similarly correlated with the mRNA

expression of each of the MS-identified proteins (Fig. 10). The reasoning for this analysis was that

proteins that interact with NRAD1 to regulate gene expression may regulate at least some of the same genes as NRAD1 in patient tumors, evidenced by the correlations of co-expressed genes. The Spearman correlations derived from cBioPortal are generated by individually plotting the mRNA expression of each protein-coding gene (~20,200) against the mRNA expression of a gene of interest (i.e., NRAD1) in 816 patient breast tumors[293,294]. Thus, gene mRNAs whose expression is positively or negatively correlated with NRAD1 mRNA expression in patient tumors may be upregulated or downregulated by NRAD1 respectively.

MS-identified hits with an mRNA expression correlation profile similar to that of NRAD1 may contribute to its regulatory functions in breast cancer. To this end, our analysis revealed that, S100A8 exhibited an mRNA expression correlation profile highly similar to that of NRAD1, where heatmap clustering revealed that of the four proteins of interest, mRNAs correlated with S100A8 were most closely related to those correlated with NRAD1 in patient breast tumors (Fig. 10). This provided evidence to support the potential role of S100A8 in NRAD1-mediated gene regulation.

To further explore similarities in the mRNA expression correlation profiles between S100A8, HRNR, HIST1H1C, and HIST1H1E and those of NRAD1, we plotted all of the Spearman correlations of the genes co-expressed with each of the proteins of interest against the Spearman correlations of the genes co-expressed with NRAD1 and calculated the correlation coefficient and p-value of this relationship (Fig. 11A, B, C, and D). Intriguingly, this analysis showed a strong and significant positive correlation between NRAD1 and S100A8 (R = 0.76, p < 2.2e-16) (Fig. 11A), indicating a high degree of overlap in the mRNAs either positively or negatively correlated with NRAD1 and S100A8. The correlations of genes co-expressed with HRNR, HIST1H1C, or HIST1H1E were not well correlated with the correlations of genes-co expressed with NRAD1 (Fig. 11B, C, and D). These findings suggest that NRAD1 and S100A8 may commonly regulate gene expression in breast cancer, supporting the potential role of S100A8 in NRAD1-mediated gene regulation. Given these results, S100A8 was further studied for its putative functional role with NRAD1.

To begin to identify the extent of the possible common gene regulation between NRAD1 and

S100A8, we used transcriptomic methods including a microarray and the same RNA-sequencing data

from patient breast tumors available from cBioPortal. Specifically, MDA-MB-468 cells were treated with

an anti-NRAD1 GapmeR to inhibit NRAD1 expression and RNA was isolated for microarray analysis

(Affymetrix Human Gene 2.0 ST Array) to identify genome-wide gene expression changes in response to

NRAD1 knockdown (Fig. 12). Genes that were up- or downregulated following NRAD1 knockdown

were also investigated in patient breast tumors using cBioPortal. From the microarray, I identified a series

of genes that were up- or downregulated upon NRAD1 knockdown (at least 1.2-fold change between the

control GapmeR and GapmeR #4, p-value < 0.05) and assessed their Spearman correlations with NRAD1

and S100A8 in patient breast tumors using RNA-sequencing data accessed via cBioPortal (TCGA Breast

Invasive Carcinoma, Cell 2015 dataset) (Fig. 13). This analysis identified several genes that were

upregulated or downregulated by NRAD1 in the microarray and positively (upregulated in microarray) or

negatively (downregulated in microarray) correlated with both NRAD1 and S100A8 in patient tumors

(Fig. 13, Appendix 2). This suggests that S100A8 may regulate many of the same genes as NRAD1,

thereby suggesting that S100A8 contributes to NRAD1-mediated gene regulation.

5.5 S100A8 binds NRAD1 and contributes to NRAD1-medaited gene regulation in MDA-MB-468 cells

To confirm the results from the lncRNA pulldown and ChIRP-MS suggesting that NRAD1 binds

S100A8 in MDA-MB-468 cells, we performed nuclear RNA immunoprecipitation with cross-linking of

chromatin prepared from MDA-MB-468 cells (Fig. 14). NRAD1 RIP was performed in parallel with

NEAT1 RIP, which served as a positive control (Fig. 15A). NEAT1 RIP was performed using an

antibody specific to EZH2, a protein known to interact with NEAT1[359]. QPCR analysis revealed that

NEAT1 expression was dramatically elevated when RIP was performed with the EZH2 antibody relative

to RIP with the IgG control (Fig. 15A). Given that the expected result was achieved using the

NEAT1/EZH2 positive control, we were confident that the results observed using S100A8 to retrieve

NRAD1 were valid. Importantly, mirroring the results of the positive control experiment, NRAD1

expression was significantly higher when RIP was performed using the S100A8 antibody relative to the

IgG control (Fig. 15B). Notably, NRAD1 expression was more than 100-fold enriched (average of three technical replicates) in RIP samples using the S100A8 antibody relative to those performed using IgG (Fig. 15B). These results suggest that S100A8 binds NRAD1 in MDA-MB-468 cells and confirms the results from the NRAD1 pulldown and ChIRP-MS experiments (Fig. 9).

In comparison, Huarte and colleagues (2010) performed a lncRNA-directed pulldown followed by mass spectrometry and cross-linked RIP to confirm the interaction between lncRNA-p21 and HNRNPK, demonstrating approximately 30-fold enrichment of lncRNA-p21 using HNRNPK-specific antibodies relative to negative control antibodies[370]. They further showed that lncRNA-p21 mediates transcriptional repression through the physical interaction with HNRNPK. We therefore gain confidence in the reliability of our lncRNA pulldown, ChIRP-MS, and RIP experiments.

Given the results of our RIP experiment demonstrating the interaction between S100A8 and NRAD1, we sought to investigate the potential regulatory role of S100A8 with NRAD1 and ALDH1A3 in MDA-MB-468 cells. Vidovic and colleagues (2020) previously identified a series of genes co-regulated by both NRAD1 and ALDH1A3 in MDA-MB-468 cells[302] (Fig. 16C). We investigated the potential role of S100A8 as a co-regulator of genes regulated by NRAD1 and ALDH1A3 by knocking down its expression using shRNAs and performing QPCR on the NRAD1 and ALDH1A3-regulated genes in MDA-MB-468 cells (Fig. 17 and 18). S100A8 knockdown was confirmed using QPCR and Western blotting, where of the two separate shRNAs used, only one significantly knocked down S100A8 at the gene and protein levels (Fig. 17A and B). Therefore, gene expression experiments were performed using only one stable S100A8 knockdown cell line. Strikingly, QPCR analysis of NRAD1 and ALDH1A3-upregulated genes revealed that the expression of all genes were downregulated upon S100A8 knockdown (6 of 7 significantly downregulated) (Fig. 18A). Similarly, 10 of 11 NRAD1 and ALDH1A3 downregulated genes were upregulated upon S100A8 knockdown (9 of 11 significantly upregulated) (Fig. 18B). This data suggests that most genes co-regulated by NRAD1 and ALDH1A3 in MDA-MB-468 cells are also regulated by S100A8, providing a functional role for S100A8 in NRAD1 and ALDH1A3-mediated gene regulation and supporting the hypothesis that NRAD1 regulates gene expression via

interactions with the protein partner S100A8 in MDA-MB-468 cells.

Notably, these results will need to be replicated using a second shRNA demonstrating significant knockdown at the gene and protein levels in MDA-MB-468 cells. Further, similar microarray analyses investigating changes in genome-wide gene expression in response to NRAD1 and ALDH1A3 knockdown should be performed in an additional TNBC cell line (i.e., SUM149) to identify NRAD1 and ALDH1A3 co-regulated genes, whose expression would be investigated via QPCR in SUM149 S100A8 knockdown cells. Following these experiments, microarray analysis of MDA-MB-468 and SUM149 cells with stable S100A8 knockdown vs. control cells will unveil genome-wide gene expression changes induced by knockdown of S100A8, which can be compared to those of NRAD1 and ALDH1A3, revealing the extent of NRAD1, ALDH1A3, and S100A8 co-regulated gene expression.   While ALDH1A3, NRAD1, and S100A8 were shown to co-regulate gene expression in MDA-MB-468 cells, the degree to which these changes are linked to cancer stemness and/or cancer progression remains unclear. Notably, gene ontology (GO) analyses on NRAD1-regulated genes revealed that NRAD1 downregulates genes involved in developmental and differentiation processes, and upregulates genes involved in alpha amino acid metabolism and lipid metabolism[302]. Intriguingly, these processes are known to play roles in CSC maintenance[108]. For example, alpha-amino acid metabolism has been shown to contribute to CSC activity in pancreatic cancer[371]. Li and colleagues (2017) showed that blocking the non-canonical glutamine metabolism pathway hindered pancreatic cancer stem cell (PCSC) characteristics (CSC self-renewal and expression of stemness-associated genes) and enhanced the effects of ionizing radiation treatment among PCSCs *in vitro* and *in vivo*[371]. Lipid metabolism also plays a role in CSC activity. Specifically, inhibition of fatty acid oxidation has been shown to reduce mammosphere formation of breast cancer cells[372]. Thus, NRAD1-mediated changes in gene expression may indeed contribute to CSC activity in TNBC. In support of these processes, NRAD1 knockdown in TNBC cells led to decreased cell viability and reduced mammosphere-forming potential[302].

Of the ALDH1A3/NRAD1/S100A8-upregulated genes (Fig. 18A), ALDH2, a member of the aldehyde dehydrogenase family has been shown to upregulate stemness factors, NANOG, SOX2, and KLF4 in embryonic stem cells[373]. Further, Chen et al. (2019) showed that FOXM1 silencing inhibited NANOG, SOX2, and OCT4 expression in liver cancer stem cell (LCSC) populations by decreasing ALDH2 expression[374]. Importantly, FOXM1 silencing and associated decrease of ALDH2 reduced tumorigenesis of LCSCs *in vivo*. Another ALDH1A3/NRAD1/S100A8-upregulated gene, PDZK1IP1, interacts with and inhibits NUMB, a negative regulator of the Notch pathway, leading to Notch pathway activation and cancer stemness in different human cell lines and PDX models[375]. Finally, Zheng et al. (2019) showed that the transcription factor SIX2 binds the promoter of CYP4Z1 and CYP4Z2P, promoting the progression of a ceRNA network mediated by CYP4Z1 and CYP4ZP2, leading to stemness and chemoresistance of breast cancer cells[376].

Thus, ALDH1A3, NRAD1, and S100A8 significantly upregulate genes involved in CSC activity, providing evidence that NRAD1-mediated gene regulation does contribute to its roles in CSC functions and breast cancer progression, likely through its induction by ALDH1A3 and its interaction with S100A8.

Given the result that S100A8 may contribute to NRAD1 and ALDH1A3-mediated gene regulation in MDA-MB-468 cells, we sought to uncover the extent of common gene regulation between ALDH1A3 and S100A8 (Fig. 19B). Our reasoning for this was that ALDH1A3 induces NRAD1, where NRAD1 contributes to ALDH1A3-mediated changes in gene expression[302]. We therefore expected the correlation of the correlations of genes co-expressed with ALDH1A3 vs. NRAD1 to be positive, which is indeed what we uncovered, a strong positive correlation (R = 0.81, p < 2.2e-16) showing that genes co-expressed with NRAD1 are also co-expressed with ALDH1A3 in patient breast tumors (Fig. 19A). This evidence points to co-regulation of gene expression among NRAD1 and ALDH1A3 and supports the previous findings[302]. Given this result, we expected the same analysis between ALDH1A3 and S100A8 to also reveal a positive correlation if S100A8 contributes to NRAD1 and ALDH1A3-mediated gene regulation. Indeed, this analysis revealed a strong positive correlation between S100A8 and ALDH1A3 (R = 0.61, p < 2.2e-16) (Fig. 19B), demonstrating that S100A8 and ALDH1A3 commonly regulate some

genes in patient breast tumors, further supporting the role of S100A8 in NRAD1 and ALDH1A3-mediated gene regulation.

5.6 S100A8 is an important contributor to breast cancer progression.

S100A8 (MRP8 or Calgranulin A) is a calcium-binding protein belonging to the S100 family of proteins that contributes to modulation of the inflammatory response in numerous human pathologies[377,378]. In addition, S100A8 is implicated in the development and progression of several human cancers, including breast cancer[379–384]. Importantly, S100A8 most often forms a heterodimer with the closely related protein, S100A9 (MRP14 or Calgranulin B). This S100A8/A9 complex, known as calprotectin, is released from neutrophils and monocytes and regulates inflammation via promotion of leukocyte recruitment and cytokine secretion[378].

S100A8/A9 is known to play a critical role in breast cancer progression where it has been shown to bind the receptor for advanced glycation end products (RAGE) to promote invasion of breast cancer cells[385]. Moreover, S100A8/A9 binding to RAGE induced EMT *in vitro* and promoted lung metastasis *in vivo*. Thus, the interaction of S100A8/A9 with RAGE plays a critical role in breast cancer invasion and metastasis and may represent a therapeutic target. Similarly, Moon and colleagues (2008) showed that S100A8/A9 contribute to H-Ras-mediated breast cancer cell invasion and migration[386]. More recently, S100A8/A9 was shown to increase breast cancer cell growth and to promote metastasis via binding the melanoma cell adhesion molecule (MCAM) cell surface receptor[387]. Importantly, S100A8/A9 binding to MCAM induced the activation of ETS translocation variant 4 (ETV4), promoting EMT and metastasis *in vivo* via upregulation of the ZEB1 transcription factor. Based on the evidence from the literature demonstrating the role of S100A8/A9 in breast cancer cell invasion and metastasis, experiments investigating the possible contribution of NRAD1 to S100A8/A9-mediated invasion/metastasis are warranted, as NRAD1 has not yet been associated with cancer cell invasion, migration, or metastasis in breast cancer.

Importantly, some literature suggests clinical implications for elevated S100A8 expression in

breast cancers[384,388]. Wang and colleagues (2018) assessed breast cancer patient data and showed that elevated S100A8 expression was found in patients with disease relapse and was associated with significantly lower overall and disease-free survival[384]. Further, they showed that TNBC patients had significantly higher expression of S100A8 relative to patients with other breast cancer subtypes. S100A8 may therefore represent a marker for relapse among TNBC patients. Miller et al. (2017) used patient breast tumor samples to determine that elevated S100A8 expression in breast tumor cells and stroma is a significant indicator of poor clinical outcome among early-stage patients[388]. Together, in addition to its functional role in breast cancer progression, S100A8 may represent a clinically valuable biomarker for breast cancer. In addition, Bao et al. (2016) showed that S100A8/A9 expression is elevated in basal-like and HER2+ breast cancers relative to other subtypes[389]. These results support our analysis of S100A8 and S100A9 expression in TNBC and Basal-like breast cancer using breast tumor RNA-seq data from TCGA (Fig. 24A, B, C, and D). It also supports our finding that S100A8 binds TNBC/Basal-like breast cancer-enriched lncRNA NRAD1 in MDA-MB-468 cells and contributes to NRAD1-mediated gene regulation (Fig. 15B, 18A and B). Bao and colleagues further provided evidence supporting the association of S100A8/A9 expression and secretion with loss of the estrogen receptor in breast cancer, which may contribute to the poor prognosis of Basal-like subtypes[389]. Specifically, ESR1 (estrogen receptor 1) and GATA3 (GATA binding protein 3) expression were negatively correlated with S100A8/A9 expression in patient breast tumors, evidenced via microarray analysis. ESR1 encodes estrogen receptor alpha (ERα), the major ER isoform in breast tissue, while GATA3 expression is positively correlated with ERα expression[390]. To this end, GATA3 regulates ERα expression via binding the ERα promoter[390]. Moreover, elevated GATA3 expression has been shown to suppress breast cancer metastasis by promoting mesenchymal-epithelial transition (MET) of invasive breast cancer cells[391]. Intriguingly, in our analyses of genes regulated by NRAD1 that were similarly co-expressed with NRAD1, ALDH1A3, S100A8, and S100A9 in patient breast tumors (Fig. 22), GATA3 was found to be downregulated by NRAD1 and negatively correlated with each of the four mRNAs. This may be consistent with an S100A8/A9-mediated loss of the estrogen receptor in association with NRAD1.

In line with my hypothesis that NRAD1 interacts with S100A8 to regulate gene expression via chromatin interactions, some papers have demonstrated that S100A8 and S100A9 can regulate gene expression via chromatin binding, providing further support for their potential roles in NRAD1-mediated gene regulation. Specifically, Schonthaler et al. (2013) showed that S100A8 and S100A9 regulate transcription of complement factor C3 via chromatin binding and remodelling at the C3 promoter in psoriasis[392]. They further demonstrated that S100A8/A9 can function in the nucleus, interacting with histones and nucleosomes, potentially contributing to epigenetic alteration of gene expression. Most recently, Song and colleagues (2021) showed nuclear S100A8/A9 plays an important role in breast cell transformation[393]. Specifically, nuclear S100A8 and S100A9 were separately shown to interact with chromatin at a combined ~30,000 target sites (R=0.88 between S100A8 and S100A9 binding levels), where binding levels at almost all sites were increased following transformation. Importantly, S100A8/A9 chromatin binding was associated with oncogenic transcriptional activation. Together, these results demonstrate that S100A8 can interact with chromatin and regulate gene expression, supporting its potential role with NRAD1.

Inhibition of S100A8/A9 has been investigated in the treatment of numerous inflammatory conditions and may therefore be of interest as a therapeutic target in human cancers. For example, inhibition of S100A8/A9 binding to RAGE using paquinimod (an S100A9 inhibitor) reduced diabetes-induced thrombocytosis (elevated platelet count) and atherogenesis (formation of plaques in the arteries) in diabetic mice[394]. Tasquinimod binds S100A9 and/or S100A8/A9, blocking S100A9 interactions with RAGE or TLR4 (Toll-like receptor 4), another receptor well-studied for its interactions with S100A8/A9, inhibiting the release of TNF-α (Tumor necrosis factor alpha) in *in vivo* models of autoimmune disorders[395].

Importantly, the use of Tasquinimod in human cancers, particularly prostate cancer, has demonstrated success in pre-clinical and clinical studies. For example, Jennbacken et al. (2011) showed that Tasquinimod inhibited the formation of lung and lymph node metastases in xenograft models of castration-resistant prostate cancer (CRPC) tumors by inhibiting tumor establishment[396]. Tasquinimod has

also been shown to enhance cancer immunotherapies in murine models. Specifically, Tasquinimod enhanced the antitumor effects of two different immunotherapies in mouse models of cancer: a tumor vaccine for prostate cancer and a tumor-targeted antigen for melanoma[397]. Tasquinimod was also shown to significantly reduce disease progression and improve progression-free survival (PFS) in patients with metastatic CRPC in a phase II clinical trial[398]. Further, in a phase III trial, Tasquinimod improved radiographic progression-free survival[399]. Intriguingly, Tasquinimod has been shown to target histone deacetylase HDAC4, inhibiting histone deacetylation, cell survival, and angiogenesis[400]. Importantly, Tasquinimod demonstrated effectiveness as a single agent against human breast tumor xenografts[400]. HDAC inhibitors have demonstrated encouraging results for TNBC treatment[401]. For example, the HDAC inhibitor, Entinostat has been shown to reduce the percentage of CD44^{high}/CD24^{low} CSCs and ALDH activity within TNBC cell populations[402]. Further, Entinostat could inhibit primary tumor formation and lung metastasis *in vivo*[402]. Thus, given its anti-HDAC activity, Tasquinimod may also be effective against breast CSCs.

The role of S100A8 and S100A9 in CSCs is not well-characterized, with a limited number of studies investigating their role in CSC activity. To this end, S100A9 expression was found to be significantly higher in glioma stem cells relative to differentiated cells, where S100A9 knockdown reduced both glioma stem cell proliferation and *in vivo* tumor growth, marking S100A9 as a potentially important player in CSC activity and an attractive therapeutic target[403]. In support of this, Wei and colleagues (2020) recently showed that tumor-associated macrophages (TAMs) could upregulate S100A9 expression, which enhanced the stemness characteristics of hepatocellular carcinoma (HCC) cells via activation of the RAGE/NF-κB axis[404]. While S100A8 has not been directly linked to CSC activity, its complexing with S100A9 certainly implies a role for S100A8 in CSC function which warrants further investigation. Some evidence for a role for S100A8 in CSC activity can be demonstrated by its contribution to Wnt/β-catenin signaling, a pathway well-characterized for its involvement in CSC function. To this end, S100A8 and S100A9 were associated with colorectal cancer progression via contributions to cell survival and migration through their activation of the Wnt/β-catenin pathway[405].

Together, S100A8 and S100A9 are critically involved in cancer progression and may regulate aspects of CSC activity. Further experiments are needed to elucidate the extent to which S100A8 and S100A9 contribute to breast CSC functions.

5.7 Investigating S100A9 and ALDH1A3 as key players in NRAD1/S100A8-mediated gene regulation.

Due to the well-characterized close association between S100A9 and S100A8 and our identification of S100A9 via NRAD1 pulldown, we reasoned that S100A9 would contribute to the regulatory interactions between ALDH1A3, NRAD1, and S100A8 in breast cancer. Given that S100A8 similarly regulates most genes regulated by both ALDH1A3 and NRAD1 (Fig. 18), we further investigated the role of ALDH1A3 with NRAD1 and S100A8 and its possible association with S100A9. Our analyses of patient breast tumor RNA-seq data revealed strong similarities among the correlations of the genes co-expressed with ALDH1A3, NRAD1, S100A8, and S100A9, where the co-expression correlation profile of ALDH1A3 most closely resembled that of NRAD1 and the co-expression correlation profiles of S100A8 and S100A9 were almost identical (Fig. 21). This result supports the previous finding that ALDH1A3 induces NRAD1, which contributes to ALDH1A3-mediated gene regulation and the known association of S100A8 and S100A9 in breast cancer[302,385]. Supporting the co-expression correlation profiles, the Spearman correlations of the genes co-expressed with S100A8 were almost perfectly correlated with those co-expressed with S100A9 (R = 0.99, p < 2.2e-16) (Fig. 20A), which supports our reasoning that the correlation of the Spearman correlations of genes co-expressed with two different mRNAs of interest would inform us about their potential common gene regulation. Further, the Spearman correlations of the genes co-expressed with NRAD1 or ALDH1A3 were strongly correlated with those co-expressed with S100A9, where the correlations (NRAD1 vs. S100A9 and ALDH1A3 vs. S100A9) were higher than those of NRAD1 vs. S100A8 and ALDH1A3 vs. S100A8 (Fig. 20B and C). This data suggests that the activity of S100A8 occurs with S100A9. In support of this, Ertem and colleagues (2016) identified S100A9, ALDH1A3, MMP7, and NPPB as critical oncogenes downregulated in colon tumors following treatment with tolfenamic acid (Clotam)[406].

Finally, in support of the potential co-regulatory roles of ALDH1A3, NRAD1, S100A8, and S100A9 in breast cancer, we identified 1080 genes upregulated by NRAD1 and positively correlated- or downregulated by NRAD1 and negatively correlated with each of ALDH1A3, NRAD1, S100A8, and S100A9 in patient breast tumors (Fig. 22, Appendix 3). This data provides further evidence that ALDH1A3, NRAD1, S100A8, and S100A9 co-regulate gene expression in breast cancer and supports the role of S100A8/A9 in NRAD1-mediated gene regulation.

To provide further support for the roles of S100A8 and S100A9 in NRAD1-mediated gene regulation, we assessed their expression in TNBC/basal-like patient tumors using RNA-seq data from the TCGA (Fig. 24A, B, C, and D). ALDH1A3 is a functional breast CSC marker enriched in TNBC/Basal-like breast cancers that induces breast CSC-associated lncRNA NRAD1, which was previously shown to be enriched in TNBC/basal-like breast cancers and Aldefluor[high] breast CSC populations[117,302]. We first confirmed these findings using patient tumor RNA-seq data from the TGCA (Cell, 2015) dataset. NRAD1 and ALDH1A3 expression was found to be significantly higher in both basal-like and TNBC patient tumors relative to non-basal or non-TNBC tumors, confirming the results of previous studies (Fig. 23A, B, C, and D). If S100A8 and S100A9 contribute to NRAD1-mediated gene regulation, they should be similarly enriched in TNBC/basal-like breast cancers. Akin to ALDH1A3 and NRAD1, S100A8 and S100A9 expression was significantly higher in basal-like patient tumors relative to non-basal tumors (Fig. 24B and D). S100A8 and S100A9 expression was elevated in TNBC tumors relative to non-TNBC tumors, although not significantly (Fig. 24A and C). The elevated expression of S100A8 and S100A9 in basal-like and TNBC patient tumors therefore supports their roles with NRAD1.

Given that NRAD1, S100A8, and S100A9 expression is significantly higher in basal-like patient tumors relative to non-basal tumors, we performed the co-expression Spearman correlation analysis to assess the correlation of Spearman correlations of genes co-expressed with NRAD1 vs. those co-expressed with S100A8 or S100A9 in basal-like patient tumors (Fig. 25A and B). This analysis revealed positive correlations between NRAD1 vs. S100A8 (Fig. 25A) and NRAD1 vs. S100A9 (Fig. 25B), demonstrating that the positive correlations observed previously in patient tumors of all subtypes were not

due entirely to non-basal subtypes. This data also supports our findings showing the regulatory association of S100A8 with NRAD1 in MDA-MB-468 cells and supports the likely role of S100A8/A9 in NRAD1-mediated gene regulation in TNBC/basal-like breast cancers. Further experiments are needed to validate the role of S100A9 in NRAD1-mediated gene regulation in breast cancer.

5.8 S100A8 and S100A9 may contribute to NRAD1 activity in ovarian cancer

Some studies have shown that NRAD1 contributes to the progression of ovarian cancer[339,340]. We therefore wondered whether S100A8/A9 could contribute to NRAD1-mediated oncogenic activity through regulation of gene expression, as we have shown in breast cancer. We first wondered whether NRAD1 expression was higher in ovarian cancer patient tumors relative to normal ovarian tissue to determine the potential targetability of NRAD1 in ovarian tumors. To this end, I investigated NRAD1 expression using microarray data from normal and ovarian tumor tissues accessible from NCBI's gene expression omnibus (GEO) datasets platform (GSE38666)[352]. These analyses showed that NRAD1 expression was significantly higher in ovarian tumor tissue relative to normal ovarian tissue (Fig. 26A and B). Similarly, previous studies have shown that S100A8 expression is also higher in ovarian cancer tissues relative to non-cancerous tissues[362]. Upon plotting the Spearman correlations of the genes co-expressed with S100A8 or S100A9 against those co-expressed with NRAD1 in ovarian cancer patient tumors using RNA-seq data extracted from the TCGA, we showed a strong positive correlation between the Spearman correlations of genes co-expressed with NRAD1 and those co-expressed with S100A8 and S100A9, supporting the possible role of S100A8/A9 in NRAD1-mediated gene regulation in ovarian cancer (Fig. 27A and B). We gain confidence in the importance of this interaction due to the evidence to support its existence in two different human cancers. Given this data, further experiments should be conducted to validate the possible association of S100A8/A9 with NRAD1 in ovarian cancer and to link this axis to ovarian cancer progression.

5.9 Limitations (Data Chapter 1)

Here, we have shown that S100A8 interacts with breast CSC-associated lncRNA, NRAD1 and contributes to ALDH1A3 and NRAD1-induced gene regulation in breast cancer cell lines. However, this regulation of gene expression has not yet been linked to breast cancer stemness and/or breast cancer progression. Thus, to provide a functional role for S100A8 in NRAD1-mediated gene regulation and cancer progression, experiments linking ALDH1A3/NRAD1/S100A8-regulated genes to oncogenic activity in breast cancer will be paramount.

Our NRAD1 pulldown and ChIRP-MS experiments were both performed in MDA-MB-468 cells only. Thus, performing these experiments once more in an additional TNBC cell line such as SUM149 may strengthen the reliability of our data. Further, given that the ChIRP-MS experiment identified only a modest number of proteins relative to the NRAD1 pulldown, we were limited by the number of proteins that were detected using both methods which affected the proteins that were selected for downstream analyses. To rectify this, NRAD1 ChIRP-MS could be optimized and performed again in MDA-MB-468 cells which may increase the number of NRAD1-associated proteins identified, thereby, increasing the number of unique proteins to validate in downstream experiments.

With respect to our S100A8 knockdown experiments, stable S100A8 knockdown was achieved using only one cell line and one shRNA. Thus, to improve the reliability of our gene expression data, a second S100A8 knockdown will need to be generated in MDA-MB-468 cells, and S100A8 knockdown will need to be performed using the same two shRNAs in SUM149 cells. QPCR experiments investigating the expression of ALDH1A3 and NRAD1-regulated genes in the context of S100A8 knockdown should be performed in MDA-MB-468 cells using the second shRNA knockdown and in SUM149cells using both knockdowns.

Finally, our data and the evidence from the literature points to the importance of the S100A8/A9 complex in breast cancer progression, although we do not yet have *in vitro* data to support the interaction of S100A9 with NRAD1. Experiments investigating the physical interaction of S100A9 and NRAD1 and the possible role of S100A9 in NRAD1-mediated gene regulation should be performed.

5.10 Future directions (Data Chapter 1)

The oncogenic activity mediated by the functional CSC marker ALDH1A3 is likely caused by gene expression changes induced by downstream effectors such as NRAD1[302]. ALDH1A3 critically regulates some normal adult physiological functions such as vision, learning and memory, immune function, adult neurogenesis, and epithelial function and differentiation[407,408] and thus, targeting downstream, cancer-specific effectors such as NRAD1 may represent a superior alternative to the direct targeting of ALDH1A3. The TNBC and CSC-enriched lncRNA NRAD1 is associated with poor patient outcomes and contributes to ALDH1A3-mediated gene expression in TNBC cell-lines[302]. NRAD1 inhibition has been shown to reduce cell survival, mammosphere formation, and *in vivo* tumor growth, demonstrating its promise as a possible therapeutic target. We have shown here that ALDH1A3 and NRAD1-mediated changes in gene expression may occur through the physical association of NRAD1 with the oncogenic protein S100A8. However, further work is needed to discern how S100A8 contributes to NRAD1-mediated oncogenic activity in breast cancer.

First, given the evidence from the literature and our analyses showing the close association of S100A8 and S100A9 with NRAD1 in breast cancer, the role of S100A9 in NRAD1-mediated gene regulation should be investigated. To this end, the result from our NRAD1 pulldown demonstrating S100A9 binding to NRAD1 using two uniquely synthesized transcripts should be validated using nuclear RIP with cross-linking coupled with QPCR, as was performed to validate the interaction of NRAD1 and S100A8. Further, S100A9 should be knocked down using two separate shRNAs in two TNBC cell lines (MDA-MB-468 and SUM149), where the expression of ALDH1A3 and NRAD1-regulated genes should be subsequently explored via QPCR in the context of S100A9 knockdown. Similarities in gene regulation between S100A8 and S100A9 would support their formation of a complex associated with NRAD1. Only S100A9 has been linked to CSC activity and our co-expression analyses revealed that the correlations of genes co-expressed with S100A9 exhibit a stronger correlation with those co-expressed with NRAD1 relative to S100A8 vs. NRAD1. Thus, it is possible that S100A9 is more important than S100A8 in NRAD1-mediated gene regulation and cancer progression. Our gene expression analyses in the context of

S100A9 knockdown may therefore show an enhanced effect in ALDH1A3/NRAD1-mediated gene regulation relative to S100A8.

S100A8 and S100A9 play important roles in invasion and metastasis in breast cancer[385]. Importantly, the role of NRAD1 in invasion and metastasis has not yet been explored in breast cancer and merits investigation. ALDH1A3 is known to promote metastasis[117,409], and the Marcato lab has shown that ALDH1A3 regulates genes in the plasminogen activation pathway, increasing the synthesis of ECM-degrading protein plasmin from plasminogen, contributing to invasion and possibly ALDH1A3-mediated metastasis (Marcato et al., unpublished). Thus, ALDH1A3-induced metastasis may occur through downstream effectors such as NRAD1 and its interacting proteins S100A8/A9.

While we have shown that S100A8 co-regulates gene expression with NRAD1 and ALDH1A3 in MDA-MB-468 cells, it remains unknown whether these changes in gene expression contribute to breast cancer progression or stemness. Thus, knockdown of ALDH1A3/NRAD1/S100A8-upregulated genes followed by experiments exploring the effect of gene knockdown on cell proliferation, survival, tumor growth, and/or cancer stemness would reveal whether these co-regulated genes mediate breast cancer progression. Given the known functions of ALDH2, PDZK1IP1, and CYP4Z1 in CSC activity, these genes would represent attractive starting points[373–376].

The putative roles of S100A8 and S100A9 in promoting breast CSC activity have not yet been investigated. To assess whether S100A8/A9 contribute to NRAD1-mediated breast cancer stemness, S100A8 and S100A9 should be investigated using the experiments performed to identify NRAD1 as a CSC-associated lncRNA[302]. To this end, the Aldefluor assay should be used to identify CSC populations among TNBC cell lines, upon which the expression of S100A8 and S100A9 should be assessed via QPCR of Aldefluor[High] vs. Aldefluor[Low] populations. Assuming positive results, S100A8 and S100A9 knockdown cells should be assessed for their mammosphere forming potential in TNBC cell lines. Given our confirmation of S100A8 and NRAD1 binding, results from these experiments similarly demonstrating S100A8/A9 contribution to CSC activity hints at the importance of the S100A8/NRAD1 association in the promotion of cancer stemness.

A limiting dilution assay, the premier experiment to determine the frequency of CSCs in solid tumors, should be performed with S100A8 and S100A9 knockdown in TNBC cell lines to investigate the potential importance of S100A8/A9 in tumor initiation, indicative of their roles in CSC activity[122]. To this end, TNBC cells with S100A8 or S100A9 knockdown, or control cells should be injected into NOD/SCID mice at decreasing concentrations to calculate the minimum number of cells required to form tumors. If S100A8 or S100A9 knockdown cells require more cells to initiate tumors relative to control cells, it is likely that S100A8/A9 is necessary for tumor-initiating/CSC activity. Further, given the availability of S100A9 inhibitors such as Tasquinimod, *in vivo* experiments investigating this drug as a single agent and in combination with chemotherapies in breast cancer may yield promising results.

Finally, to explore the dependency upon S100A8/A9 for NRAD1 chromatin binding, NRAD1 ChIRP-seq should be performed using S100A8 or S100A9 knockdown and control cells[356]. ChIRP-seq is used to elucidate specific genomic regions bound by an RNA of interest. Thus, in the S100A8/A9 knockdown background, reduced NRAD1 binding to NRAD1-bound and regulated genes would support the role of S100A8/A9 in NRAD1 chromatin binding. ChIP-seq is used to identify the genome-wide binding sites of proteins[410]. Thus, S100A8/A9 ChIP-seq using cells with or without NRAD1 knockdown would determine the importance of NRAD1 in S100A8/A9 DNA-binding. Together, these experiments will determine the co-dependency of S100A8/A9 and NRAD1 for chromatin binding and gene regulation in breast cancer.

5.11 NRAD1 regulates miRNA expression in MDA-MB-468 and SUM149 cells (Data Chapter 2).

LncRNAs are known to modulate gene expression by performing regulatory roles aligning with one of several functional archetypes including transcriptional enhancers or guides, transcription factor decoys, molecular scaffolds, and competitive endogenous RNAs (ceRNAs) that sponge miRNAs, proteins, and other molecules[228].

Significant light has been recently cast on the interactions of lncRNAs with sequence complimentary miRNAs in cancer, where most of the literature is focused on the roles of oncogenic lncRNAs as miRNA sponges that sequester miRNAs with tumor suppressive functions. In addition, to their roles as miRNA sponges, lncRNAs can also act as miRNA precursors that regulate miRNA biogenesis, or as competitors against miRNAs for binding sites on target mRNAs[239]. Accordingly, the lncRNA LOC554202 is the host gene for the breast cancer metastasis-suppressing miRNA, miR-31 and regulates its transcriptional activity[286]. LOC554202 and miR-31 experience promoter hypermethylation in TNBC cell lines, downregulating their expression and contributing to invasion and metastasis. MiRNAs can also regulate lncRNAs by binding to them, resulting in lncRNA degradation by the RISC (RNA-induced silencing complex) complex. For example, miR-21 negatively regulates the tumor suppressor lncRNA, GAS5 through binding a specific site on exon 4 of GAS5[411]. Similar interactions have also been shown with MALAT1/miR-1[277] and XIST/miR-7[412]. While these non-sponging lncRNA-miRNA interactions have been demonstrated in some cancers, the further investigation of these mechanisms in human cancers is warranted. Taken together, lncRNA-miRNA axes critically regulate gene expression in human cancers and are potential therapeutic targets[366]. We therefore sought to explore the role of NRAD1 in regulation of miRNA expression in TNBC cell lines to provide support for a novel mechanism through which NRAD1 may regulate gene expression.

ChIRP-seq experiments previously revealed that NRAD1 has enriched chromatin interactions among the genes it regulates; however, more than 40% of NRAD1-regulated genes are not bound by NRAD1 and thus, alternative regulatory mechanisms may exist[302]. Given that NRAD1 is expressed in both the cytoplasm and the nucleus (Fig. 4), it is possible that NRAD1 functions as a ceRNA in the

cytoplasm to regulate gene expression. To investigate the possible alternative mechanisms through which NRAD1 regulates gene expression in breast cancer, I studied the interactions between NRAD1 and miRNAs in two separate TNBC cell lines (MDA-MB-468 and SUM149) by transiently inhibiting NRAD1 expression and assessing changes in miRNA expression in response to NRAD1 inhibition. To accomplish this, we used the Affymetrix GeneChip miRNA 4.0 Array, allowing us to quantify changes in miRNA expression (control vs. knockdown) following NRAD1 knockdown. These analyses revealed that NRAD1 knockdown indeed regulates miRNA expression in MDA-MB-468 and SUM149 cells (Fig. 28A, B, C, and D). To this end, NRAD1 knockdown with GapmeR #3 and NRAD1 knockdown with GapmeR #4 led to the upregulation and downregulation of miRNAs in both cell lines (Fig. 28A, B, C, and D, Table 6). Furthermore, GapmeR #3 and GapmeR #4 induced similar changes in miRNA expression within cell lines and demonstrated some similarities across both cell lines, supporting the reliability of our data (Fig. 29).

To narrow the list of NRAD1-regulated miRNAs to identify key miRNAs with which NRAD1 may interact to regulate gene expression in TNBC, we identified the miRNAs up- or downregulated by NRAD1 knockdown with both GapmeRs in both cell lines (FC $\geq$ 1.3 or $\leq$ -1.3, p < 0.05) (Fig. 30 and 31). Downstream analyses should focus on these miRNAs, given the increased likelihood of their regulation by NRAD1. This led to the identification of two miRNAs, miR-4485-3p and miR-595, both of which were highly upregulated following NRAD1 knockdown in both cell lines, far surpassing our fold-expression change cutoff of $\geq$ 1.3 (Fig. 30). While our subsequent analyses will focus on NRAD1-regulated miRNAs similarly regulated by knockdown with both GapmeRs in both cell lines, the miRNAs commonly regulated by different GapmeRs within a cell line remain important targets for further analyses. Accordingly, we identified 13 and 20 miRNAs upregulated by both GapmeRs in MDA-MB-468 cells and SUM149 cells respectively (Fig. 30, Table 7 and Table 8). Our analysis of miRNAs downregulated by both GapmeRs in MDA-MB-468 and SUM149 cells identified miR-1226-3p and miR-4458, both of which also demonstrated elevated fold-expression changes relative to the control in both cell lines (Fig. 31). Similarly, 16 and 24 miRNAs were downregulated by treatment of MDA-MB-468

and SUM149 cells with both GapmeRs respectively, illustrating cell line-specific changes in miRNA expression which warrant further analyses (Fig. 31, Table 9 and Table 10).

5.12 Investigating miR-595 and miR-4485-3p as putative targets for NRAD1-mediated sponging

The miRNA-sponging roles of lncRNAs are well-characterized in breast cancer, where miRNA-sequestering lncRNAs are known to play important roles in regulation of gene expression, promoting breast cancer progression. While the non-sponging interactions between lncRNAs and miRNAs are far-less studied, these mechanisms represent potential untapped sources of information with respect to the far-reaching diversity of regulatory lncRNA-miRNA interactions in breast cancer. Given that NRAD1 is known to act as a miRNA sponge in cancer and our identification of NRAD1 expression in the cytoplasmic fractions of MDA-MB-468 cells, suggesting its potential role as a miRNA sponge in breast cancer, we focused our analyses on the role of NRAD1 as a miRNA sponge. To this end, our downstream analyses explored the miRNAs upregulated following GapmeR 3 and 4 treatment of MDA-MB-468 and SUM149 cells (Fig. 30). The miRNAs upregulated in the absence of NRAD1 are downregulated when NRAD1 is expressed, leading to our hypothesis that NRAD1 downregulates the expression of these miRNAs by binding and sequestering them. Our subsequent exploration was therefore concentrated on miR-4485-3p and miR-595 as putatively NRAD1-sponged miRNAs (Fig. 30).

To provide further support for our selection of these miRNAs, we consulted the literature to uncover their functions in cancers. MiRNAs, sponged by oncogenic lncRNAs should have tumor suppressive functions if their sponging by a lncRNAs promotes oncogenic activity. Accordingly, ectopically expressed miR-4485-3p reduced tumorigenicity in MDA-MB-231 mouse xenograft models, illustrating its role as a tumor suppressor[413]. Further, miR-595 was also shown to act as a tumor suppressor in ovarian cancer. Specifically, miR-595 overexpression decreased cell proliferation, colony formation, invasion, and sensitized ovarian cancer cells to cisplatin[414]. Intriguingly, miR-595 was found to target the functional CSC marker ABCB1 in ovarian cancer[414]. Given that both miRNAs have been shown to act as tumor suppressors, we gained confidence in their potential interactions with NRAD1.

Generally, miRNAs interact with the 3′ untranslated region (3′ UTR) of target mRNAs to induce mRNA degradation and translational repression[415]. However, miRNAs may also bind the 5' UTR, promoter regions, or coding sequences[415]. MiRNAs bind to mRNAs and lncRNAs through a miRNA response element (MRE), a seed region in the mRNA or lncRNA sequence that usually binds nucleotides 2-8 of the 5' end of the miRNA with a high degree of complementarity[415]. With respect to lncRNA-mediated sponging of miRNAs, different sponging effects may be observed[416]. For example, miRNAs may bind lncRNAs with either complete or partial complementarity. In addition, lncRNAs may possess multiple MREs, allowing their association with different miRNAs[416].

To provide further support for the putative NRAD1-mediated sponging of miR-595 and/or miR-4485-3p, we explored the potential binding of miR-4485-3p and miR-595 to NRAD1 transcripts (Fig. 32A and B). To investigate this, we identified the short and long NRAD1 transcript sequences using the Ensembl database and identified the sequences of each miRNA using miRBase. Our analysis demonstrated support for the binding of both miRNAs to both NRAD1 transcripts, with evidence for the 5' miRNA seed region in both miRNAs (Fig. 32A and B). Importantly, miR-595 binding to NRAD1 transcripts appeared to be stronger, given its longer stretches of consecutively complimentary nucleotides at its 5' end (Fig. 32A). Intriguingly, miR-4485-3p binds the same sequence in both NRAD1 transcripts, which may indicate an increased likelihood for its true binding of NRAD1 (Fig. 32B). Together, while this analysis provides support for miRNA binding to NRAD1, the experimental validation of these interactions is required.

Given the evidence that miR-4485-3p and miR-595 may bind NRAD1, we explored the degree to which NRAD1 may sponge either of these miRNAs to regulate gene expression (Fig. 33). We first used TargetScan to identify the predicted mRNA targets of each miRNA. We then assessed changes in gene expression in response to NRAD1 knockdown in MDA-MB-468 cells using the Affymetrix Human Gene 2.0 ST array platform (control vs. GapmeR #4), identifying genes downregulated following NRAD1 knockdown (upregulated by NRAD1), applying a cutoff of ≤ -1.2, p < 0.05. Next, we identified the mRNAs that were predicted targets of miR-4485-3p or miR-595 and also upregulated by NRAD1 in the

microarray. miRNA targets that are upregulated by NRAD1 may suggest that NRAD1 upregulates these

mRNAs by sponging the miRNA that otherwise targets it. To this end, NRAD1 was shown to upregulate

53 mRNAs that are also predicted targets of miR-4485-3p (Fig. 33A) and 409 mRNAs that are also

predicted targets of miR-595 (Fig. 33B). These results support our hypothesis that NRAD1 sponges miR-

4485-3p and/or miR-595 to regulate gene expression in breast cancer. While the mRNA targets of miR-

595 showed increased overlap with NRAD1-upregulated genes relative to the mRNA targets of miR-

4485-3p, there are significantly more mRNAs predicted to be targeted by miR-595 vs. miR-4485-3p

(3535 vs. 526). When looking at the mRNA targets that overlap with NRAD1-upregulated mRNAs, the

percentages of predicted mRNA targets out of all mRNA targets are highly similar for both miRNAs

(10% of predicted miR-4485-3p-targeted mRNAs are also upregulated by NRAD1 and 11.6% of

predicted miR-595-targed mRNAs are also upregulated by NRAD1).

To narrow the list of mRNAs upregulated by NRAD1 that are predicted targets of miR-595 or

miR-4485-3p, we plotted the TargetScan binding score for the miRNA with its predicted mRNA target

versus the fold-expression change of the mRNA, identified via microarray in MDA-MB-468 cells with

NRAD1 knockdown (fold-change $\geq$ 1.2, p < 0.05) (Fig. 34 and 35). The TargetScan binding score

indicates the predicted efficacy of mRNA targeting by a miRNA, where higher binding scores correspond

to a higher likelihood of targeting. We used this analysis to identify mRNAs with both a high fold-

expression change following NRAD1 knockdown and a TargetScan binding score of >0.2. Through these

analyses, we identified 44 mRNAs that were upregulated by NRAD1 (fold-change $\geq$ 1.2, p < 0.05) and

were predicted targets of miR-595 (TargetScan binding score >0.2) (Fig. 34) and 39 mRNAs that were

upregulated by NRAD1 and were predicted targets of miR-4485-3p (Fig. 35). Following experiments that

confirm binding of miR-595 or miR-4485-3p to NRAD1, further analyses should interrogate the mRNAs

upregulated by NRAD1 (fold-change $\geq$ 1.2, p < 0.05) that are also predicted targets of the miRNA

(TargetScan binding score >0.2) to validate the interaction of the miRNA with the mRNA *in vitro*.

Validated targets may represent mRNAs regulated by NRAD1 through its sponging of the miRNA of

interest.

In addition to the miRNAs upregulated following NRAD1 knockdown by both GapmeRs in MDA-MB-468 and SUM149 cells, the miRNAs downregulated by treatment of both MDA-MB-468 and SUM149 cells with both GapmeRs (miR-1226-3p and miR-4458) represent important targets for further analyses (Fig. 31). While downregulation of these miRNAs upon NRAD1 knockdown does not suggest NRAD1-mediated regulation by sponging, NRAD1-mediated upregulation of these miRNAs may serve an important role in breast cancer. This may be particularly important regarding miR-4458, given its striking downregulation following NRAD1 knockdown by both GapmeRs in both cell lines (Fig. 31).

5.13 ALDH1A3 regulates miRNA expression in MDA-MB-468 and SUM149 cells.

NRAD1 is induced by ALDH1A3 and its product, retinoic acid in breast cancer[302]. Thus, we performed miRNA microarray analysis in the context of ALDH1A3 knockdown to investigate whether there is a further connection between ALDH1A3 and NRAD1 via common regulation of miRNA expression. Intriguingly, we showed here that ALDH1A3 regulates the expression of miRNAs in both MDA-MB-468 and SUM149 cells (Fig. 36A, B, C, and D, Table 11). Unlike NRAD1, we did not identify any miRNAs upregulated or downregulated following NRAD1 knockdown by both shRNAs in both cell lines that met our cutoff criteria (fold-change ≥ 1.3 or ≤ -1.3, $p < 0.05$) (Fig. 38A and B). However, several miRNAs were found to be significantly up- or downregulated by both shRNA-61 and shRNA-63 (fold-change ≥ 1.3 or ≤ -1.3, $p < 0.05$) within either MDA-MB-468 or SUM149 cells, which warrant further investigation (Fig. 38A and B, Tables 12, 13, 14, and 15).

5.14 Investigating NRAD1 and ALDH1A3 co-regulated miRNA expression in MDA-MB-468 and SUM149 cells.

Intriguingly, when we assessed similarities in miRNA expression following knockdown of NRAD1 (GapmeR #3 and GapmeR #4) and ALDH1A3 (shRNA-61 and shRNA-63) in parallel without a cutoff, we uncovered similarities in miRNA expression between knockdown of NRAD1 and ALDH1A3 within a cell line (Fig. 39). Specifically, the miRNA expression profile following knockdown of one gene

149

(i.e., NRAD1) was more closely related to the miRNA expression profile after knockdown of the other gene (i.e. ALDH1A3) within a cell line than to itself in the other cell line (Fig. 39). For example, the miRNA expression landscape after NRAD1 knockdown with GapmeRs 3 and 4 in SUM149 cells was most similar to that of ALDH1A3 knockdown with shRNAs 61 and 63 in SUM149 cells. We observed the same result in MDA-MB-468 cells (Fig. 39). Thus, the role of ALDH1A3 in NRAD1-mediated miRNA expression changes should be investigated in a cell line specific manner.

To study the interactions between miRNAs and a lncRNA or protein, it is the most efficient use of time and resources to select miRNAs that are highly up- or downregulated following knockdown of the lncRNA or protein. Thus, while similarities in miRNA expression in response to knockdown of NRAD1 and ALDH1A3 exist in the absence of a cutoff or at a low fold-change threshold (Fig. 39), despite their possible biological relevance, their experimental validation would be challenging. We therefore applied the cutoff (fold-change $\geq$ 1.3 or $\leq$ -1.3, p < 0.05) to identify miRNAs commonly regulated by NRAD1 and ALDH1A3. We were specifically interested in miRNAs upregulated or downregulated by knockdown of both NRAD1 (both GapmeRs) and ALDH1A3 (both shRNAs) within a cell line (Fig. 40). To this end, the analysis showed only one miRNA, miR-4458, that fit these criteria (Fig. 40B). Specifically, miR-4458 was shown to be downregulated by knockdown of NRAD1 with both GapmeRs and knockdown of ALDH1A3 with both shRNAs in SUM149 cells (Fig. 40B). This miRNA may therefore represent an miRNA regulated by ALDH1A3 through its induction of NRAD1. Together, these analyses suggest that overall, ALDH1A3 does not regulate miRNA expression through its induction of NRAD1 at a level that causes meaningful changes in gene expression. ALDH1A3-mediated induction of NRAD1 is therefore most important in terms of NRAD1-mediated gene regulation in the nucleus. However, our data shows that ALDH1A3 regulates miRNA expression in both MDA-MB-468 and SUM149 cells (Fig. 36), and thus, it is possible that ALDH1A3 induces the expression of lncRNAs other than NRAD1 which interact with the identified miRNAs to influence gene expression in breast cancer.

5.15 Limitations and Future Directions (Data Chapter 2)

Through our analysis of the miRNA landscape following knockdown of NRAD1 or ALDH1A3, we have shown for the first time that both significantly regulate miRNA expression in different breast cancer cell lines. We have identified key miRNAs that may be sponged by NRAD1 to regulate gene expression in MDA-MB-468 and SUM149 cells. The results from this study provide preliminary data to support the further investigation of the identified miRNAs as NRAD1-sponged targets in MDA-MB-468 and SUM149 cells, contributing to regulation of gene expression.

Validation of the miRNA microarray using QPCR in MDA-MB-468 and SUM149 cells is first required to confirm upregulation of miR-595 and miR-4485-3p following NRAD1 knockdown. Next, *in vitro* studies are needed to confirm the interactions of NRAD1 with miR-595 and miR-4485-3p in MDA-MB-468 and SUM149 cells. To this end, a dual-glo luciferase assay should be performed[417]. To perform this assay, a base substitution of around 5 nucleotides should be made in the proposed binding site of the miRNA of interest on the NRAD1 transcript of interest. Wild-type NRAD1 transcripts and mutated transcripts should be cloned into pmiRGLO, an expression vector containing a luciferase reporter gene. Upon transfection of recombinant constructs into MDA-MB-468 or SUM149 cells, mimics of the miRNA of interest should be added to the transfected cells where bioluminescence is subsequently quantified. The absence of a bioluminescent signal indicates miRNA binding, while bioluminescence demonstrates impaired binding ability.

Following confirmation of miRNA binding to NRAD1, the targeting of mRNAs of interest by each miRNA should be investigated. To accomplish this, the dual-glo luciferase assay should be performed, where the sequence of the mRNA of interest is cloned into pmiRGLO. Following addition of the appropriate miRNA mimics, if the miRNA binds the mRNA, no bioluminescence would be observed, thereby confirming the interaction of the miRNA with the mRNA.

Antagomirs, also known as anti-miRs, are oligonucleotides that bind miRNAs and block their binding to mRNAs[418]. Antagomirs can be used to assess the roles of NRAD1/miRNA interactions in NRAD1-induced oncogenic activity including cell proliferation, apoptosis, and mammosphere formation.

Accordingly, NRAD1 knockdown induces apoptosis, and reduces cell proliferation and mammosphere formation in TNBC cell lines[302]. Thus, NRAD1 knockdown in combination with antagomir treatment would determine the role of possible NRAD1-mediated miRNA sponging in the oncogenic function. For example, if NRAD1 knockdown and antagomir treatment results in mammosphere formation, then inhibited mammosphere formation in response to NRAD1 knockdown may be attributed to miRNA-mediated mRNA functions.

Furthermore, given that GapmeR-mediated NRAD1 knockdown inhibits both the long and short transcripts, we are unable to discern whether changes in miRNA expression in response to NRAD1 knockdown are due to inhibition of the short or long transcript variants. The assays investigating miRNA binding to different NRAD1 transcripts may shed light on this. Finally, given the regulation of miRNAs by ALDH1A3 in MDA-MB-468 and SUM149 cells, the identification of novel ALDH1A3-induced lncRNAs may uncover key players in regulation of miRNA expression and mRNA stability in breast cancer in an ALDH1A3-dependant manner.

5.16 Conclusions

In conclusion, TNBC is an aggressive breast cancer subtype that requires significant research to elucidate novel therapeutic targets through which the disease can be treated. A considerable focus has been placed on the potential treatment of TNBCs by targeting TNBC-enriched breast CSCs, which contribute to poor clinical outcomes among patients with this subtype. To this end, the Marcato lab previously identified lncRNA NRAD1 as a key mediator of TNBC progression and stemness that may represent a promising therapeutic target for TNBCs. To build on the previous work, we sought to uncover the mechanisms through which NRAD1 may regulate gene expression in TNBC. We hypothesized that NRAD1 regulates gene expression by interacting with protein partners in the nucleus and by sequestering tumor suppressor miRNAs in the cytoplasm. Our analyses identified the protein S100A8 as an NRAD1-binding protein that co-regulates gene expression with ALDH1A3 and NRAD1 in MDA-MB-468 cells. Further experiments will be needed to assess the degree to which S100A8 contributes to NRAD1-

mediated TNBC progression and stemness. We also show that NRAD1 regulates the expression of miRNAs in MDA-MB-468 and SUM149 TNBC cells, where subsequent analyses identified miRNAs, miR-595 and miR-4485-3p as potential targets for NRAD1-mediated sponging and gene expression regulation. Future experiments will confirm these interactions *in vitro*.

BIBLIOGRAPHY

1. Fares, J., Fares, M. Y., Khachfe, H. H., Salhab, H. A. & Fares, Y. Molecular principles of metastasis: a hallmark of cancer revisited. *Signal Transduction and Targeted Therapy* vol. 5 1–17 (2020).

2. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2019. *CA. Cancer J. Clin.* **69**, 7–34 (2019).

3. DeVita, V. T. & Chu, E. A history of cancer chemotherapy. *Cancer Research* vol. 68 8643–8653 (2008).

4. Pantel, K., Alix-Panabières, C. & Riethdorf, S. Cancer micrometastases. *Nature Reviews Clinical Oncology* vol. 6 339–351 (2009).

5. Esfahani, K. *et al.* A review of cancer immunotherapy: From the past, to the present, to the future. *Curr. Oncol.* **27**, 87–97 (2020).

6. Smith, L. *et al.* Members of the Canadian Cancer Statistics Advisory Committee Analytic leads. (2019).

7. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* vol. 100 57–70 (2000).

8. Balmain, A., Barrett, J. C., Moses, H. & Renan, M. J. How many mutations are required for tumorigenesis? implications from human cancer data. *Mol. Carcinog.* **7**, 139–146 (1993).

9. Rudolph, K. L. *et al.* Longevity, stress response, and cancer in aging telomerase-deficient mice. *Cell* **96**, 701–712 (1999).

10. Kinzler, K. W. & Vogelstein, B. Lessons from hereditary colorectal cancer. *Cell* vol. 87 159–170 (1996).

11. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell* vol. 144 646–674 (2011).

12. Stratton, M. R., Campbell, P. J. & Futreal, P. A. The cancer genome. *Nature* vol. 458 719–724 (2009).

13. Reddy, E. P., Reynolds, R. K., Santos, E. & Barbacid, M. A point mutation is responsible for the acquisition of transforming properties by the T24 human bladder carcinoma oncogene. *Nature* **300**, 149–152 (1982).

14. Vogelstein, B. & Kinzler, K. W. Cancer genes and the pathways they control. *Nature Medicine* vol. 10 789–799 (2004).

15. Croce, C. M. Oncogenes and Cancer. *N. Engl. J. Med.* **358**, 502–511 (2008).

16. Lowe, S. W., Cepero, E. & Evan, G. Intrinsic tumour suppression. *Nature* vol. 432 307–315 (2004).

17. Serrano, M., Lin, A. W., McCurrach, M. E., Beach, D. & Lowe, S. W. Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16(INK4a). *Cell* **88**, 593–602 (1997).

18. Friedberg, E. C. DNA damage and repair. *Nature* vol. 421 436–440 (2003).

19. Zhong, Q., Peng, H. L., Zhao, X., Zhang, L. & Hwang, W. T. Effects of BRCA1- And BRCA2-

related mutations on ovarian and breast cancer survival: A meta-analysis. *Clin. Cancer Res.* **21**, 211–220 (2015).

20. Li, C. I., Uribe, D. J. & Daling, J. R. Clinical characteristics of different histologic types of breast cancer. *Br. J. Cancer* **93**, 1046–1052 (2005).

21. Borst, M. J. & Ingold, J. A. Metastatic patterns of invasive lobular versus invasive ductal carcinoma of the breast. *Surgery* **114**, 637–642 (1993).

22. Burstein, H. J., Polyak, K., Wong, J. S., Lester, S. C. & Kaelin, C. M. Ductal Carcinoma in Situ of the Breast. *N. Engl. J. Med.* **350**, 1430–1441 (2004).

23. Joshi, H. & Press, M. F. Molecular oncology of breast cancer. in *The Breast: Comprehensive Management of Benign and Malignant Diseases* 282-307.e5 (Elsevier Inc., 2018). doi:10.1016/B978-0-323-35955-9.00022-2.

24. Hammond, M. E. H. *et al.* American Society of Clinical Oncology/College Of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *J. Clin. Oncol.* **28**, 2784–95 (2010).

25. Waks, A. G. & Winer, E. P. Breast Cancer Treatment. *JAMA* **321**, 288 (2019).

26. Osborne, C. K. Tamoxifen in the Treatment of Breast Cancer. *N. Engl. J. Med.* **339**, 1609–1618 (1998).

27. Abe, O. *et al.* Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: Patient-level meta-analysis of randomised trials. *Lancet* **378**, 771–784 (2011).

28. Smith, I. E. & Dowsett, M. Aromatase Inhibitors in Breast Cancer. *N. Engl. J. Med.* **348**, 2431–2442 (2003).

29. Fisher, B. *et al.* Tamoxifen for the Prevention of Breast Cancer: Current Status of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. *JNCI J. Natl. Cancer Inst.* **97**, 1652–1662 (2005).

30. Cuzick, J. *et al.* Tamoxifen for prevention of breast cancer: Extended long-term follow-up of the IBIS-I breast cancer prevention trial. *Lancet Oncol.* **16**, 67–75 (2015).

31. Goss, P. E. *et al.* Exemestane for Breast-Cancer Prevention in Postmenopausal Women. *N. Engl. J. Med.* **364**, 2381–2391 (2011).

32. Cuzick, J. *et al.* Anastrozole for prevention of breast cancer in high-risk postmenopausal women (IBIS-II): An international, double-blind, randomised placebo-controlled trial. *Lancet* **383**, 1041–1048 (2014).

33. Davies, C. *et al.* Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years after diagnosis of oestrogen receptor-positive breast cancer: ATLAS, a randomised trial. *Lancet* **381**, 805–816 (2013).

34. Bradley, R. *et al.* Aromatase inhibitors versus tamoxifen in early breast cancer: Patient-level meta-analysis of the randomised trials. *Lancet* **386**, 1341–1352 (2015).

35. Piccart-Gebhart, M. J. *et al.* Trastuzumab after Adjuvant Chemotherapy in HER2-Positive Breast Cancer. *N. Engl. J. Med.* **353**, 1659–1672 (2005).

36. Albain, K. *et al.* Comparisons between different polychemotherapy regimens for early breast cancer: Meta-analyses of long-term outcome among 100 000 women in 123 randomised trials.

Lancet **379**, 432–444 (2012).

37. Voelker, R. Another Targeted Therapy for ERBB2-Positive Breast Cancer. *JAMA* **323**, 408 (2020).

38. Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000).

39. Van De Vijver, M. J. *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347**, 1999–2009 (2002).

40. Sørlie, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10869–10874 (2001).

41. Herschkowitz, J. I. *et al.* Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol.* **8**, R76 (2007).

42. Prat, A. *et al.* Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res.* **12**, R68 (2010).

43. Dent, R. *et al.* Triple-Negative Breast Cancer: Clinical Features and Patterns of Recurrence. *Clin. Cancer Res.* **13**, 4429–4434 (2007).

44. Liedtke, C. *et al.* Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. *J. Clin. Oncol.* **26**, 1275–81 (2008).

45. Foulkes, W. D., Smith, I. E. & Reis-Filho, J. S. Triple-Negative Breast Cancer. *N. Engl. J. Med.* **363**, 1938–1948 (2010).

46. Bauer, K. R., Brown, M., Cress, R. D., Parise, C. A. & Caggiano, V. Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype. *Cancer* **109**, 1721–1728 (2007).

47. Millikan, R. C. *et al.* Epidemiology of basal-like breast cancer. *Breast Cancer Res. Treat.* **109**, 123–139 (2008).

48. Carey, L. A. *et al.* The triple negative paradox: Primary tumor chemosensitivity of breast cancer subtypes. *Clin. Cancer Res.* **13**, 2329–2334 (2007).

49. Denkert, C., Liedtke, C., Tutt, A. & von Minckwitz, G. Molecular alterations in triple-negative breast cancer—the road to new treatment strategies. *Lancet* **389**, 2430–2442 (2017).

50. Anders, C. K. & Carey, L. A. Biology, metastatic patterns, and treatment of patients with triple-negative breast cancer. *Clin. Breast Cancer* **9**, S73–S81 (2009).

51. Dent, R. *et al.* Pattern of metastatic spread in triple-negative breast cancer. *Breast Cancer Res. Treat.* **115**, 423–428 (2009).

52. Hassan, N., Rutsch, N., Győrffy, B., Espinoza-Sánchez, N. A. & Götte, M. SETD3 acts as a prognostic marker in breast cancer patients and modulates the viability and invasion of breast cancer cells. *Sci. Rep.* **10**, 1–16 (2020).

53. von Minckwitz, G. *et al.* Definition and Impact of Pathologic Complete Response on Prognosis After Neoadjuvant Chemotherapy in Various Intrinsic Breast Cancer Subtypes. (2012) doi:10.1200/JCO.2011.41.3161.

54. Audeh, M. W. *et al.* Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and recurrent ovarian cancer: A proof-of-concept trial. *Lancet* **376**, 245–251

(2010).

55. Davies, H. *et al.* HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Anne Vincent-Salomon* **23**, 31 (2017).

56. Robson, M. *et al.* Olaparib for Metastatic Breast Cancer in Patients with a Germline *BRCA* Mutation. *N. Engl. J. Med.* **377**, 523–533 (2017).

57. Loi, S. *et al.* Prognostic and Predictive Value of Tumor-Infiltrating Lymphocytes in a Phase III Randomized Adjuvant Breast Cancer Trial in Node-Positive Breast Cancer Comparing the Addition of Docetaxel to Doxorubicin With Doxorubicin-Based Chemotherapy: BIG 02-98. *J Clin Oncol* **31**, (2013).

58. Adams, S. *et al.* Prognostic value of tumor-infiltrating lymphocytes in triple-negative breast cancers from two phase III randomized adjuvant breast cancer trials: ECOG 2197 and ECOG 1199. *J. Clin. Oncol.* **32**, 2959–2966 (2014).

59. Denkert, C. *et al.* Tumor-Infiltrating Lymphocytes and Response to Neoadjuvant Chemotherapy With or Without Carboplatin in Human Epidermal Growth Factor Receptor 2-Positive and Triple-Negative Primary Breast Cancers. (2014) doi:10.1200/JCO.2014.58.1967.

60. Schmid, P. *et al.* Atezolizumab and Nab-Paclitaxel in Advanced Triple-Negative Breast Cancer. *N. Engl. J. Med.* **379**, 2108–2121 (2018).

61. Wang, C. *et al.* Prognostic value of androgen receptor in triple negative breast cancer: A meta-analysis. *Oncotarget* **7**, 46482–46491 (2016).

62. Gucalp, A. *et al.* Phase II trial of bicalutamide in patients with androgen receptor-positive, estrogen receptor-negative metastatic breast cancer. *Clin. Cancer Res.* **19**, 5505–5512 (2013).

63. Lehmann, B. D. *et al.* PIK3CA mutations in androgen receptor-positive triple negative breast cancer confer sensitivity to the combination of PI3K and androgen receptor inhibitors. *Breast Cancer Res.* **16**, 1–14 (2014).

64. Bonnefoi, H. *et al.* A phase II trial of abiraterone acetate plus prednisone in patients with triple-negative androgen receptor positive locally advanced or metastatic breast cancer (UCBG 12-1). *Ann. Oncol.* **27**, 812–818 (2016).

65. Honeth, G. *et al.* The CD44+/CD24-phenotype is enriched in basal-like breast tumors. *Breast Cancer Res.* **10**, (2008).

66. Giatromanolaki, A., Sividris, E., Fiska, A. & Koukourakis, M. I. The CD44+/CD242 phenotype relates to 'triple-negative' state and unfavorable prognosis in breast cancer patients. doi:10.1007/s12032-010-9530-3.

67. Ricardo, S. *et al.* Breast cancer stem cell markers CD44, CD24 and ALDH1: Expression distribution within intrinsic molecular subtype. *J. Clin. Pathol.* **64**, 937–944 (2011).

68. Idowu, M. O. *et al.* CD44 +/CD24 -/low cancer stem/progenitor cells are more abundant in triple-negative invasive breast carcinoma phenotype and are associated with poor outcome. *Hum. Pathol.* **43**, 364–373 (2012).

69. Li, H. *et al.* Stem cell marker aldehyde dehydrogenase 1 (ALDH1)-expressing cells are enriched in triple-negative breast cancer. *Int. J. Biol. Markers* **28**, (2013).

70. Wu, Y., Sarkissyan, M., Elshimali, Y. & Vadgama, J. V. Triple Negative Breast Tumors in African-American and Hispanic/Latina Women Are High in CD44+, Low in CD24+, and Have

Loss of PTEN. *PLoS One* **8**, (2013).

71. Ma, F. *et al.* Aldehyde dehydrogenase 1 (ALDH1) expression is an independent prognostic factor in triple negative breast cancer (TNBC). *Med. (United States)* **96**, (2017).

72. Wang, H. *et al.* CD44+/CD24- phenotype predicts a poor prognosis in triple-negative breast cancer. *Oncol. Lett.* **14**, 5890–5898 (2017).

73. Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J. & Clarke, M. F. Prospective identification of tumorigenic breast cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 3983–3988 (2003).

74. Thomas, M. L., Coyle, K. M., Sultan, M., Vaghar-Kashani, A. & Marcato, P. Chemoresistance in Cancer Stem Cells and Strategies to Overcome Resistance. (2014) doi:10.4172/2167-7700.1000125.

75. Dalerba, P., Cho, R. W. & Clarke, M. F. Cancer Stem Cells: Models and Concepts. *Annu. Rev. Med.* **58**, 267–284 (2007).

76. Batlle, E. & Clevers, H. Cancer stem cells revisited. *Nat. Med.* **23**, 1124–1134 (2017).

77. Bonnet, D. & Dick, J. E. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat. Med.* **3**, 730–737 (1997).

78. Dalerba, P. *et al.* Phenotypic characterization of human colorectal cancer stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 10158–10163 (2007).

79. Li, F., Tiede, B., Massagué, J. & Kang, Y. Beyond tumorigenesis: Cancer stem cells in metastasis. *Cell Research* vol. 17 3–14 (2007).

80. Baccelli, I. *et al.* Identification of a population of blood circulating tumor cells from breast cancer patients that initiates metastasis in a xenograft assay. *Nat. Biotechnol.* **31**, 539–544 (2013).

81. Phillips, T. M., McBride, W. H. & Pajonk, F. The Response of CD24 −/low /CD44 + Breast Cancer–Initiating Cells to Radiation. *JNCI J. Natl. Cancer Inst.* **98**, 1777–1785 (2006).

82. Lin, Y., Zhong, Y., Guan, H., Zhang, X. & Sun, Q. CD44+/CD24- phenotype contributes to malignant relapse following surgical resection and chemotherapy in patients with invasive ductal carcinoma. *J. Exp. Clin. Cancer Res.* **31**, 1–9 (2012).

83. Dylla, S. J. *et al.* Colorectal Cancer Stem Cells Are Enriched in Xenogeneic Tumors Following Chemotherapy. *PLoS One* **3**, e2428 (2008).

84. Levina, V., Marrangoni, A. M., DeMarco, R., Gorelik, E. & Lokshin, A. E. Drug-Selected Human Lung Cancer Stem Cells: Cytokine Network, Tumorigenic and Metastatic Properties. *PLoS One* **3**, e3077 (2008).

85. Shafee, N. *et al.* Cancer stem cells contribute to cisplatin resistance in Brca1/p53-mediated mouse mammary tumors. *Cancer Res.* **68**, 3243–3250 (2008).

86. Rizzo, S. *et al.* Ovarian cancer stem cell-like side populations are enriched following chemotherapy and overexpress EZH2. *Mol. Cancer Ther.* **10**, 325–335 (2011).

87. Chang, L. *et al.* Acquisition of epithelialmesenchymal transition and cancer stem cell phenotypes is associated with activation of the PI3K/Akt/mTOR pathway in prostate cancer radioresistance. *Cell Death Dis.* **4**, e875–e875 (2013).

88. Abubaker, K. *et al.* Short-term single treatment of chemotherapy results in the enrichment of ovarian cancer stem cell-like cells leading to an increased tumor burden. *Mol. Cancer* **12**, 24 (2013).

89. Hirschmann-Jax, C. *et al.* A distinct 'side population' of cells with high drug efflux capacity in human tumor cells. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 14228–14233 (2004).

90. Bao, S. *et al.* Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. *Nature* **444**, 756–760 (2006).

91. Eyler, C. E. & Rich, J. N. Survival of the fittest: Cancer stem cells in therapeutic resistance and angiogenesis. *Journal of Clinical Oncology* vol. 26 2839–2845 (2008).

92. Liu, P. P. *et al.* Metabolic regulation of cancer cell side population by glucose through activation of the Akt pathway. *Cell Death Differ.* **21**, 124–135 (2014).

93. Sultan, M. *et al.* Epigenetic Silencing of TAP1 in Aldefluor [+] Breast Cancer Stem Cells Contributes to Their Enhanced Immune Evasion. *Stem Cells* **36**, 641–654 (2018).

94. porter, R. K., McNamee, N. & Martinez, V. G. 2-Deoxy-D-Glucose inhibits aggressive triple-negative breast cancer cells by targeting glycolysis and the cancer stem cell phenotype. doi:10.1038/s41598-019-39789-9.

95. Walsh, H. R., Cruickshank, B. M., Brown, J. M. & Marcato, P. The Flick of a Switch: Conferring Survival Advantage to Breast Cancer Stem Cells Through Metabolic Plasticity. *Front. Oncol.* **9**, 753 (2019).

96. Chiba, T. *et al.* Side population purified from hepatocellular carcinoma cells harbors cancer stem cell-like properties. *Hepatology* **44**, 240–251 (2006).

97. Haraguchi, N. *et al.* Characterization of a Side Population of Cancer Cells from Human Gastrointestinal System. *Stem Cells* **24**, 506–513 (2006).

98. Loebinger, M. R. *et al.* Squamous cell cancers contain a side population of stem-like cells that are made chemosensitive by ABC transporter blockade. *Br. J. Cancer* **98**, 380–387 (2008).

99. Wright, M. H. *et al.* Brca1 breast tumors contain distinct CD44+/CD24-and CD133+cells with cancer stem cell characteristics. *Breast Cancer Res.* **10**, (2008).

100. Dean, M. ABC transporters, drug resistance, and cancer stem cells. *J. Mammary Gland Biol. Neoplasia* **14**, 3–9 (2009).

101. Okamoto, A. *et al.* Expansion and characterization of cancer stem-like cells in squamous cell carcinoma of the head and neck. *Oral Oncol.* **45**, 633–639 (2009).

102. Zhu, Z. *et al.* Cancer stem/progenitor cells are highly enriched in CD133 [+] CD44 [+] population in hepatocellular carcinoma. *Int. J. Cancer* **126**, NA-NA (2009).

103. Leccia, F. *et al.* ABCG2, a novel antigen to sort luminal progenitors of BRCA1- breast cancer cells. *Mol. Cancer* **13**, 1–13 (2014).

104. Vasiliou, V. & Nebert, D. W. Analysis and update of the human aldehyde dehydrogenase (ALDH) gene family. *Hum. Genomics* **2**, 138–143 (2005).

105. Cheung, A. M. S. *et al.* Aldehyde dehydrogenase activity in leukemic blasts defines a subgroup of acute myeloid leukemia with adverse prognosis and superior NOD/SCID engrafting potential. *Leukemia* **21**, 1423–1430 (2007).

106.   Ginestier, C. *et al.* ALDH1 Is a Marker of Normal and Malignant Human Mammary Stem Cells and a Predictor of Poor Clinical Outcome. *Cell Stem Cell* **1**, 555–567 (2007).

107.   Tanei, T. *et al.* Association of breast cancer stem cells identified by aldehyde dehydrogenase 1 expression with resistance to sequential paclitaxel and epirubicin-based chemotherapy for breast cancers. *Clin. Cancer Res.* **15**, 4234–4241 (2009).

108.   Peiris-Pagès, M., Martinez-Outschoorn, U. E., Pestell, R. G., Sotgia, F. & Lisanti, M. P. Cancer stem cell metabolism. *Breast Cancer Research* vol. 18 55 (2016).

109.   Jagust, P., De Luxán-Delgado, B., Parejo-Alonso, B. & Sancho, P. Metabolism-based therapeutic strategies targeting cancer stem cells. *Frontiers in Pharmacology* vol. 10 203 (2019).

110.   Koppenol, W. H., Bounds, P. L. & Dang, C. V. Otto Warburg's contributions to current concepts of cancer metabolism. *Nature Reviews Cancer* vol. 11 325–337 (2011).

111.   Lee, Y. *et al.* Cd44+ cells in head and neck squamous cell carcinoma suppress t-cell-mediated immunity by selective constitutive and inducible expression of PD-L1. *Clin. Cancer Res.* **22**, 3571–3581 (2016).

112.   Peitzsch, C., Tyutyunnykova, A., Pantel, K. & Dubrovska, A. Cancer stem cells: The root of tumor recurrence and metastases. *Seminars in Cancer Biology* vol. 44 10–24 (2017).

113.   Mueller, M. T. *et al.* Combined Targeted Treatment to Eliminate Tumorigenic Cancer Stem Cells in Human Pancreatic Cancer. *Gastroenterology* **137**, 1102–1113 (2009).

114.   Li, S. Y. *et al.* Combination therapy with epigenetic-targeted and chemotherapeutic drugs delivered by nanoparticles to enhance the chemotherapy response and overcome resistance by breast cancer stem cells. *J. Control. Release* **205**, 7–14 (2015).

115.   Yun, E. J. *et al.* Targeting Cancer Stem Cells in Castration-Resistant Prostate Cancer. *Clin. Cancer Res.* **22**, 670–679 (2016).

116.   Li, Y., Atkinson, K. & Zhang, T. Combination of chemotherapy and cancer stem cell targeting agents: Preclinical and clinical studies. *Cancer Letters* vol. 396 103–109 (2017).

117.   Marcato, P. *et al.* Aldehyde Dehydrogenase Activity of Breast Cancer Stem Cells Is Primarily Due To Isoform ALDH1A3 and Its Expression Is Predictive of Metastasis. *Stem Cells* **29**, 32–45 (2011).

118.   Feng, J. *et al.* Aldehyde dehydrogenase 1 is a tumor stem cell-Associated marker in lung cancer. *Mol. Cancer Res.* **7**, 330–338 (2009).

119.   Landen, C. N. *et al.* Targeting aldehyde dehydrogenase cancer stem cells in ovarian cancer. *Mol. Cancer Ther.* **9**, 3186–3199 (2010).

120.   Kim, M. P. *et al.* ALDH activity selectively defines an enhanced tumor-initiating cell population relative to CD133 expression in human pancreatic adenocarcinoma. *PLoS One* **6**, (2011).

121.   Singh, S. K. *et al.* Identification of human brain tumour initiating cells. *Nature* **432**, 396–401 (2004).

122.   Hu, Y. & Smyth, G. K. ELDA: Extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *J. Immunol. Methods* **347**, 70–78 (2009).

123.   Schatton, T., Frank, N. Y. & Frank, M. H. Identification and targeting of cancer stem cells. *BioEssays* **31**, 1038–1049 (2009).

124. Yang, X., Sarvestani, S. K., Moeinzadeh, S., He, X. & Jabbari, E. Three-Dimensional-Engineered Matrix to Study Cancer Stem Cells and Tumorsphere Formation: Effect of Matrix Modulus. *Tissue Eng. Part A* **19**, 669–684 (2013).

125. Abbaszadegan, M. R. *et al.* Isolation, identification, and characterization of cancer stem cells: A review. *J. Cell. Physiol.* **232**, 2008–2018 (2017).

126. Lee, C. H., Yu, C. C., Wang, B. Y. & Chang, W. W. Tumorsphere as an effective in vitro platform for screening anticancer stem cell drugs. *Oncotarget* **7**, 1215–1226 (2016).

127. Nagle, P. W., Plukker, J. T. M., Muijs, C. T., van Luijk, P. & Coppes, R. P. Patient-derived tumor organoids for prediction of cancer treatment response. *Seminars in Cancer Biology* vol. 53 258–264 (2018).

128. Li, X. *et al.* Ex Vivo Organoid Cultures Reveal the Importance of the Tumor Microenvironment for Maintenance of Colorectal Cancer Stem Cells. *Cancers (Basel).* **12**, 923 (2020).

129. Visvader, J. E. Cells of origin in cancer. *Nature* vol. 469 314–322 (2011).

130. Chaffer, C. L. & Weinberg, R. A. How does multistep tumorigenesis really proceed? *Cancer Discov.* **5**, 22–24 (2015).

131. Mani, S. A. *et al.* The Epithelial-Mesenchymal Transition Generates Cells with Properties of Stem Cells. *Cell* **133**, 704–715 (2008).

132. Polyak, K. & Weinberg, R. A. Transitions between epithelial and mesenchymal states: Acquisition of malignant and stem cell traits. *Nature Reviews Cancer* vol. 9 265–273 (2009).

133. Liu, S. *et al.* Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. *Stem Cell Reports* **2**, 78–91 (2014).

134. Brabletz, T., Kalluri, R., Nieto, M. A. & Weinberg, R. A. EMT in cancer. *Nature Reviews Cancer* vol. 18 128–134 (2018).

135. Charafe-Jauffret, E. *et al.* Aldehyde dehydrogenase 1-positive cancer stem cells mediate metastasis and poor clinical outcome in inflammatory breast cancer. *Clin. Cancer Res.* **16**, 45–55 (2010).

136. Charafe-Jauffret, E. *et al.* Breast cancer cell lines contain functional cancer stem sells with metastatic capacity and a distinct molecular signature. *Cancer Res.* **69**, 1302–1313 (2009).

137. Zhong, Y. *et al.* ALDH1 is a better clinical indicator for relapse of invasive ductal breast cancer than the CD44 + /CD24 2 phenotype. doi:10.1007/s12032-014-0864-0.

138. Miyoshi, Y. *et al.* Differences in expression of the cancer stem cell marker aldehyde dehydrogenase 1 among estrogen receptor-positive/human epidermal growth factor receptor type 2-negative breast cancer cases with early, late, and no recurrence. *Breast Cancer Res.* **18**, 73 (2016).

139. Wei, W. *et al.* Relationship of CD44+CD24-/low breast cancer stem cells and axillary lymph node metastasis. *J. Transl. Med.* **10 Suppl 1**, 1–6 (2012).

140. Chute, J. P. *et al.* Inhibition of aldehyde dehydrogenase and retinoid signaling induces the expansion of human hematopoietic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 11707–11712 (2006).

141. Marcato, P. *et al.* Aldehyde dehydrogenase 1A3 influences breast cancer progression via

differential retinoic acid signaling. *Mol. Oncol.* **9**, 17–31 (2015).

142. Croker, A. K. & Allan, A. L. Inhibition of aldehyde dehydrogenase (ALDH) activity reduces chemotherapy and radiation resistance of stem-like ALDH hiCD44 + human breast cancer cells. *Breast Cancer Res. Treat.* **133**, 75–87 (2012).

143. Suman, S., Das, T. P. & Damodaran, C. Silencing NOTCH signaling causes growth arrest in both breast cancer stem cells and breast cancer cells. *Br. J. Cancer* **109**, 2587–2596 (2013).

144. Sims-Mourtada, J., Opdenaker, L. M., Davis, J., Arnold, K. M. & Flynn, D. Taxane-induced hedgehog signaling is linked to expansion of breast cancer stem-like populations after chemotherapy. *Mol. Carcinog.* **54**, 1480–1493 (2015).

145. Yue, Z. *et al.* LGR4 modulates breast cancer initiation, metastasis, and cancer stem cells. *FASEB J.* **32**, 2422–2437 (2018).

146. Takebe, N., Warren, R. Q. & Ivy, S. P. Breast cancer growth and metastasis: Interplay between cancer stem cells, embryonic signaling pathways and epithelial-to-mesenchymal transition. *Breast Cancer Research* vol. 13 1–11 (2011).

147. Harrison, H. *et al.* Regulation of breast cancer stem cell activity by signaling through the Notch4 receptor. *Cancer Res.* **70**, 709–718 (2010).

148. Qiu, M. *et al.* Specific inhibition of Notch1 signaling enhances the antitumor efficacy of chemotherapy in triple negative breast cancer through reduction of cancer stem cells. *Cancer Lett.* **328**, 261–270 (2013).

149. Zhao, D. *et al.* NOTCH-induced aldehyde dehydrogenase 1A1 deacetylation promotes breast cancer stem cells. *J. Clin. Invest.* **124**, 5453–5465 (2014).

150. Mamaeva, V. *et al.* Inhibiting notch activity in breast cancer stem cells by glucose functionalized nanoparticles carrying γ-secretase inhibitors. *Mol. Ther.* **24**, 926–936 (2016).

151. Olsauskas-Kuprys, R., Zlobin, A. & Osipo, C. Gamma secretase inhibitors of Notch signaling. *OncoTargets and Therapy* vol. 6 943–955 (2013).

152. Kondratyev, M. *et al.* Gamma-secretase inhibitors target tumor-initiating cells in a mouse model of ERBB2 breast cancer. *Oncogene* **31**, 93–103 (2012).

153. Pandya, K. *et al.* Targeting both Notch and ErbB-2 signalling pathways is required for prevention of ErbB-2-positive breast tumour recurrence. *Br. J. Cancer* **105**, 796–806 (2011).

154. Takebe, N. *et al.* Targeting Notch, Hedgehog, and Wnt pathways in cancer stem cells: Clinical update. *Nature Reviews Clinical Oncology* vol. 12 445–464 (2015).

155. Means-Powell, J. *et al.* Abstract P2-14-04: A Phase Ib Dose Escalation Trial of RO4929097 (a γ-secretase inhibitor) in Combination with Exemestane in Patients with ER + Metastatic Breast Cancer. in *Cancer Research* vol. 72 P2-14-04-P2-14–04 (American Association for Cancer Research (AACR), 2012).

156. Schott, A. F. *et al.* Preclinical and clinical studies of gamma secretase inhibitors with docetaxel on human breast tumors. *Clin. Cancer Res.* **19**, 1512–1524 (2013).

157. Wend, P., Holland, J. D., Ziebold, U. & Birchmeier, W. Wnt signaling in stem and cancer stem cells. *Seminars in Cell and Developmental Biology* vol. 21 855–863 (2010).

158. King, T. D., Suto, M. J. & Li, Y. The wnt/β-catenin signaling pathway: A potential therapeutic

target in the treatment of triple negative breast cancer. *J. Cell. Biochem.* **113**, 13–18 (2012).

159. Duchartre, Y., Kim, Y. M. & Kahn, M. The Wnt signaling pathway in cancer. *Critical Reviews in Oncology/Hematology* vol. 99 141–149 (2016).

160. Jang, G. B. *et al.* Blockade of Wnt/β-catenin signaling suppresses breast cancer metastasis by inhibiting CSC-like phenotype. *Sci. Rep.* **5**, 12465 (2015).

161. Liu, J. *et al.* Targeting Wnt-driven cancer through the inhibition of Porcupine by LGK974. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 20224–20229 (2013).

162. DiMeo, T. A. *et al.* A novel lung metastasis signature links Wnt signaling with cancer cell self-renewal and epithelial-mesenchymal transition in basal-like breast cancer. *Cancer Res.* **69**, 5364–5373 (2009).

163. Liu, J. *et al.* PD08-11: Targeting Porcupine, a Critical Node for Wnt Signalling in Cancer. in *Cancer Research* vol. 71 PD08-11-PD08-11 (American Association for Cancer Research (AACR), 2011).

164. Rodon, J. *et al.* Phase 1 study of single-agent WNT974, a first-in-class Porcupine inhibitor, in patients with advanced solid tumours. *Br. J. Cancer* (2021) doi:10.1038/s41416-021-01389-8.

165. Diamond, J. R. *et al.* Phase Ib clinical trial of the anti-frizzled antibody vantictumab (OMP-18R5) plus paclitaxel in patients with locally advanced or metastatic HER2-negative breast cancer. *Breast Cancer Res. Treat.* **184**, 53–62 (2020).

166. Bhateja, P., Cherian, M., Majumder, S. & Ramaswamy, B. The Hedgehog Signaling Pathway: A Viable Target in Breast Cancer? *Cancers (Basel).* **11**, 1126 (2019).

167. Ramaswamy, B. *et al.* Hedgehog signaling is a novel therapeutic target in tamoxifen-resistant breast cancer aberrantly activated by PI3K/AKT pathway. *Cancer Res.* **72**, 5048–5059 (2012).

168. Fan, P. *et al.* Genistein decreases the breast cancer stem-like cell population through Hedgehog pathway. *Stem Cell Res. Ther.* **4**, 1–10 (2013).

169. Yang, N. *et al.* Inhibition of Sonic Hedgehog Signaling Pathway by Thiazole Antibiotic Thiostrepton Attenuates the CD44+/CD24-Stem-Like Population and Sphere-Forming Capacity in Triple-Negative Breast Cancer. *Cell. Physiol. Biochem.* **38**, 1157–1170 (2016).

170. Koike, Y. *et al.* Anti-cell growth and anti-cancer stem cell activities of the non-canonical hedgehog inhibitor GANT61 in triple-negative breast cancer cells. *Breast Cancer* **24**, 683–693 (2017).

171. Liu, S. *et al.* Hedgehog signaling and Bmi-1 regulate self-renewal of normal and malignant human mammary stem cells. *Cancer Res.* **66**, 6063–6071 (2006).

172. Kubo, M. *et al.* Hedgehog signaling pathway is a new therapeutic target for patients with breast cancer. *Cancer Res.* **64**, 6071–6074 (2004).

173. Dubey, S. Vismodegib: The first drug approved for advanced and metastatic basal cell carcinoma Determinants of blood donation in undergraduate Medical Students in Uttar Pradesh, India View project RADIO-PATHOLOGICAL CORRELATION WITH GUIDED FNAC OR BIOPSY OF INTRA ABDOMINO-PELVIC MASSES View project. *Artic. J. Postgrad. Med.* **59**, 59 (2013).

174. Addition of Vismodegib to Neoadjuvant Chemotherapy in Triple Negative Breast Cancer Patients - Full Text View - ClinicalTrials.gov. https://clinicaltrials.gov/ct2/show/NCT02694224.

175. Fiori, M. E., Villanova, L. & De Maria, R. Cancer stem cells: at the forefront of personalized medicine and immunotherapy. *Current Opinion in Pharmacology* vol. 35 1–11 (2017).

176. Luna, J. I., Grossenbacher, S. K., Murphy, W. J. & Canter, R. J. Targeting Cancer Stem Cells with Natural Killer Cell Immunotherapy. *Expert Opinion on Biological Therapy* vol. 17 313–324 (2017).

177. Codd, A. S., Kanaseki, T., Torigo, T. & Tabi, Z. Cancer stem cells as targets for immunotherapy. *Immunology* **153**, 304–314 (2018).

178. Guo, Y., Feng, K., Wang, Y. & Han, W. Targeting cancer stem cells by using chimeric antigen receptor-modified T cells: a potential and curable approach for cancer treatment. *Protein and Cell* vol. 9 516–526 (2018).

179. Zheng, F. *et al.* Cancer Stem Cell Vaccination With PD-L1 and CTLA-4 Blockades Enhances the Eradication of Melanoma Stem Cells in a Mouse Tumor Model. *J. Immunother.* **41**, 361–368 (2018).

180. Zhang, Y. *et al.* Targeting radiation-resistant prostate cancer stem cells by B7-H3 CAR T cells. *Mol. Cancer Ther.* **20**, 577–588 (2021).

181. Witt, K. *et al.* Cripto-1 plasmid DNA vaccination targets metastasis and cancer stem cells in murine mammary carcinoma. *Cancer Immunol. Res.* **6**, 1417–1425 (2018).

182. Sultan, M. *et al.* Hide-and-seek: the interplay between cancer stem cells and the immune system. *Carcinogenesis* **38**, 107–118 (2017).

183. Quaglino, E., Conti, L. & Cavallo, F. Breast cancer stem cell antigens as targets for immunotherapy. *Seminars in Immunology* vol. 47 101386 (2020).

184. Ribas, A. & Wolchok, J. D. Cancer immunotherapy using checkpoint blockade. *Science* vol. 359 1350–1355 (2018).

185. Wei, S. C., Duffy, C. R. & Allison, J. P. Fundamental mechanisms of immune checkpoint blockade therapy. *Cancer Discovery* vol. 8 1069–1086 (2018).

186. Chambers, C. A., Kuhns, M. S., Egen, J. G. & Allison, J. P. C TLA -4-M EDIATED I NHIBITION IN R EGULATION OF T C ELL R ESPONSES : Mechanisms and Manipulation in Tumor Immunotherapy . *Annu. Rev. Immunol.* **19**, 565–594 (2001).

187. Peggs, K. S., Quezada, S. A., Korman, A. J. & Allison, J. P. Principles and use of anti-CTLA4 antibody in human cancer immunotherapy. *Current Opinion in Immunology* vol. 18 206–213 (2006).

188. Peggs, K. S., Quezada, S. A., Chambers, C. A., Korman, A. J. & Allison, J. P. Blockade of CTLA-4 on both effector and regulatory T cell compartments contributes to the antitumor activity of anti-CTLA-4 antibodies. *J. Exp. Med.* **206**, 1717–1725 (2009).

189. Hodi, F. S. *et al.* Biologic activity of cytotoxic T lymphocyte-associated antigen 4 antibody blockade in previously vaccinated metastatic melanoma and ovarian carcinoma patients. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 4712–4717 (2003).

190. Ribas, A. *et al.* Antitumor Activity in Melanoma and Anti-Self Responses in a Phase I Trial With the Anti-Cytotoxic T Lymphocyte-Associated Antigen 4 Monoclonal Antibody CP-675,206. *J Clin Oncol* **23**, 8968–8977.

191. Hodi, F. S. *et al.* Improved Survival with Ipilimumab in Patients with Metastatic Melanoma. *N.*

Engl. J. Med. **363**, 711–723 (2010).

192. Robert, C. *et al.* Ipilimumab plus Dacarbazine for Previously Untreated Metastatic Melanoma. *N. Engl. J. Med.* **364**, 2517–2526 (2011).

193. Brahmer, J. R. *et al.* Safety and Activity of Anti–PD-L1 Antibody in Patients with Advanced Cancer. *N. Engl. J. Med.* **366**, 2455–2465 (2012).

194. Okazaki, T., Chikuma, S., Iwai, Y., Fagarasan, S. & Honjo, T. A rheostat for immune responses: The unique properties of PD-1 and their advantages for clinical application. *Nature Immunology* vol. 14 1212–1218 (2013).

195. Chen, B. J. *et al.* PD-L1 expression is characteristic of a subset of aggressive B-cell lymphomas and virus-associated malignancies. *Clin. Cancer Res.* **19**, 3462–3473 (2013).

196. Massi, D. *et al.* PD-L1 marks a subset of melanomas with a shorter overall survival and distinct genetic and morphological characteristics. *Ann. Oncol.* **25**, 2433–2442 (2014).

197. Muenst, S. *et al.* Expression of programmed death ligand 1 (PD-L1) is associated with poor prognosis in human breast cancer. *Breast Cancer Res. Treat.* **146**, 15–24 (2014).

198. Lee, Y. & Sunwoo, J. PD-L1 is preferentially expressed on CD44+ tumor-initiating cells in head and neck squamous cell carcinoma. *J. Immunother. Cancer* **2**, P270 (2014).

199. Wu, Y. *et al.* Increased PD-L1 expression in breast and colon cancer stem cells. *Clin. Exp. Pharmacol. Physiol.* **44**, 602–604 (2017).

200. Hsu, J. M. *et al.* STT3-dependent PD-L1 accumulation on cancer stem cells promotes immune evasion. *Nat. Commun.* **9**, 1–17 (2018).

201. Mansour, F. A., Al-Mazrou, A., Al-Mohanna, F., Al-Alwan, M. & Ghebeh, H. PD-L1 is overexpressed on breast cancer stem cells through notch3/mTOR axis. *Oncoimmunology* **9**, 1729299 (2020).

202. Gupta, H. B. *et al.* Tumor B7-H1 regulates cancer stem cell generation and virulence. *J. Immunol.* **196**, (2016).

203. Wolchok, J. D. *et al.* Nivolumab plus Ipilimumab in Advanced Melanoma. *N. Engl. J. Med.* **369**, 122–133 (2013).

204. Garon, E. B. *et al.* Pembrolizumab for the Treatment of Non–Small-Cell Lung Cancer. *N. Engl. J. Med.* **372**, 2018–2028 (2015).

205. Schmid, P. *et al.* Atezolizumab and Nab-Paclitaxel in Advanced Triple-Negative Breast Cancer. *N. Engl. J. Med.* **379**, 2108–2121 (2018).

206. Xu, Q. *et al.* Antigen-Specific T-Cell Response from Dendritic Cell Vaccination Using Cancer Stem-Like Cell-Associated Antigens. *Stem Cells* **27**, 1734–1740 (2009).

207. Ning, N. *et al.* Cancer stem cell vaccination confers significant antitumor immunity. *Cancer Res.* **72**, 1853–1864 (2012).

208. Li, Q. *et al.* Generation of a novel dendritic-cell vaccine using melanoma and squamous cancer stem cells. *J. Vis. Exp.* 50561 (2014) doi:10.3791/50561.

209. Vaccine Therapy in Treating Patients With HER2-Negative Stage III-IV Breast Cancer - Full Text View - ClinicalTrials.gov. https://clinicaltrials.gov/ct2/show/NCT02157051.

210. Wang, D. *et al.* CRISPR Screening of CAR T Cells and Cancer Stem Cells Reveals Critical Dependencies for Cell-Based Therapies. *Cancer Discov.* candisc.1243.2020 (2020) doi:10.1158/2159-8290.cd-20-1243.

211. Zhu, X. *et al.* Patient-derived glioblastoma stem cells are killed by CD133-specific CAR T cells but induce the T cell aging marker CD57. *Oncotarget* **6**, 171–184 (2015).

212. Yang, D. *et al.* T cells expressing NKG2D chimeric antigen receptors efficiently eliminate glioblastoma and cancer stem cells. *J. Immunother. Cancer* **7**, 171 (2019).

213. Byrd, T. T. *et al.* TEM8/ANTXR1-Specific CAR T cells as a targeted therapy for triple-negative breast cancer. *Cancer Res.* **78**, 489–500 (2018).

214. Seitz, C. M. *et al.* GD2-targeted chimeric antigen receptor T cells prevent metastasis formation by elimination of breast cancer stem-like cells. *Oncoimmunology* **9**, 1683345 (2020).

215. Senbanjo, L. T. & Chellaiah, M. A. CD44: A Multifunctional Cell Surface Adhesion Receptor Is a Regulator of Progression and Metastasis of Cancer Cells. *Front. Cell Dev. Biol.* **5**, 18 (2017).

216. Chen, C., Zhao, S., Karnad, A. & Freeman, J. W. The biology and role of CD44 in cancer progression: Therapeutic implications. *Journal of Hematology and Oncology* vol. 11 1–23 (2018).

217. Iida, J. *et al.* DNA Aptamers against Exon v10 of CD44 Inhibit Breast Cancer Cell Migration. *PLoS One* **9**, e88712 (2014).

218. Muntimadugu, E., Kumar, R., Saladi, S., Rafeeqi, T. A. & Khan, W. CD44 targeted chemotherapy for co-eradication of breast cancer stem cells and cancer cells using polymeric nanoparticles of salinomycin and paclitaxel. *Colloids Surfaces B Biointerfaces* **143**, 532–546 (2016).

219. Liu, P. *et al.* Disulfiram targets cancer stem-like cells and reverses resistance and cross-resistance in acquired paclitaxel-resistant triple-negative breast cancer cells. *Br. J. Cancer* **109**, 1876–1885 (2013).

220. Yip, N. C. *et al.* Disulfiram modulated ROS-MAPK and NFB pathways and targeted breast cancer cells with cancer stem cell-like properties. *Br. J. Cancer* **104**, 1564–1574 (2011).

221. Sun, T. *et al.* Induction of immunogenic cell death in radiation-resistant breast cancer stem cells by repurposing anti-alcoholism drug disulfiram. *Cell Commun. Signal.* **18**, 36 (2020).

222. Thomas, M. L. *et al.* Citral reduces breast tumor growth by inhibiting the cancer stem cell marker ALDH1A3. *Mol. Oncol.* **10**, 1485–1496 (2016).

223. Kim, R. J. *et al.* High aldehyde dehydrogenase activity enhances stem cell features in breast cancer cells by activating hypoxia-inducible factor-2α. *Cancer Lett.* **333**, 18–31 (2013).

224. Morgan, C. A., Parajuli, B., Buchman, C. D., Dria, K. & Hurley, T. D. N,N-diethylaminobenzaldehyde (DEAB) as a substrate and mechanism-based inhibitor for human ALDH isoenzymes. *Chem. Biol. Interact.* **234**, 18–28 (2015).

225. Collins, F. S., Morgan, M. & Patrinos, A. The Human Genome Project: Lessons from large-scale biology. *Science* vol. 300 286–290 (2003).

226. Kung, J. T. Y., Colognori, D. & Lee, J. T. Long noncoding RNAs: Past, present, and future. *Genetics* vol. 193 651–669 (2013).

227. Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489**, 101–108 (2012).

228. Fang, Y. & Fullwood, M. J. Roles, Functions, and Mechanisms of Long Non-coding RNAs in Cancer. *Genomics. Proteomics Bioinformatics* **14**, 42–54 (2016).

229. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775–1789 (2012).

230. Gutschner, T. & Diederichs, S. The hallmarks of cancer: a long non-coding RNA point of view. *RNA Biol.* **9**, 703–19 (2012).

231. Rinn, J. L. & Chang, H. Y. Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **81**, 145–66 (2012).

232. Ulitsky, I. Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. *Nat. Rev. Genet.* **17**, 601–614 (2016).

233. Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223–227 (2009).

234. Cabili, M. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25**, 1915–1927 (2011).

235. Katayama, S. *et al.* Molecular biology: Antisense transcription in the mammalian transcriptome. *Science (80-.).* **309**, 1564–1566 (2005).

236. Harrison, P. M., Zheng, D., Zhang, Z., Carriero, N. & Gerstein, M. Transcribed processed pseudogenes in the human genome: An intermediate form of expressed retrosequence lacking protein-coding ability. *Nucleic Acids Res.* **33**, 2374–2383 (2005).

237. Reis, E. M. *et al.* Antisense intronic non-coding RNA levels correlate to the degree of tumor differentiation in prostate cancer. *Oncogene* **23**, 6684–6692 (2004).

238. Nakaya, H. I. *et al.* Genome mapping and expression analyses of human intronic noncoding RNAs reveal tissue-specific patterns and enrichment in genes related to regulation of transcription. *Genome Biol.* **8**, R43 (2007).

239. Fernandes, J., Acuña, S., Aoki, J., Floeter-Winter, L. & Muxel, S. Long Non-Coding RNAs in the Regulation of Gene Expression: Physiology and Disease. *Non-Coding RNA* **5**, 17 (2019).

240. Volders, P. J. *et al.* Lncipedia 5: Towards a reference set of human long non-coding rnas. *Nucleic Acids Res.* **47**, D135–D139 (2019).

241. Ponjavic, J., Ponting, C. P. & Lunter, G. Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Res.* **17**, 556–565 (2007).

242. Johnsson, P., Lipovich, L., Grandér, D. & Morris, K. V. Evolutionary conservation of long non-coding RNAs; Sequence, structure, function. *Biochimica et Biophysica Acta - General Subjects* vol. 1840 1063–1071 (2014).

243. Carlevaro-Fita, J. *et al.* Cancer LncRNA Census reveals evidence for deep functional conservation of long noncoding RNAs in tumorigenesis. *Commun. Biol.* **3**, 1–16 (2020).

244. Marín-Béjar, O. *et al.* The human lncRNA LINC-PINT inhibits tumor cell invasion through a highly conserved sequence element. *Genome Biol.* **18**, 202 (2017).

245. Iyer, M. K. *et al.* The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.* **47**, 199–208 (2015).

246. Clemson, C. M., McNeil, J. A., Willard, H. F. & Lawrence, J. B. XIST RNA paints the inactive X chromosome at interphase: Evidence for a novel RNA involved in nuclear/chromosome structure. *J. Cell Biol.* **132**, 259–275 (1996).

247. Rocha, S. T. da, Edwards, C. A., Ito, M., Ogata, T. & Ferguson-Smith, A. C. Genomic imprinting at the mammalian Dlk1-Dio3 domain. *Trends in Genetics* vol. 24 306–316 (2008).

248. Mohamed, J. S., Gaughwin, P. M., Lim, B., Robson, P. & Lipovich, L. Conserved long noncoding RNAs transcriptionally regulated by Oct4 and Nanog modulate pluripotency in mouse embryonic stem cells. *RNA* **16**, 324–337 (2010).

249. Ng, S.-Y., Johnson, R. & Stanton, L. W. Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. *EMBO J.* **31**, 522–533 (2012).

250. Ng, S. Y., Bogu, G. K., Soh, B. S. & Stanton, L. W. The long noncoding RNA RMST interacts with SOX2 to regulate neurogenesis. *Mol. Cell* **51**, 349–359 (2013).

251. Rani, N. *et al.* A Primate lncRNA Mediates Notch Signaling during Neuronal Development by Sequestering miRNA. *Neuron* **90**, 1174–1188 (2016).

252. Bond, A. M. *et al.* Balanced gene regulation by an embryonic brain ncRNA is critical for adult hippocampal GABA circuitry. *Nat. Publ. Gr.* (2009) doi:10.1038/nn.2371.

253. Coghlan, S. *et al.* GABA system dysfunction in autism and related disorders: From synapse to symptoms. *Neuroscience and Biobehavioral Reviews* vol. 36 2044–2055 (2012).

254. Puts, N. A. J. *et al.* Reduced GABA and altered somatosensory function in children with autism spectrum disorder. *Autism Res.* **10**, 608–619 (2017).

255. Gale, K. GABA and epilepsy: basic concepts from preclinical research. *Epilepsia* **33 Suppl 5**, S3-12 (1992).

256. Gonzalez-Burgos, G., Hashimoto, T. & Lewis, D. A. Alterations of cortical GABA neurons and network oscillations in schizophrenia. *Current Psychiatry Reports* vol. 12 335–344 (2010).

257. Batista, P. J. & Chang, H. Y. Long noncoding RNAs: Cellular address codes in development and disease. *Cell* vol. 152 1298–1307 (2013).

258. Zhou, S. *et al.* The regulatory roles of lncRNAs in the process of breast cancer invasion and metastasis. *Biosci. Rep.* **38**, (2018).

259. Willard, H. F. *X Chromosome Inactivation, Minireview XIST, and Pursuit of the X-Inactivation Center. Cell* vol. 86 (1996).

260. Rinn, J. L. *et al.* Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs. *Cell* **129**, 1311–1323 (2007).

261. Gabory, A., Ripoche, M.-A., Yoshimizu, T. & Dandolo, L. The <i>H19</i> gene: regulation and function of a non-coding RNA. *Cytogenet. Genome Res.* **113**, 188–193 (2006).

262. Prensner, J. R. & Chinnaiyan, A. M. The emergence of lncRNAs in cancer biology. *Cancer Discovery* vol. 1 391–407 (2011).

263. Gutschner, T. & Diederichs, S. The hallmarks of cancer: A long non-coding RNA point of view. *RNA Biology* vol. 9 703–709 (2012).

264. Huarte, M. The emerging role of lncRNAs in cancer. *Nature Medicine* vol. 21 1253–1261 (2015).

265. Luo, M. & Liang, C. LncRNA LINC00483 promotes gastric cancer development through regulating MAPK1 expression by sponging miR-490-3p. *Biol Res* **53**, 14 (2020).

266. Gutschner, T. *et al.* The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res.* **73**, 1180–1189 (2013).

267. Shi, S. J. *et al.* LncRNA-ATB promotes trastuzumab resistance and invasionmetastasis cascade in breast cancer. *Oncotarget* **6**, 11652–11663 (2015).

268. Brown, J. M., Wasson, M.-C. D. & Marcato, P. The Missing Lnc: The Potential of Targeting Triple-Negative Breast Cancer and Cancer Stem Cells by Inhibiting Long Non-Coding RNAs. *Cells* **9**, 763 (2020).

269. Xu, R. *et al.* LncRNA XIST/miR-200c regulates the stemness properties and tumourigenicity of human bladder cancer stem cell-like cells. *Cancer Cell Int.* **18**, 1–10 (2018).

270. Zhao, Y., Feng, C., Li, Y., Ma, Y. & Cai, R. LncRNA H19 promotes lung cancer proliferation and metastasis by inhibiting miR-200a function. *Mol. Cell. Biochem.* **460**, 1–8 (2019).

271. Bhan, A. & Mandal, S. S. LncRNA HOTAIR: A master regulator of chromatin dynamics and cancer. *Biochimica et Biophysica Acta - Reviews on Cancer* vol. 1856 151–164 (2015).

272. Gupta, R. A. *et al.* Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464**, 1071–1076 (2010).

273. Kogo, R. *et al.* Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. *Cancer Res.* **71**, 6320–6326 (2011).

274. Luo, M. *et al.* Long non-coding RNA H19 increases bladder cancer metastasis by associating with EZH2 and inhibiting E-cadherin expression. *Cancer Lett.* **333**, 213–221 (2013).

275. Abdollahzadeh, R. *et al.* Competing endogenous RNA (ceRNA) cross talk and language in ceRNA regulatory networks: A new look at hallmarks of breast cancer. *J. Cell. Physiol.* **234**, 10080–10100 (2019).

276. Chou, J. *et al.* MALAT1 induced migration and invasion of human breast cancer cells by competitively binding MIR-1 with cdc42. *Biochem. Biophys. Res. Commun.* **472**, 262–269 (2016).

277. Jin, C., Yan, B., Lu, Q., Lin, Y. & Ma, L. Reciprocal regulation of Hsa-miR-1 and long noncoding RNA MALAT1 promotes triple-negative breast cancer development. *Tumor Biol.* **37**, 7383–7394 (2016).

278. Wang, Y. *et al.* MiR-204/ZEB2 axis functions as key mediator for MALATI-induced epithelial-mesenchymal transition in breast cancer. *Tumor Biol.* **39**, 1–8 (2017).

279. Zheng, A. *et al.* Long non-coding RNA LUCAT1/miR-5582-3p/TCF7L2 axis regulates breast cancer stemness via Wnt/β-catenin pathway. *J. Exp. Clin. Cancer Res.* **38**, 305 (2019).

280. Tang, T. *et al.* LncCCAT1 Promotes Breast Cancer Stem Cell Function through Activating WNT/β-catenin Signaling. *Theranostics* **9**, 7384–7402 (2019).

281. Zhong, Y. X., Li, W. S., Liao, L. S. & Liang, L. LncRNA CCAT1 promotes cell proliferation and differentiation via negative modulation of miRNA-218 in human DPSCs. *Eur. Rev. Med. Pharmacol. Sci.* **23**, 3575–3583 (2019).

282. Wang, R., Zhang, T., Yang, Z., Jiang, C. & Seng, J. Long non-coding RNA FTH1P3 activates paclitaxel resistance in breast cancer through miR-206/ABCB1. *J. Cell. Mol. Med.* **22**, 4068–4075 (2018).

283. Zheng, P. *et al.* Long noncoding RNA CASC2 promotes paclitaxel resistance in breast cancer through regulation of miR-18a-5p/CDK19. *Histochem. Cell Biol.* **152**, 281–291 (2019).

284. Chang, L., Hu, Z., Zhou, Z. & Zhang, H. Linc00518 Contributes to Multidrug Resistance Through Regulating the MiR-199a/MRP1 Axis in Breast Cancer. *Cell. Physiol. Biochem.* **48**, 16–28 (2018).

285. Wang, Y. *et al.* Long noncoding RNA AC073284.4 suppresses epithelial–mesenchymal transition by sponging miR-18b-5p in paclitaxel-resistant breast cancer cells. *J. Cell. Physiol.* **234**, 23202–23215 (2019).

286. Augoff, K., McCue, B., Plow, E. F. & Sossey-Alaoui, K. MiR-31 and its host gene lncRNA LOC554202 are regulated by promoter hypermethylation in triple-negative breast cancer. *Mol. Cancer* **11**, 5 (2012).

287. Sun, B. *et al.* Research progress on the interactions between long non-coding RNAs and microRNAs in human cancer (Review). *Oncol. Lett.* **19**, 595–605 (2020).

288. Alles, J. *et al.* An estimate of the total number of true human miRNAs. *Nucleic Acids Res.* **47**, 3353–3364 (2019).

289. Yang, L. *et al.* LncRNAs regulate cancer metastasis via binding to functional proteins. *Oncotarget* vol. 9 1426–1443 (2018).

290. Xing, Z. *et al.* LncRNA directs cooperative epigenetic regulation downstream of chemokine signals. *Cell* **159**, 1110–1125 (2014).

291. Gooding, A. J. *et al.* The lncRNA BORG Drives Breast Cancer Metastasis and Disease Recurrence. *Sci. Rep.* **7**, (2017).

292. Gooding, A. J. *et al.* The lncRNA BORG facilitates the survival and chemoresistance of triple-negative breast cancers. *Oncogene* **38**, 2020–2041 (2019).

293. Cerami, E. *et al.* The cBio Cancer Genomics Portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).

294. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, 1–1 (2013).

295. Loewer, S. *et al.* Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat. Genet.* **42**, 1113–1117 (2010).

296. Deng, J. *et al.* Long Non-Coding RNA HOTAIR Regulates the Proliferation, Self-Renewal Capacity, Tumor Formation and Migration of the Cancer Stem-Like Cell (CSC) Subpopulation Enriched from Breast Cancer Cells. *PLoS One* **12**, e0170860 (2017).

297. Zhou, M. *et al.* LncRNA-Hh Strengthen Cancer Stem Cells Generation in Twist-Positive Breast Cancer via Activation of Hedgehog Signaling Pathway. *Stem Cells* **34**, 55–66 (2016).

298. Brannan, C. I., Dees, E. C., Ingram, R. S. & Tilghman, S. M. The product of the H19 gene may function as an RNA. *Mol. Cell. Biol.* **10**, 28–36 (1990).

299. Peng, F. *et al.* Glycolysis gatekeeper PDK1 reprograms breast cancer stem cells under hypoxia. *Oncogene* **37**, 1062–1074 (2018).

300. Shima, H. *et al.* Lnc RNA H19 is associated with poor prognosis in breast cancer patients and promotes cancer stemness. *Breast Cancer Res. Treat.* **170**, 507–516 (2018).

301. Zhang, Y. *et al.* Long noncoding RNA LINP1 regulates repair of DNA double-strand breaks in triple-negative breast cancer. *Nat. Struct. Mol. Biol.* **23**, 522–530 (2016).

302. Vidovic, D. *et al.* ALDH1A3-regulated long non-coding RNA NRAD1 is a potential novel target for triple-negative breast tumors and cancer stem cells. *Cell Death Differ.* **27**, 363–378 (2020).

303. Peng, F. *et al.* H19/let-7/LIN28 reciprocal negative regulatory circuit promotes breast cancer stem cell maintenance. *Cell Death Dis.* **8**, (2017).

304. Shin, V. Y. *et al.* Long non-coding RNA NEAT1 confers oncogenic role in triple-negative breast cancer through modulating chemoresistance and cancer stemness. *Cell Death Dis.* **10**, (2019).

305. Bamodu, O. A. *et al.* Aberrant KDM5B expression promotes aggressive breast cancer through MALAT1 overexpression and downregulation of hsa-miR-448. *BMC Cancer* **16**, 160 (2016).

306. Latorre, E. *et al.* The ribonucleic complex HuR-MALAT1 represses CD133 expression and suppresses epithelial-mesenchymal transition in breast cancer. *Cancer Res.* **76**, 2626–2636 (2016).

307. Xing, Z., Park, P. K., Lin, C. & Yang, L. LncRNA BCAR4 wires up signaling transduction in breast cancer. *RNA Biol.* **12**, 681–689 (2015).

308. Sha, S., Yuan, D., Liu, Y., Han, B. & Zhong, N. Targeting long non-coding RNA DANCR inhibits triple negative breast cancer progression. *Biol. Open* **6**, 1310–1316 (2017).

309. Hou, P. *et al.* LincRNA-ROR induces epithelial-to-mesenchymal transition and contributes to breast cancer tumorigenesis and metastasis. *Cell Death Dis.* **5**, (2014).

310. Tu, Z., Schmöllerl, J., Cuiffo, B. G. & Karnoub, A. E. Microenvironmental Regulation of Long Noncoding RNA LINC01133 Promotes Cancer Stem Cell-Like Phenotypic Traits in Triple-Negative Breast Cancers. *Stem Cells* **37**, 1281–1292 (2019).

311. Li, H. *et al.* Long noncoding RNA linc00617 exhibits oncogenic activity in breast cancer. *Mol. Carcinog.* **56**, 3–17 (2017).

312. Song, X. *et al.* LncRNA SPRY4-IT1 regulates breast cancer cell stemness through competitively binding miR-6882-3p with TCF7L2. *J. Cell. Mol. Med.* jcmm.14786 (2019) doi:10.1111/jcmm.14786.

313. Jia, X. *et al.* KLF5 regulated lncRNA RP1 promotes the growth and metastasis of breast cancer via repressing p27kip1 translation. *Cell Death Dis.* **10**, 373 (2019).

314. Lu, G. *et al.* Long noncoding RNA LINC00511 contributes to breast cancer tumourigenesis and stemness by inducing the miR-185-3p/E2F1/Nanog axis. *J. Exp. Clin. Cancer Res.* **37**, 289 (2018).

315. Zhang, Z. *et al.* Long non-coding RNA FEZF1-AS1 promotes breast cancer stemness and tumorigenesis via targeting miR-30a/Nanog axis. *J. Cell. Physiol.* **233**, 8630–8638 (2018).

316. Keshavarz, M. & Asadi, M. H. Long non-coding <scp>RNA ES</scp> 1 controls the proliferation of breast cancer cells by regulating the Oct4/Sox2/miR-302 axis. *FEBS J.* **286**, 2611–2623 (2019).

317. García-Venzor, A. *et al.* Microenvironment-regulated lncRNA-HAL is able to promote stemness in breast cancer cells. *Biochim. Biophys. Acta - Mol. Cell Res.* **1866**, 118523 (2019).

318. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in caenorhabditis

elegans. *Nature* **391**, 806–811 (1998).

319. Moore, C. B., Guthrie, E. H., Huang, M. T. H. & Taxman, D. J. Short hairpin RNA (shRNA): design, delivery, and assessment of gene knockdown. *Methods Mol. Biol.* **629**, 141–158 (2010).

320. Barata, P., Sood, A. K. & Hong, D. S. RNA-targeted therapeutics in cancer clinical trials: Current status and future directions. *Cancer Treatment Reviews* vol. 50 35–47 (2016).

321. Lennox, K. A. & Behlke, M. A. Cellular localization of long non-coding RNAs affects silencing by RNAi more than by antisense oligonucleotides. *Nucleic Acids Res.* **44**, 863–877 (2016).

322. Vaidya, A. M. *et al.* Systemic Delivery of Tumor-Targeting siRNA Nanoparticles against an Oncogenic LncRNA Facilitates Effective Triple-Negative Breast Cancer Therapy. *Bioconjug. Chem.* **30**, 907–919 (2019).

323. Schoch, K. M. & Miller, T. M. Antisense Oligonucleotides: Translation from Mouse Models to Human Neurodegenerative Diseases. *Neuron* vol. 94 1056–1070 (2017).

324. Eder, P. S., Devine, R. J., Dagle, J. M. & Walder, J. A. Substrate Specificity and Kinetics of Degradation of Antisense Oligonucleotides by a 3' Exonuclease in Plasma. *Antisense Res. Dev.* **1**, 141–151 (1991).

325. Geary, R. S., Norris, D., Yu, R. & Bennett, C. F. Pharmacokinetics, biodistribution and cell uptake of antisense oligonucleotides. *Advanced Drug Delivery Reviews* vol. 87 46–51 (2015).

326. Lennox, K. A. & Behlke, M. A. Tips for successful lncRNA knockdown using gapmers. in *Methods in Molecular Biology* vol. 2176 121–140 (Humana Press Inc., 2020).

327. Bennett, C. F. & Swayze, E. E. RNA targeting therapeutics: Molecular mechanisms of antisense oligonucleotides as a therapeutic platform. *Annual Review of Pharmacology and Toxicology* vol. 50 259–293 (2010).

328. Passini, M. A. *et al.* Antisense oligonucleotides delivered to the mouse CNS ameliorate symptoms of severe spinal muscular atrophy. *Sci. Transl. Med.* **3**, 72ra18-72ra18 (2011).

329. Miller, T. M. *et al.* An antisense oligonucleotide against SOD1 delivered intrathecally for patients with SOD1 familial amyotrophic lateral sclerosis: A phase 1, randomised, first-in-man study. *Lancet Neurol.* **12**, 435–442 (2013).

330. Miller, T. *et al.* Phase 1–2 Trial of Antisense Oligonucleotide Tofersen for SOD1 ALS . *N. Engl. J. Med.* **383**, 109–119 (2020).

331. Kordasiewicz, H. B. *et al.* Sustained Therapeutic Reversal of Huntington's Disease by Transient Repression of Huntingtin Synthesis. *Neuron* **74**, 1031–1044 (2012).

332. Leavitt, B. *et al.* Discovery and Early Clinical Development of ISIS-HTTRx, the First HTT-Lowering Drug to Be Tested in Patients with Huntington's Disease (PL01.002). *Neurology* **86**, (2016).

333. Gutschner, T. *et al.* The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res.* **73**, 1180–1189 (2013).

334. Xing, Z. *et al.* LncRNA directs cooperative epigenetic regulation downstream of chemokine signals. *Cell* **159**, 1110–1125 (2014).

335. Amodio, N. *et al.* Drugging the lncRNA MALAT1 via LNA gapmeR ASO inhibits gene expression of proteasome subunits and triggers anti-multiple myeloma activity. *Leukemia* **32**,

1948–1957 (2018).

336. Leucci, E. *et al.* Melanoma addiction to the long non-coding RNA SAMMSON. *Nature* **531**, 518–522 (2016).

337. Yang, J. *et al.* CRISPR/Cas9-mediated noncoding RNA editing in human cancers. *RNA Biol.* **15**, 35–43 (2018).

338. Zhen, S. *et al.* Inhibition of long non-coding RNA UCA1 by CRISPR/Cas9 attenuated malignant phenotypes of bladder cancer. *Oncotarget* **8**, 9634–9646 (2017).

339. Ruan, Z. & Zhao, D. Long intergenic noncoding RNA LINC00284 knockdown reduces angiogenesis in ovarian cancer cells *via* up-regulation of MEST through NF-κB1. *FASEB J.* **33**, 12047–12059 (2019).

340. Wang, S. *et al.* Construction and Investigation of an LINC00284-Associated Regulatory Network in Serous Ovarian Carcinoma. *Dis. Markers* **2020**, (2020).

341. Xing, C. *et al.* Identification of Potential Biomarkers Involved in Gastric Cancer Through Integrated Analysis of Non-Coding RNA Associated Competing Endogenous RNAs Network. *Clin. Lab.* **64**, 1661–1669 (2018).

342. Wang, H. *et al.* Comprehensive analysis of aberrantly expressed profiles of lncRNAs and miRNAs with associated ceRNA network in muscleinvasive bladder cancer. *Oncotarget* **7**, 86174–86185 (2016).

343. Zhao, Y. *et al.* Construction and investigation of lncRNA-associated ceRNA regulatory network in papillary thyroid cancer. *Oncol. Rep.* **39**, 1197–1206 (2018).

344. Dar, A. A. *et al.* miRNA-205 suppresses melanoma cell proliferation and induces senescence via regulation of E2F1 protein. *J. Biol. Chem.* **286**, 16606–16614 (2011).

345. Elgamal, O. A. *et al.* Tumor Suppressive Function of mir-205 in Breast Cancer Is Linked to HMGB3 Regulation. *PLoS One* **8**, e76402 (2013).

346. Xiao, Y., Humphries, B., Yang, C. & Wang, Z. MiR-205 dysregulations in breast cancer: The complexity and opportunities. *Non-coding RNA* vol. 5 53 (2019).

347. Yan, D., Wu, F., Peng, C. & Wang, M. Silencing of LINC00284 inhibits cell proliferation and migration in oral squamous cell carcinoma by the miR-211-3p/MAFG axis and FUS/KAZN axis. *Cancer Biol. Ther.* **22**, 149–163 (2021).

348. Cailleau, R., Olivé, M. & Cruciger, Q. V. J. Long-term human breast carcinoma cell lines of metastatic origin: Preliminary characterization. *In Vitro* **14**, 911–915 (1978).

349. Forozan, F. *et al.* Molecular cytogenetic analysis of 11 new breast cancer cell lines. *Br. J. Cancer* **81**, 1328–1334 (1999).

350. Marín-Béjar, O. & Huarte, M. RNA pulldown protocol for in vitro detection and identification of RNA-associated proteins. *Methods Mol. Biol.* **1206**, 87–95 (2015).

351. Chu, C. & Chang, H. Y. ChIRP-MS: RNA-Directed proteomic discovery. in *Methods in Molecular Biology* vol. 1861 37–45 (Humana Press Inc., 2018).

352. Lili, L. N., Matyunina, L. V., Walker, L. D., Benigno, B. B. & McDonald, J. F. Molecular profiling predicts the existence of two functionally distinct classes of ovarian cancer stroma. *Biomed Res. Int.* **2013**, (2013).

353. Howe, K. L. *et al.* Ensembl 2021. *Nucleic Acids Res.* **49**, D884–D891 (2021).

354. Kozomara, A., Birgaoanu, M. & Griffiths-Jones, S. MiRBase: From microRNA sequences to function. *Nucleic Acids Res.* **47**, D155–D162 (2019).

355. Agarwal, V., Bell, G. W., Nam, J. W. & Bartel, D. P. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* **4**, (2015).

356. Chu, C., Quinn, J. & Chang, H. Y. Chromatin isolation by RNA purification (ChIRP). *J. Vis. Exp.* 3912 (2012) doi:10.3791/3912.

357. Fu, S. J. *et al.* Hornerin promotes tumor progression and is associated with poor prognosis in hepatocellular carcinoma. *BMC Cancer* **18**, 815 (2018).

358. Zhang, Y. *et al.* Characterization of long noncoding RNA-associated proteins by RNA-immunoprecipitation. in *Methods in Molecular Biology* vol. 1402 19–26 (Humana Press Inc., 2016).

359. Chen, Q. *et al.* Long noncoding RNA NEAT1, regulated by the EGFR pathway, contributes to glioblastoma progression through the WNT/b-catenin pathway by scaffolding EZH2. *Clin. Cancer Res.* **24**, 684–695 (2018).

360. Bresnick, A. R., Weber, D. J. & Zimmer, D. B. S100 proteins in cancer. *Nature Reviews Cancer* vol. 15 96–109 (2015).

361. Shabani, F., Farasat, A., Mahdavi, M. & Gheibi, N. Calprotectin (S100A8/S100A9): a key protein between inflammation and cancer. *Inflammation Research* vol. 67 801–812 (2018).

362. Cortesi, L. *et al.* Protein expression patterns associated with advanced stage ovarian cancer. *Electrophoresis* **32**, 1992–2003 (2011).

363. Saghaeian Jazi, M., Mansour Samaei, N., Ghanei, M., Behgam Shadmehr, M. & Javad Mowla, S. Identification of new SOX2OT transcript variants highly expressed in human cancer cell lines and down regulated in stem cell differentiation. doi:10.1007/s11033-015-3939-x.

364. Chang, X., Zhang, H., Yang, Q. & Pang, L. LncRNA SOX2OT affects cervical cancer cell growth, migration and invasion by regulating SOX2. *Cell Cycle* **19**, 1391–1403 (2020).

365. Zhu, Y. *et al.* LncRNA NEAT1 remodels chromatin to promote the 5-Fu resistance by maintaining colorectal cancer stemness. *Cell Death Dis.* **11**, 1–11 (2020).

366. Venkatesh, J., Wasson, M.-C. D., Brown, J. M., Fernando, W. & Marcato, P. LncRNA-miRNA axes in breast cancer: Novel points of interaction for strategic attack. *Cancer Lett.* **509**, 81–88 (2021).

367. Guo, X. *et al.* Long non-coding RNAs function annotation: A global prediction method based on bi-colored networks. *Nucleic Acids Res.* **41**, e35 (2013).

368. Guo, Q. *et al.* Comprehensive analysis of lncRNA-mRNA co-expression patterns identifies immune-associated lncRNA biomarkers in ovarian cancer malignant progression. *Sci. Rep.* **5**, 17683 (2015).

369. Wu, G. & Stein, L. A network module-based method for identifying cancer prognostic signatures. *Genome Biol.* **13**, R112 (2012).

370. Huarte, M. *et al.* A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* **142**, 409–419 (2010).

371. Li, D. *et al.* Inhibition of glutamine metabolism counteracts pancreatic cancer stem cell features and sensitizes cells to radiotherapy. *Oncotarget* **6**, 31151–31163 (2015).

372. De Francesco, E. M., Maggiolini, M., Tanowitz, H. B., Sotgia, F. & Lisanti, M. P. Targeting hypoxic cancer stem cells (CSCs) with Doxycycline: Implications for optimizing anti-angiogenic therapy. *Oncotarget* **8**, 56126–56142 (2017).

373. Serio, R. N., Lu, C., Gross, S. S. & Gudas, L. J. Different Effects of Knockouts in ALDH2 and ACSS2 on Embryonic Stem Cell Differentiation. *Alcohol. Clin. Exp. Res.* **43**, 1859–1871 (2019).

374. Chen, L. *et al.* Silencing transcription factor FOXM1 represses proliferation, migration, and invasion while inducing apoptosis of liver cancer stem cells by regulating the expression of ALDH2. *IUBMB Life* **72**, 285–295 (2020).

375. Garcia-Heredia, J. M., Lucena-Cacace, A., Verdugo-Sivianes, E. M., Perez, M. & Carnero, A. The cargo protein MAP17 (PDZK1IP1) regulates the cancer stem cell pool activating the Notch pathway by abducting NUMB. *Clin. Cancer Res.* **23**, 3871–3883 (2017).

376. Zheng, L. *et al.* Transcriptional factor six2 promotes the competitive endogenous RNA network between CYP4Z1 and pseudogene CYP4Z2P responsible for maintaining the stemness of breast cancer cells. *J. Hematol. Oncol.* **12**, 23 (2019).

377. Gebhardt, C., Németh, J., Angel, P. & Hess, J. S100A8 and S100A9 in inflammation and cancer. *Biochem. Pharmacol.* **72**, 1622–1631 (2006).

378. Wang, S. *et al.* S100A8/A9 in inflammation. *Frontiers in Immunology* vol. 9 1298 (2018).

379. Kwon, C. H., Moon, H. J., Park, H. J., Choi, J. H. & Park, D. Y. S100A8 and S100A9 Promotes Invasion and Migration through p38 Mitogen-Activated Protein Kinase-Dependent NF-κB Activation in Gastric Cancer Cells. *Mol. Cells* **35**, 226–234 (2013).

380. Grebhardt, S., Müller-Decker, K., Bestvater, F., Hershfinkel, M. & Mayer, D. Impact of S100A8/A9 expression on prostate cancer progression in vitro and in vivo. *J. Cell. Physiol.* **229**, 661–671 (2014).

381. Mohammed AlQahtani, and H. *et al.* Impact of S100A8 Expression on Kidney Cancer Progression and Molecular Docking Studies for Kidney Cancer Therapeutics.

382. Lim, S. Y., Yuzhalin, A. E., Gordon-Weeks, A. N. & Muschel, R. J. Tumor-infiltrating monocytes/macrophages promote tumor invasion and migration by upregulating S100A8 and S100A9 expression in cancer cells. *Oncogene* **35**, 5735–5745 (2016).

383. Nedjadi, T. *et al.* S100A8 and S100A9 proteins form part of a paracrine feedback loop between pancreatic cancer cells and monocytes. *BMC Cancer* **18**, 1255 (2018).

384. Wang, D. *et al.* Clinical Significance of Elevated S100A8 Expression in Breast Cancer Patients. *Front. Oncol.* **8**, 496 (2018).

385. Yin, C. *et al.* RAGE-binding S100A8/A9 promotes the migration and invasion of human breast cancer cells through actin polymerization and epithelial-mesenchymal transition. *Breast Cancer Res. Treat.* **142**, 297–309 (2013).

386. Moon, A. *et al.* Global gene expression profiling unveils S100A8/A9 as candidate markers in H-ras-mediated human breast epithelial cell invasion. *Mol. Cancer Res.* **6**, 1544–1553 (2008).

387. Chen, Y. *et al.* Critical role of the MCAM-ETV4 axis triggered by extracellular S100A8/A9 in breast cancer aggressiveness. *Neoplasia (United States)* **21**, 627–640 (2019).

388.   Miller, P. *et al.* Elevated S100A8 protein expression in breast cancer cells and breast tumor stroma is prognostic of poor disease outcome. *Breast Cancer Res. Treat.* **166**, 85–94 (2017).

389.   Bao, Y., Wang, A. & Mo, J. S100A8/A9 is associated with estrogen receptor loss in breast cancer. *Oncol. Lett.* **11**, 1936–1942 (2016).

390.   Eeckhoute, J. *et al.* Positive cross-regulatory loop ties GATA-3 to estrogen receptor α expression in breast cancer. *Cancer Res.* **67**, 6477–6483 (2007).

391.   Yan, W., Cao, Q. J., Arenas, R. B., Bentley, B. & Shao, R. GATA3 inhibits breast cancer metastasis through the reversal of epithelial-mesenchymal transition. *J. Biol. Chem.* **285**, 14042–14051 (2010).

392.   Schonthaler, H. B. *et al.* S100A8-S100A9 Protein Complex Mediates Psoriasis by Regulating the Expression of Complement Factor C3. *Immunity* **39**, 1171–1181 (2013).

393.   Song, R. & Struhl, K. S100A8/S100A9 cytokine acts as a transcriptional coactivator during breast cellular transformation. *Sci. Adv.* **7**, eabe5357 (2021).

394.   Kraakman, M. J. *et al.* Neutrophil-derived S100 calcium-binding proteins A8/A9 promote reticulated thrombocytosis and atherogenesis in diabetes. *J. Clin. Invest.* **127**, 2133–2147 (2017).

395.   Björk, P. *et al.* Identification of human S100A9 as a novel target for treatment of autoimmune disease via binding to quinoline-3-carboxamides. *PLoS Biol.* **7**, 0800–0812 (2009).

396.   Jennbacken, K. *et al.* Inhibition of metastasis in a castration resistant prostate cancer model by the quinoline-3-carboxamide tasquinimod (ABR-215050). *Prostate* **72**, 913–924 (2012).

397.   Shen, L. *et al.* Tasquinimod modulates suppressive myeloid cells and enhances cancer immunotherapies in murine models. *Cancer Immunol. Res.* **3**, 136–148 (2015).

398.   Pili, R. *et al.* Phase II Randomized, Double-Blind, Placebo-Controlled Study of Tasquinimod in Men With Minimally Symptomatic Metastatic Castrate-Resistant Prostate Cancer. *Genitourin. Cancers Symp.* **29**, (2011).

399.   Sternberg, C. *et al.* Randomized, double-blind, placebo-controlled phase III study of tasquinimod in men with metastatic castration-resistant prostate cancer. in *Journal of Clinical Oncology* vol. 34 2636–2643 (American Society of Clinical Oncology, 2016).

400.   Isaacs, J. T. *et al.* Tasquinimod is an allosteric modulator of HDAC4 survival signaling within the compromised cancer microenvironment. *Cancer Res.* **73**, 1386–1399 (2013).

401.   Fedele, P., Orlando, L. & Cinieri, S. Targeting triple negative breast cancer with histone deacetylase inhibitors. *Expert Opinion on Investigational Drugs* vol. 26 1199–1206 (2017).

402.   Schech, A., Kazi, A., Yu, S., Shah, P. & Sabnis, G. Histone deacetylase inhibitor entinostat inhibits tumor-initiating cells in triple-negative breast cancer cells. *Mol. Cancer Ther.* **14**, 1848–1857 (2015).

403.   Chen, S. *et al.* Comparative proteomics of glioma stem cells and differentiated tumor cells identifies S100A9 as a potential therapeutic target. *J. Cell. Biochem.* **114**, 2795–2808 (2013).

404.   Wei, R. *et al.* S100 calcium-binding protein A9 from tumor-associated macrophage enhances cancer stem cell-like properties of hepatocellular carcinoma. *Int. J. Cancer* **148**, 1233–1244 (2021).

405.   Duan, L. *et al.* S100A8 and S100A9 Are Associated with Colorectal Carcinoma Progression and

Contribute to Colorectal Carcinoma Cell Survival and Migration via Wnt/β-Catenin Pathway. *PLoS One* **8**, e62092 (2013).

406. Ertem, F. U. *et al.* Oncogenic targets Mmp7, S100a9, Nppb and Aldh1a3 from transcriptome profiling of FAP and Pirc adenomas are downregulated in response to tumor suppression by Clotam. *Int. J. Cancer* **140**, 460–468 (2017).

407. Lane, M. A. & Bailey, S. J. Role of retinoid signalling in the adult brain. *Progress in Neurobiology* vol. 75 275–293 (2005).

408. Vilhais-Neto, G. C. & Pourquié, O. Retinoic acid (DOI:10.1016/j.cub.2007.12.042). *Current Biology* vol. 18 550–552 (2008).

409. Yamashita, D. *et al.* Identification of ALDH1A3 as a viable therapeutic target in breast cancer metastasis-initiating cells. *Mol. Cancer Ther.* **19**, 1134–1147 (2020).

410. Park, P. J. ChIP-seq: Advantages and challenges of a maturing technology. *Nature Reviews Genetics* vol. 10 669–680 (2009).

411. Zhang, Z. *et al.* Negative regulation of lncRNA GAS5 by miR-21. *Cell Death Differ.* **20**, 1558–1568 (2013).

412. Li, M. *et al.* MiR-7 reduces the BCSC subset by inhibiting XIST to modulate the miR-92b/Slug/ESA axis and inhibit tumor growth. *Breast Cancer Res.* **22**, (2020).

413. Sripada, L. *et al.* hsa-miR-4485 regulates mitochondrial functions and inhibits the tumorigenicity of breast cancer cells. *J. Mol. Med.* **95**, 641–651 (2017).

414. Tian, S., Zhang, M., Chen, X., Liu, Y. & Lou, G. MicroRNA-595 sensitizes ovarian cancer cells to cisplatin by targeting ABCB1. *Oncotarget* **7**, 87091–87099 (2016).

415. O'Brien, J., Hayder, H., Zayed, Y. & Peng, C. Overview of microRNA biogenesis, mechanisms of actions, and circulation. *Frontiers in Endocrinology* vol. 9 402 (2018).

416. Sun, B. *et al.* Research progress on the interactions between long non-coding RNAs and microRNAs in human cancer (Review). *Oncol. Lett.* **19**, 595–605 (2020).

417. Yang, T. *et al.* MicroRNA-15a induces cell apoptosis and inhibits metastasis by targeting BCL2L2 in non-small cell lung cancer. *Tumor Biol.* **36**, 4357–4365 (2015).

418. Song, M. S. & Rossi, J. J. The anti-miR21 antagomir, a therapeutic tool for colorectal cancer, has a potential synergistic effect by perturbing an angiogenesis-associated miR30. *Front. Genet.* **4**, 301 (2013).

APPENDICES

Appendix 1. Proteins identified via mass spectrometry following NRAD1 pulldown with T7 and Sp6-synthesized NRAD1 transcripts.

Gene Symbol	Protein Name
KRT1	[Master Protein] Keratin, type II cytoskeletal 1
KRT9	[Master Protein] Keratin, type I cytoskeletal 9
KRT10	[Master Protein] Keratin, type I cytoskeletal 10
KRT2	[Master Protein] Keratin, type II cytoskeletal 2 epidermal
HSPA8	[Master Protein] Heat shock cognate 71 kDa protein
LMNA	[Master Protein] Prelamin-A/C
ANXA2	[Master Protein] Annexin A2
HNRNPK	[Master Protein] Heterogeneous nuclear ribonucleoprotein K
HNRNPM	[Master Protein] Heterogeneous nuclear ribonucleoprotein M
HNRNPA2B1	[Master Protein] Heterogeneous nuclear ribonucleoproteins A2/B1
RNH1	[Master Protein] Ribonuclease inhibitor
HNRNPA1	[Master Protein] Heterogeneous nuclear ribonucleoprotein A1
ILF2	[Master Protein] Interleukin enhancer-binding factor 2
ACTB	[Master Protein] Actin, cytoplasmic 1
HNRNPC	[Master Protein] Heterogeneous nuclear ribonucleoproteins C1/C2
NCL	[Master Protein] Nucleolin
RPS3A	[Master Protein] 40S ribosomal protein S3a
ILF3	[Master Protein] Interleukin enhancer-binding factor 3
NPM1	[Master Protein] Nucleophosmin
EEF1A1	[Master Protein] Elongation factor 1-alpha 1
RBMX	[Master Protein] RNA-binding motif protein, X chromosome
KRT5	[Master Protein] Keratin, type II cytoskeletal 5
RPL23A	[Master Protein] 60S ribosomal protein L23a
RPS18	[Master Protein] 40S ribosomal protein S18
RPS4X	[Master Protein] 40S ribosomal protein S4, X isoform
PTBP1	[Master Protein] Polypyrimidine tract-binding protein 1
RPS19	[Master Protein] 40S ribosomal protein S19
RPL26	[Master Protein] 60S ribosomal protein L26
C11orf98	[Master Protein] Chromosome 11 open reading frame 48, isoform CRA_c
PABPC1	[Master Protein] Polyadenylate-binding protein 1
HNRNPU	[Master Protein] Heterogeneous nuclear ribonucleoprotein U
HSPA1B	[Master Protein] Heat shock 70 kDa protein 1B
RPS3	[Master Protein] 40S ribosomal protein S3
HIST1H4A	[Master Protein] Histone H4
RPL35	[Master Protein] 60S ribosomal protein L35
RPS25	[Master Protein] 40S ribosomal protein S25
DHX9	[Master Protein] ATP-dependent RNA helicase A
EZR	[Master Protein] Ezrin
HNRNPL	[Master Protein] Heterogeneous nuclear ribonucleoprotein L
KRT16	[Master Protein] Keratin, type I cytoskeletal 16
LMNB1	[Master Protein] Lamin-B1
RPS13	[Master Protein] 40S ribosomal protein S13

Appendix 1. Proteins identified via mass spectrometry following NRAD1 pulldown with T7 and Sp6-synthesized NRAD1 transcripts.

RPS7	[Master Protein] 40S ribosomal protein S7
AHNAK	[Master Protein] Neuroblast differentiation-associated protein AHNAK
DEK	[Master Protein] Protein DEK
HNRNPA3	[Master Protein] Heterogeneous nuclear ribonucleoprotein A3
RPL31	[Master Protein] 60S ribosomal protein L31
SERBP1	[Master Protein] Plasminogen activator inhibitor 1 RNA-binding protein
SUB1	[Master Protein] Activated RNA polymerase II transcriptional coactivator p15
ALYREF	[Master Protein] THO complex subunit 4
EIF2S3	[Master Protein] Eukaryotic translation initiation factor 2 subunit 3
KRT14	[Master Protein] Keratin, type I cytoskeletal 14
RBBP4	[Master Protein] Histone-binding protein RBBP4
RPL18	[Master Protein] 60S ribosomal protein L18 (Fragment)
RPLP2	[Master Protein] 60S acidic ribosomal protein P2
RPS11	[Master Protein] 40S ribosomal protein S11
RPS9	[Master Protein] 40S ribosomal protein S9
SRP14	[Master Protein] Signal recognition particle 14 kDa protein
SYNCRIP	[Master Protein] Heterogeneous nuclear ribonucleoprotein Q
TRIM28	[Master Protein] Transcription intermediary factor 1-beta
DDX17	[Master Protein] Probable ATP-dependent RNA helicase DDX17
FLNA	[Master Protein] Filamin-A
HSPA5	[Master Protein] 78 kDa glucose-regulated protein
RPL10A	[Master Protein] 60S ribosomal protein L10a
RPL17-C18orf32	[Master Protein] Protein RPL17-C18orf32
RPL22	[Master Protein] 60S ribosomal protein L22
RPL30	[Master Protein] 60S ribosomal protein L30
RPL7A	[Master Protein] 60S ribosomal protein L7a
RPS8	[Master Protein] 40S ribosomal protein S8
YBX1	[Master Protein] Nuclease-sensitive element-binding protein 1 (Fragment)
ALB	[Master Protein] Serum albumin
DCD	[Master Protein] Dermcidin
DDX5	[Master Protein] Probable ATP-dependent RNA helicase DDX5
HIST1H1C	[Master Protein] Histone H1.2
HNRNPAB	[Master Protein] Heterogeneous nuclear ribonucleoprotein A/B
KHDRBS1	[Master Protein] KH domain-containing, RNA-binding, signal transduction-associated protein 1
RPL13	[Master Protein] 60S ribosomal protein L13
RPL3	[Master Protein] 60S ribosomal protein L3
RPLP0	[Master Protein] 60S acidic ribosomal protein P0
RPS14	[Master Protein] 40S ribosomal protein S14
RPS20	[Master Protein] 40S ribosomal protein S20
SRSF3	[Master Protein] Serine/arginine-rich splicing factor 3
TUBA1C	[Master Protein] Tubulin alpha-1C chain
TUBB	[Master Protein] Tubulin beta chain
ELAVL1	[Master Protein] ELAV-like protein 1
ERH	[Master Protein] Enhancer of rudimentary homolog

Appendix 1. Proteins identified via mass spectrometry following NRAD1 pulldown with T7 and Sp6-synthesized NRAD1 transcripts.

FAU	[Master Protein] 40S ribosomal protein S30
HIST1H1B	[Master Protein] Histone H1.5
HIST1H1E	[Master Protein] Histone H1.4
HNRNPA0	[Master Protein] Heterogeneous nuclear ribonucleoprotein A0
HNRNPH1	[Master Protein] Heterogeneous nuclear ribonucleoprotein H
HNRNPR	[Master Protein] Heterogeneous nuclear ribonucleoprotein R
HRNR	[Master Protein] Hornerin
RPL28	[Master Protein] 60S ribosomal protein L28
RPL8	[Master Protein] 60S ribosomal protein L8
RPS16	[Master Protein] 40S ribosomal protein S16
RPSA	[Master Protein] 40S ribosomal protein SA
CBX3	[Master Protein] Chromobox protein homolog 3
CFL1	[Master Protein] Cofilin-1
EEF1D	[Master Protein] Elongation factor 1-delta
EEF1G	[Master Protein] Elongation factor 1-gamma
EIF2S1	[Master Protein] Eukaryotic translation initiation factor 2 subunit 1
HMGN2	[Master Protein] Non-histone chromosomal protein HMG-17
MATR3	[Master Protein] Matrin-3
RPL19	[Master Protein] 60S ribosomal protein L19
RPL27	[Master Protein] 60S ribosomal protein L27
RPL6	[Master Protein] 60S ribosomal protein L6
RPL9	[Master Protein] 60S ribosomal protein L9
RPS17	[Master Protein] 40S ribosomal protein S17
RPS21	[Master Protein] 40S ribosomal protein S21
RPS27A	[Master Protein] Ubiquitin-40S ribosomal protein S27a
RPS28	[Master Protein] 40S ribosomal protein S28
RPS5	[Master Protein] 40S ribosomal protein S5
RPS6	[Master Protein] 40S ribosomal protein S6
S100A8	[Master Protein] Protein S100-A8
S100A9	[Master Protein] Protein S100-A9
SRRM1	[Master Protein] Serine/arginine repetitive matrix protein 1
VDAC2	[Master Protein] Voltage-dependent anion-selective channel protein 2
YWHAZ	[Master Protein] 14-3-3 protein zeta/delta
BUB3	[Master Protein] Mitotic checkpoint protein BUB3
CAPRIN1	[Master Protein] Caprin-1
CHTOP	[Master Protein] Chromatin target of PRMT1 protein
CIRBP	[Master Protein] Cold-inducible RNA-binding protein
EEF1B2	[Master Protein] Elongation factor 1-beta
FUS	[Master Protein] RNA-binding protein FUS
HMGN1	[Master Protein] Non-histone chromosomal protein HMG-14
HNRNPD	[Master Protein] Heterogeneous nuclear ribonucleoprotein D0
HSP90AA1	[Master Protein] Heat shock protein HSP 90-alpha
KRT6A	[Master Protein] Keratin, type II cytoskeletal 6A
NUDT21	[Master Protein] Cleavage and polyadenylation specificity factor subunit 5
PUF60	[Master Protein] Poly(U)-binding-splicing factor PUF60
RPL11	[Master Protein] 60S ribosomal protein L11

Appendix 1. Proteins identified via mass spectrometry following NRAD1 pulldown with T7 and Sp6-synthesized NRAD1 transcripts.

RPL13A	[Master Protein] 60S ribosomal protein L13a
RPL18A	[Master Protein] 60S ribosomal protein L18a
RPL24	[Master Protein] 60S ribosomal protein L24
RPL27A	[Master Protein] 60S ribosomal protein L27a
RPL29	[Master Protein] 60S ribosomal protein L29
RPL34	[Master Protein] 60S ribosomal protein L34
RPL4	[Master Protein] 60S ribosomal protein L4
RPL7	[Master Protein] 60S ribosomal protein L7
RPS12	[Master Protein] 40S ribosomal protein S12
SAP18	[Master Protein] Histone deacetylase complex subunit SAP18
SNRPA	[Master Protein] U1 small nuclear ribonucleoprotein A
SNRPD3	[Master Protein] Small nuclear ribonucleoprotein Sm D3
SRSF7	[Master Protein] Serine/arginine-rich splicing factor 7
YWHAB	[Master Protein] 14-3-3 protein beta/alpha
DSP	[Master Protein] Desmoplakin
G3BP1	[Master Protein] Ras GTPase-activating protein-binding protein 1
GAPDH	[Master Protein] Glyceraldehyde-3-phosphate dehydrogenase
GNB2L1	[Master Protein] Guanine nucleotide-binding protein subunit beta-2-like 1 (Fragment)
H2AFZ	[Master Protein] Histone H2A.Z
HIST1H2AA	[Master Protein] Histone H2A type 1-A
HMGA1	[Master Protein] High mobility group protein HMG-I/HMG-Y
HNRNPF	[Master Protein] Heterogeneous nuclear ribonucleoprotein F
MYL6	[Master Protein] Myosin light polypeptide 6
PNN	[Master Protein] Pinin
PRPF19	[Master Protein] Pre-mRNA-processing factor 19
RALY	[Master Protein] RNA-binding protein Raly
RPL14	[Master Protein] 60S ribosomal protein L14
RPL35A	[Master Protein] 60S ribosomal protein L35a
RPL38	[Master Protein] 60S ribosomal protein L38
S100A14	[Master Protein] Protein S100-A14
S100A16	[Master Protein] Protein S100-A16
SNRPA1	[Master Protein] U2 small nuclear ribonucleoprotein A'
SNRPB	[Master Protein] Small nuclear ribonucleoprotein-associated proteins B and B'
SNRPD2	[Master Protein] Small nuclear ribonucleoprotein Sm D2
SNRPE	[Master Protein] Small nuclear ribonucleoprotein E
SPTAN1	[Master Protein] Spectrin alpha chain, non-erythrocytic 1
SRSF1	[Master Protein] Serine/arginine-rich-splicing factor 1
HNRNPDL	[Master Protein] Heterogeneous nuclear ribonucleoprotein D-like
KRT6B	[Master Protein] Keratin, type II cytoskeletal 6B
KRT74	[Master Protein] Keratin, type II cytoskeletal 74
PABPN1	[Master Protein] Polyadenylate-binding protein 2
PGM2	[Master Protein] Phosphoglucomutase-2
RPL32	[Master Protein] 60S ribosomal protein L32
RPL37	[Master Protein] 60S ribosomal protein L37
RPS23	[Master Protein] 40S ribosomal protein S23
RPS29	[Master Protein] 40S ribosomal protein S29

Appendix 1. Proteins identified via mass spectrometry following NRAD1 pulldown with T7 and Sp6-synthesized NRAD1 transcripts.

S100A10	[Master Protein] Protein S100-A10
S100A6	[Master Protein] Protein S100-A6
SF3B4	[Master Protein] Splicing factor 3B subunit 4
SNRPF	[Master Protein] Small nuclear ribonucleoprotein F
SQSTM1	[Master Protein] Sequestosome-1
YWHAQ	[Master Protein] 14-3-3 protein theta

Appendix 2. NRAD1-regulated genes identified via microarray similarly correlated with NRAD1 and S100A8 in patient breast tumors. NRAD1-regulated genes identified using the Affymetrix Human Gene 2.0 ST microarray platform (control GapmeR vs. GapmeR #4; n=3) in MDA-MB-468 cells (≥ 1.2 or ≤ -1.2-fold expression change, $p < 0.05$) were assessed for their Spearman co-expression correlations with NRAD1 and S100A8 in patient breast tumors. To identify co-expression Spearman correlations, RNA-sequencing data (RNA-seq RSEM V2) of patient breast tumors (invasive breast carcinoma) was extracted from the TCGA, Cell 2015 dataset (n = 816), accessible via cBioPortal. The Spearman correlations of NRAD1-upregulated genes (green) positively correlated, and NRAD1-downregulated genes (red) negatively correlated (Spearman correlation > 0.25 or < -0.25) with NRAD1 and S100A8 in patient breast tumors are shown.

Gene Symbol	Fold-Change	NRAD1	S100A8
LEMD1	5.34	0.441709051	0.311897379
SLC15A1	3.07	0.35845799	0.462348891
SLC34A2	2.58	0.506204087	0.305216968
FAM49A	2.41	0.359673892	0.421125362
PLAGL1	2.24	0.431369595	0.34528171
STEAP3	2.2	0.285662501	0.295926391
SLC28A1	2.13	0.276124201	0.318461444
PDZK1IP1	2.11	0.293048134	0.306339503
MUC16	2.05	0.391671949	0.266493597
FMO6P	1.99	0.333407194	0.330411112
MRAS	1.92	0.369875009	0.397045795
ELF5	1.9	0.426243662	0.284900588
PLCH1	1.9	0.295110339	0.370981314
ST6GAL1	1.86	0.299624214	0.324513752
CHODL	1.84	0.417143142	0.286682146
ACSL5	1.83	0.342141801	0.321936559
B3GNT3	1.81	0.384375444	0.474996669
SPNS2	1.81	0.324272897	0.394180108
ACSL4	1.79	0.332960249	0.412127808
AKR1B15	1.76	0.267401372	0.452726685
PLCG2	1.76	0.366487676	0.318905782
KLK7	1.75	0.46426314	0.267996644
EFNA5	1.7	0.384255752	0.259002326
PDXK	1.68	0.27869899	0.33390334
ASS1	1.66	0.385990522	0.538949041
OPN3	1.64	0.322876279	0.343140456
PTK7	1.63	0.3402218	0.353536764
RCAN1	1.63	0.502748345	0.364053241
ETV5	1.59	0.324676744	0.253860562
TGM1	1.59	0.387112544	0.366098405

Appendix 2. NRAD1-regulated genes identified via microarray similarly correlated with NRAD1 and S100A8 in patient breast tumors.

LCN2	1.58	0.386502087	0.513359981
CRABP1	1.56	0.440248565	0.347682474
RARRES1	1.55	0.528423404	0.575891968
KCNG1	1.54	0.360537213	0.385768077
KRT17	1.53	0.439174291	0.276803543
ZNF438	1.53	0.258008424	0.268155418
MAML2	1.52	0.418394914	0.253669896
SLFN11	1.52	0.318563096	0.402165484
TMEM40	1.5	0.328266283	0.510367138
CD82	1.49	0.359720711	0.334420607
PLS3	1.49	0.3229367	0.340054364
S1PR2	1.49	0.342364099	0.417322324
S100A7	1.48	0.265406588	0.707605745
CNGA1	1.47	0.360871429	0.302687564
GBGT1	1.47	0.298964774	0.274727375
LINC01119	1.47	0.325904093	0.323133323
RHBDL2	1.47	0.289602535	0.344698881
UBASH3B	1.47	0.330645689	0.503516159
CPNE2	1.46	0.258540731	0.378219889
ITM2C	1.45	0.333156383	0.360095271
TNFRSF21	1.44	0.373545594	0.391293571
VGLL1	1.44	0.550226351	0.413663763
PRKCA	1.43	0.310854129	0.287547193
CDH3	1.42	0.471335895	0.397472127
DAPK1	1.42	0.296797749	0.39276861
TANK	1.42	0.393199792	0.266418587
PDE9A	1.4	0.363549382	0.268188791
PROM1	1.4	0.547684914	0.345423375
RASGRF1	1.4	0.27943511	0.461336544
FGF11	1.39	0.344874882	0.361269292
PDE7A	1.39	0.33964964	0.356600339
PML	1.39	0.315287501	0.41943279
FUT3	1.38	0.347147769	0.568361826
RGL1	1.38	0.276641144	0.309125785
CLDN1	1.37	0.360716048	0.334084021
ZIC1	1.37	0.360685466	0.251917484
ANXA8	1.36	0.402996441	0.276529962
GLYATL2	1.36	0.421085026	0.358015963

Appendix 2. NRAD1-regulated genes identified via microarray similarly correlated with NRAD1 and S100A8 in patient breast tumors.

C3	1.35	0.327805489	0.294340659
KLK8	1.35	0.433815181	0.272375804
ZSWIM4	1.35	0.356539596	0.385928698
FOLR1	1.34	0.539174905	0.25282248
SLC2A12	1.34	0.444168312	0.264989012
ST8SIA1	1.34	0.473786478	0.402533786
NCCRP1	1.33	0.385679637	0.510338372
CPAMD8	1.32	0.375849402	0.256415781
MME	1.32	0.269761967	0.256871729
BACE2	1.3	0.368608886	0.345334925
MET	1.3	0.355933311	0.29490888
CXCL16	1.29	0.35817543	0.329158421
LY6D	1.29	0.408633285	0.448922884
MICAL3	1.29	0.349434039	0.283068387
AIF1L	1.28	0.271974173	0.347361587
CEBPB	1.28	0.402756983	0.588139324
EGFR	1.28	0.392874854	0.332659327
ANXA8	1.27	0.402996441	0.276529962
WNT6	1.27	0.421826165	0.339742249
EML4	1.26	0.262549876	0.439727584
KRT4	1.25	0.262838732	0.354270895
SCPEP1	1.25	0.383279203	0.411082503
KRT5	1.24	0.400582953	0.252330047
PSAT1	1.24	0.361086601	0.519501547
PSAT1	1.24	0.361086601	0.519501547
SERPINH1	1.24	0.252570238	0.431111848
SVOPL	1.24	0.396952013	0.25823221
TTLL4	1.23	0.372367237	0.35641641
MPZL2	1.22	0.467227925	0.366842567
RASGEF1B	1.21	0.334282756	0.408935263
TOX	1.21	0.32154247	0.293972413
MSN	1.2	0.314046877	0.449804345
PLA2G4A	1.2	0.40375745	0.335674701
TBC1D1	1.2	0.310301973	0.282809841
CRNKL1	-1.24	-0.282593015	-0.286019403
RAD17	-1.26	-0.25924134	-0.398376614
GATA3	-1.27	-0.34419013	-0.498391147
TMEM161B	-1.27	-0.255762368	-0.410421278

Appendix 2. NRAD1-regulated genes identified via microarray similarly correlated with NRAD1 and S100A8 in patient breast tumors.

KCTD6	-1.46	-0.281417069	-0.353348108
SLC39A6	-1.46	-0.315443396	-0.44199857
ZNF814	-1.5	-0.300444614	-0.427061063
ZXDB	-1.55	-0.276565309	-0.339103033
GUF1	-1.56	-0.265372585	-0.272092984
DNAJC12	-1.73	-0.308825992	-0.457743806

Appendix 3. NRAD1-regulated genes identified via microarray similarly correlated with NRAD1, S100A8, S100A9, and ALDH1A3 in patient breast tumors. NRAD1-regulated genes identified using the Affymetrix Human Gene 2.0 ST microarray platform (control GapmeR vs. GapmeR #4; n=3) in MDA-MB-468 cells ($\geq$ 1.2 or $\leq$ -1.2-fold expression change, p < 0.05) were assessed for their Spearman co-expression correlations with NRAD1, S100A8, S100A9, and ALDH1A3 in patient breast tumors. To identify co-expression Spearman correlations, RNA-sequencing data (RNA-seq RSEM V2) of patient breast tumors (invasive breast carcinoma) was extracted from the TCGA, Cell 2015 dataset (n = 816), accessible via cBioPortal. The Spearman correlations of NRAD1-upregulated genes (green) positively correlated, and NRAD1-downregulated genes (red) negatively correlated (Spearman correlation > 0.25 or < -0.25) with NRAD1, S100A8, S100A9, and ALDH1A3 in patient breast tumors are shown.

Gene Symbol	Fold-Change	NRAD1	S100A8	S100A9	ALDH1A3
LEMD1	5.34	0.44170905	0.31189738	0.33176433	0.32694952
SLC15A1	3.07	0.35845799	0.46234889	0.47323661	0.3380984
SLC34A2	2.58	0.50620409	0.30521697	0.37829334	0.39127625
FAM49A	2.41	0.35967389	0.42112536	0.4305646	0.54118862
PLAGL1	2.24	0.4313696	0.34528171	0.36848496	0.56171877
MUC16	2.05	0.39167195	0.2664936	0.2921526	0.282426
FMO6P	1.99	0.33340719	0.33041111	0.31764687	0.30437588
MRAS	1.92	0.36987501	0.3970458	0.40719706	0.51656473
ELF5	1.9	0.42624366	0.28490059	0.31373038	0.31376584
ST6GAL1	1.86	0.29962421	0.32451375	0.37002105	0.3685334
CHODL	1.84	0.41714314	0.28668215	0.30614667	0.2966257
ACSL5	1.83	0.3421418	0.32193656	0.34936252	0.39783562
B3GNT3	1.81	0.38437544	0.47499667	0.51770537	0.43021111
SPNS2	1.81	0.3242729	0.39418011	0.43549967	0.51735323
ACSL4	1.79	0.33296025	0.41212781	0.4032833	0.5309088
AKR1B15	1.76	0.26740137	0.45272669	0.42242098	0.36867254
PLCG2	1.76	0.36648768	0.31890578	0.31522878	0.44266136
KLK7	1.75	0.46426314	0.26799664	0.311566	0.38251488
EFNA5	1.7	0.38425575	0.25900233	0.31361752	0.29292853
OPN3	1.64	0.32287628	0.34314046	0.36408518	0.30511865
PTK7	1.63	0.3402218	0.35353676	0.38647371	0.34363726
RCAN1	1.63	0.50274835	0.36405324	0.41049058	0.53632002
ETV5	1.59	0.32467674	0.25386056	0.29079252	0.41971075
TGM1	1.59	0.38711254	0.36609841	0.38507097	0.29067631
LCN2	1.58	0.38650209	0.51335998	0.54161025	0.28426375
CRABP1	1.56	0.44024857	0.34768247	0.40217195	0.26256244
RARRES1	1.55	0.5284234	0.57589197	0.61180152	0.53393214
KCNG1	1.54	0.36053721	0.38576808	0.41974646	0.30449391
KRT17	1.53	0.43917429	0.27680354	0.31588749	0.34233021
MAML2	1.52	0.41839491	0.2536699	0.28523714	0.57817326

Appendix 3. NRAD1-regulated genes identified via microarray similarly correlated with NRAD1, S100A8, S100A9, and ALDH1A3 in patient breast tumors.

SLFN11	1.52	0.3185631	0.40216548	0.4049523	0.45873611
TMEM40	1.5	0.32826628	0.51036714	0.5117909	0.30941292
CD82	1.49	0.35972071	0.33442061	0.40366089	0.40621867
PLS3	1.49	0.3229367	0.34005436	0.34457284	0.48148848
S1PR2	1.49	0.3423641	0.41732232	0.4297187	0.34589473
CNGA1	1.47	0.36087143	0.30268756	0.3439856	0.41816331
GBGT1	1.47	0.29896477	0.27472738	0.32655742	0.34544897
LINC01119	1.47	0.32590409	0.32313332	0.30823887	0.44407193
UBASH3B	1.47	0.33064569	0.50351616	0.50672061	0.47651249
CPNE2	1.46	0.25854073	0.37821989	0.36182205	0.46724815
ITM2C	1.45	0.33315638	0.36009527	0.38829978	0.44692083
TNFRSF21	1.44	0.37354559	0.39129357	0.4131146	0.29649844
VGLL1	1.44	0.55022635	0.41366376	0.46017947	0.43863358
PRKCA	1.43	0.31085413	0.28754719	0.32558788	0.5149627
CDH3	1.42	0.4713359	0.39747213	0.42859906	0.42911719
DAPK1	1.42	0.29679775	0.39276861	0.37457934	0.35957886
TANK	1.42	0.39319979	0.26641859	0.25653463	0.34116139
PDE9A	1.4	0.36354938	0.26818879	0.28849327	0.37773258
PROM1	1.4	0.54768491	0.34542338	0.41426198	0.45246869
RASGRF1	1.4	0.27943511	0.46133654	0.45592658	0.38596888
FGF11	1.39	0.34487488	0.36126929	0.37706393	0.34990183
PML	1.39	0.3152875	0.41943279	0.42400388	0.37146
FUT3	1.38	0.34714777	0.56836183	0.61584241	0.37021073
RGL1	1.38	0.27664114	0.30912579	0.34223565	0.47980191
CLDN1	1.37	0.36071605	0.33408402	0.35663264	0.38628133
ZIC1	1.37	0.36068547	0.25191748	0.2581401	0.26590453
ANXA8	1.36	0.40299644	0.27652996	0.29731956	0.3276508
GLYATL2	1.36	0.42108503	0.35801596	0.40193844	0.35570708
C3	1.35	0.32780549	0.29434066	0.32172147	0.44967143
KLK8	1.35	0.43381518	0.2723758	0.32360416	0.36040473
ZSWIM4	1.35	0.3565396	0.3859287	0.40848972	0.37868459
FOLR1	1.34	0.53917491	0.25282248	0.2977765	0.30968605
SLC2A12	1.34	0.44416831	0.26498901	0.29211573	0.40151459
ST8SIA1	1.34	0.47378648	0.40253379	0.40493804	0.43569316
NCCRP1	1.33	0.38567964	0.51033837	0.50272253	0.34919767
CPAMD8	1.32	0.3758494	0.25641578	0.32872077	0.33601938
MME	1.32	0.26976197	0.25687173	0.29312014	0.38879654
BACE2	1.3	0.36860889	0.34533493	0.37641969	0.34978981

Appendix 3. NRAD1-regulated genes identified via microarray similarly correlated with NRAD1, S100A8, S100A9, and ALDH1A3 in patient breast tumors.

Gene	Value			
MET	1.3	0.35593331	0.29490888	0.33118162	0.47255212
CXCL16	1.29	0.35817543	0.32915842	0.36952662	0.35195107
LY6D	1.29	0.40863329	0.44892288	0.47472041	0.31779381
MICAL3	1.29	0.34943404	0.28306839	0.3301168	0.40751641
EGFR	1.28	0.39287485	0.33265933	0.36886902	0.56681078
ANXA8	1.27	0.40299644	0.27652996	0.29731956	0.3276508
WNT6	1.27	0.42182617	0.33974225	0.38205	0.2945118
EML4	1.26	0.26254988	0.43972758	0.40224013	0.3275656
SCPEP1	1.25	0.3832792	0.4110825	0.43972133	0.44044824
KRT5	1.24	0.40058295	0.25233005	0.28137132	0.31806959
PSAT1	1.24	0.3610866	0.51950155	0.52078319	0.39046062
PSAT1	1.24	0.3610866	0.51950155	0.52078319	0.39046062
SERPINH1	1.24	0.25257024	0.43111185	0.4262534	0.30140958
SVOPL	1.24	0.39695201	0.25823221	0.29578215	0.26150685
TTLL4	1.23	0.37236724	0.35641641	0.38954715	0.31863771
MPZL2	1.22	0.46722793	0.36684257	0.40243556	0.44855577
RASGEF1B	1.21	0.33428276	0.40893526	0.39622526	0.38343163
TOX	1.21	0.32154247	0.29397241	0.30292162	0.41172062
MSN	1.2	0.31404688	0.44980435	0.42801799	0.50710467
PLA2G4A	1.2	0.40375745	0.3356747	0.36362004	0.48413769
TBC1D1	1.2	0.31030197	0.28280984	0.27857025	0.30353447
CRNKL1	-1.24	-0.282593	-0.2860194	-0.3129977	-0.2952764
RAD17	-1.26	-0.2592413	-0.3983766	-0.4170998	-0.2839327
GATA3	-1.27	-0.3441901	-0.4983911	-0.5015787	-0.4468738
TMEM161B	-1.27	-0.2557624	-0.4104213	-0.4296409	-0.2815154
KCTD6	-1.46	-0.2814171	-0.3533481	-0.3584504	-0.3625825
SLC39A6	-1.46	-0.3154434	-0.4419986	-0.4408514	-0.3330344
ZNF814	-1.5	-0.3004446	-0.4270611	-0.4332365	-0.3009686
GUF1	-1.56	-0.2653726	-0.272093	-0.3208678	-0.2584252
DNAJC12	-1.73	-0.308826	-0.4577438	-0.4842554	-0.3771835