June 2021

CircREV1 Expression in Triple-Negative Breast Cancer

Meagan P. Horton
University of South Florida

CircREV1 Expression in Triple-Negative Breast Cancer

by

Meagan P. Horton

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science
with a concentration in Cell and Molecular Biology
Department of Cell Biology, Microbiology, and Molecular Biology
College of Arts and Sciences
University of South Florida

Co-Major Professor: Margaret Park, Ph.D.
Co-Major Professor: Charles Chalfant, Ph.D.
Sandy Westerheide, Ph.D.
Florian Karreth, Ph.D.

Date of Approval:
June 7, 2021

Keywords: circular RNA, bioinformatics, alternative splicing, proliferation, triple negative breast cancer

DEDICATION

This thesis is dedicated to the people I cherish the most, and who have supported me throughout the entirety of my scientific endeavors. My mother, who has always raised me to be strong and independent by presenting herself first-hand as such. My brother, who could not be more opposite of me in career interest, yet has been by my side nevertheless. My late father, whose passing during my young teenage years fueled me to pursue the person I am today. Lastly, to the future of female STEM, may your passion never fade.

ACKNOWLEDGEMENTS

I would like to thank my outstanding mentor and principle investigator, Dr. Margaret Park, for her confidence in me as a young female scientist and her direct support and instruction as a hands-on adviser. Having joined her laboratory as an undergraduate volunteer, she presented with nothing but kindness and desire for me to learn and advance as a newly developing researcher. She inspired me to delve deeper into my comprehensive scientific understanding, in addition to captivating my love for wet-lab experimentation. I sincerely thank you for being nothing short of compassionate, encouraging, and ultimately genuine; for pushing me to strive for excellence and providing incredible leadership along the way.

My deepest gratitude extends to my mentor and former boss, Dr. Charles Chalfant, for his willingness to invest in my career as an emerging researcher and his persistence in urging me to be the dedicated scientist I am today. Allowing me to serve as a laboratory assistant in his laboratory while volunteering in Dr. Park's, instead of permitting me to endure a job in the service industry to maintain my undergraduate livelihood, he demonstrated his devotion in progressing me as an upcoming woman in STEM research. I thank you from the bottom of my heart for your generosity and assurance in me.

Furthermore, I would like to thank my committee members (current and former), Dr. Florian Karreth, Dr. Sandy Westerheide and Dr. Kristina Schmidt, for investing in my future and making time to listen and provide feedback. Without their guidance, I would have less of an understanding in molecular biology technique and experimentation, and limited perfection in scientific presentation skills.

Throughout my opportunities in the labs of Dr. Park and Dr. Chalfant, I often looked to my peers and senior scientists for advice. Without the help of my lab mates, Emily Mayo and Shaun Stevens, experimentation likely would have been much more challenging and stress-ridden. I deeply appreciate their guidance and mentorship. More so, I profoundly thank Dr. Christina Moss, who took kindly to me as an undergraduate student in the genetics lab she was my graduate TA for. Her certainty in me as a rising female scientist has been one of the top motivators in getting me to where I am today. I thank her for taking many hours establishing my “lab hands” as well as inspiring me to understand computational biology on levels I never wanted to.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Triple-negative breast cancer (TNBC) comprises only 24% of breast cancer cases, yet is the second leading cause of cancer mortality in women due to its aggressive nature (1). This increase in mortality is due to the lack of receptors for three targetable growth factors (HER2, progesterone, and estrogen receptors). Our previous studies have indicated that these cancers are highly dysregulated in respect to alternative splicing. Hence, we undertook a study aimed at identifying circular RNAs (circRNAs) generated from back-splicing events which were dysregulated in TNBC. We have identified a novel circRNA transcript, circular REV1 (circREV1), which is upregulated in our TNBC cell lines. Its overexpression may be an indicator of TNBC and its progression. The complexes formed between circRNAs and proteins or other RNA transcripts are able to dysregulate gene expression, which is one of the hallmarks of cancer (2).

Next generation sequencing of RNA collected from breast epithelium and TNBC cell lines were aligned with STAR aligner and bioinformatically analyzed for differential expression of circRNAs via circtools (3). CircREV1 was found to be overexpressed in two TNBC cell lines, while demonstrating minimal expression in breast epithelial cells. Quantitative polymerase chain reaction (qRT-PCR) analysis, following reverse transcription of RNA collected from relevant cell lines, validated these findings *in vitro*. Treatment with RNase R, to remove the linear construct of REV1, confirmed that the circular isoform is responsible for the overexpression found in the TNBC cell lines. The isoform we have honed in on is that of exon 3 back-spliced to exon 2, with the intron spanning the two being retained.

Linear REV1 is suggested to function as a scaffolding protein to recruit DNA polymerases for translesion synthesis in DNA repair mechanisms (4). It can be hypothesized that cancer may manipulate this gene to facilitate bypass of cell cycle regulation machinery. While minimal data has been published on the function of circREV1, it is predicted to be involved in a pathway which favors cancer advancement. Ultimately, our findings suggest that circREV1 upregulation may be influential in the transformation from breast epithelium to a more aggressive phenotype.

CHAPTER ONE: INTRODUCTION

Triple-Negative Breast Cancer (TNBC)

Breast cancer is a commonly occurring disease that affects many individuals, specifically women of all races and nationalities. While approximately 1 in 8 American women develop invasive breast cancer in their lifetime, it more rarely occurs in men with an expectation of 2,650 diagnoses in 2021 (5). Advancements in research and treatments have decreased death rates, in more recent years, in women over 50 years of age by 1% per year (5). However, breast cancer still remains to be the second leading cause of cancer deaths in women behind lung cancer (the leading cause of cancer deaths for men and women) (5). The severity of breast cancer increases when it is classified as Triple-Negative Breast Cancer (TNBC). This class of breast cancer accounts for around 10-15% of all breast cancers and is characterized by the lack of receptors for estrogen and progesterone, in addition to non-overexpression of the protein HER2 (6). TNBC tends to respond well to initial treatment administration, yet cancer cells that have escaped treatment make for a very aggressive disease that grows and spreads with haste. Unfortunately, without the major receptors and HER2 overexpression, there are limited options for targeted treatment, as the cancer tends to be resistant, making for a poor prognosis.

Survival statistics for different types of cancer are generated by data keeping, which is maintained by institutions, such as the National Cancer Institute (NCI). Specifically, the American Cancer Society relies on the SEER database from the NCI, which does not group cancers by stages, but rather groups them into classifications of them being localized (not spread outside the breast), regional (spread to nearby structures/lymph nodes), and distant (spread to distant structures) (6).

Five year survival rates for TNBC that are localized happens to be a promising 91%, however, it drops to 65% for regional diagnoses, and a mere 12% for distant diagnoses (6). Something to keep in mind is that these numbers are not static, and do not take all factors in to consideration – each tumor responds to treatment differently, each has their own grade, and each has a specified extent to which it has spread.

While invasive breast cancers have various treatment options, TNBC has fewer options due to the deficiency in receptors. These shortcomings mean that hormone therapies and drugs targeting such are inadequate for an efficient treatment regime (6). Because surgery is an option for those who have localized tumors, this lends to the higher survival rates seen for these classifications. Oftentimes, chemotherapy can be administered prior to surgery in order to shrink tumors on the larger end of the spectrum, and it can also be given after surgery in order to safeguard and reduce chances of the tumor returning (6). Radiation is another option, but it does depend on specific aspects of the tumor, such as how regional metastases are and what actions have been taken prior to radiation (24). For more advanced diagnoses, other molecular mechanisms must be employed, such as PARP inhibitors, progressive chemotherapies, and immunotherapy (6).

Alternative Splicing

The idea that humans create a multitude of different proteins using a set number of genes is only possible due to the occurrence of alternative splicing. DNA contains regions that are protein coding, called exons, and they are interspersed with regions that do not code for proteins, called introns. Alternative splicing is the “cutting” of precursor mRNA, in which intronic sequences are spliced out and exonic sequences are joined, producing different transcripts that go on to be translated into different proteins. The precursor mRNA for one specific gene can give rise to many

different proteins due to the inclusion or exclusion of different exons, promoting protein diversity. When exons are ligated in the order in which they appear in a gene, this is called constitutive splicing, while alternative splicing occurs when there is a diversion in such splicing, such as a skipping of exons (7). Ultimately, alternative splicing is the explanation for the inconsistency observed when it is seen that humans have approximately 25,000 protein coding genes, yet more than 90,000 different possible proteins generated (7). The idea that alternative splicing yields more proteins than the number of genes present in the human genome is visualized in Figure 1.

Non-coding RNA (ncRNA)

Non-coding RNA stems from transcribed messenger RNA from the DNA template, yet it does not go on to be translated into protein. This provides reasoning for the term "non-coding," as it does not code for functional proteins. However, just because these RNAs do not encode for functional proteins, it does not mean that they are not functional. These RNAs make up regulatory factors involved in many cellular functions, such as transfer RNAs (tRNAs) in translation and small nuclear RNAs (snRNAs) involved in splicing (8,19). In more recent years, many regulatory non-coding RNAs have been discovered, such as piwi-associated RNAs, endogenous short-interfering RNAs and microRNAs (miRNAs); these act as major regulators of gene expression through a multitude of different pathways (8,19). The discovery of ncRNAs has been pivotal in the field of gene research. Small interfering RNAs (siRNAs) can be utilized to study gene expression due to their silencing abilities, and even extends further to aspects of disease development where the roles of specific ncRNAs can be determined by the use of factors like specifically designed siRNAs. NcRNAs must hold a significant amount of importance considering they account for approximately 95% of the total RNA transcribed from the genome (9).

Circular Non-coding RNA

When considering alternative splicing and ncRNAs, one of the current areas of focus are circular RNAs (circRNAs). CircRNAs are formed from an alternative splicing event called back-splicing, where an upstream 3' splice site is joined to a downstream 5' splice site, as depicted in Figure 2 (21). Though circRNAs were initially discovered in 1976 and considered conserved byproducts of splicing in low abundance and unknown function, biotechnology has allowed for the identification of a large number of these molecules that tend to be expressed in a tissue- and developmental-specific manner (9). Not much is known about their functions, but those that have been characterized seem to act as miRNA sponges, preventing specific mRNA translation. Further research has demonstrated that circRNAs may heavily influence gene expression via transcription and RNA-binding protein (RBP) interactions. In addition to the fact that circRNAs are quite stable due to their non-exposed 3' and 5' ends, their suspected roles make them an attractive target in disease prevention and progression interference.

Circular RNAs in Cancer

CircRNA expression has shown correlation with many human diseases. Specifically, they have been associated with autophagy, apoptosis, cell cycle, and proliferation, while further studies have shown them to hold regulatory functions in diseases such as Parkinson's and Alzheimer's (neurological), cardiovascular implications, and a number of cancerous tumors (9). For breast cancer itself, one circRNA, circFoxo3, has been extensively studied, revealing its non-coding roles in the inhibition of tumor growth and angiogenesis (9, 10). Upon inducing expression of circFoxo3, TNBC cell line MDA231 displayed slowed proliferation and lessened cell survival when undergoing H_2O_2 treatment (10). These researchers also undertook mice studies to understand the

tumor formation with induced circFoxo3 expression and found that much smaller tumors formed in mice injected with cells having upregulated expression (10). This study further displayed the possibility of circRNAs acting as miRNA sponges by confirming circFoxo3's interactions with the miRNAs found to bind the Foxo3 host gene, as the host gene was more readily expressed due to the lessened silencing by the miRNAs (10). It is clear that circRNAs function for regulatory purposes, which cancer can exploit to its advantage, especially in terms of growth.

Clinical Relevance

Being the second leading cause of cancer deaths in the country, breast cancer has a high incidence affecting nearly 1 in 8, or 13%, of the female U.S. population. Of these cases, 10-15% of them are TNBC, leaving these individuals susceptible to a fast-progressing disease with an ill-perfected treatment plan. Because TNBC is so aggressive and fast growing in nature, in addition to its limited treatment options and resistance to those available, it is highly important to determine a more effective and targeted treatment regime for this subclass of breast cancer.

Hypothesis

Circular RNAs are a ncRNA molecule that hold regulatory roles in breast cancer and go on to interact in manners that promote or inhibit cancer progression. The upregulation of circREV1 in TNBC cell lines compared to immortalized breast epithelial cells promotes the cancerous nature of TNBC.

Research Aims

1) Identify circRNAs with dysregulated expression in triple-negative breast cancer via use of bioinformatic analysis.
2) Determine if circREV1 expression affects cancer-related phenotypic changes.

Figures

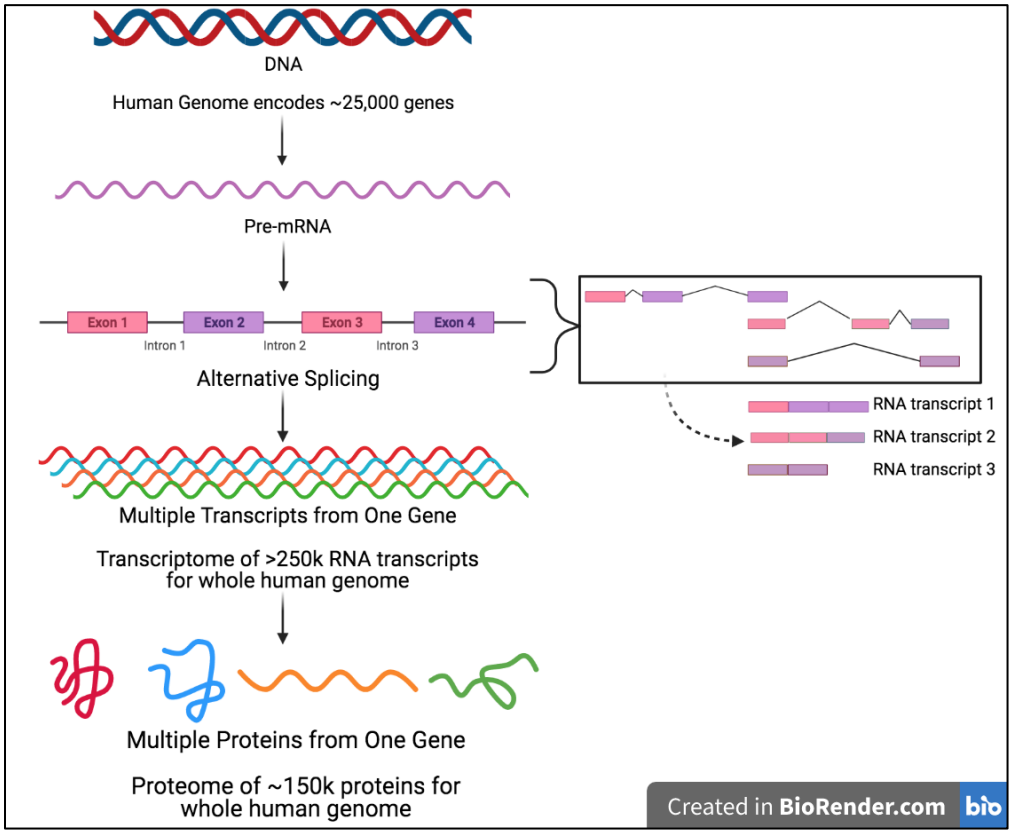


Figure 1: Alternative Splicing Yields RNA and Protein Diversity
Alternative splicing produces RNA diversity and permits the production of nearly 150,000 proteins from only ~25,000 genes in the human genome.

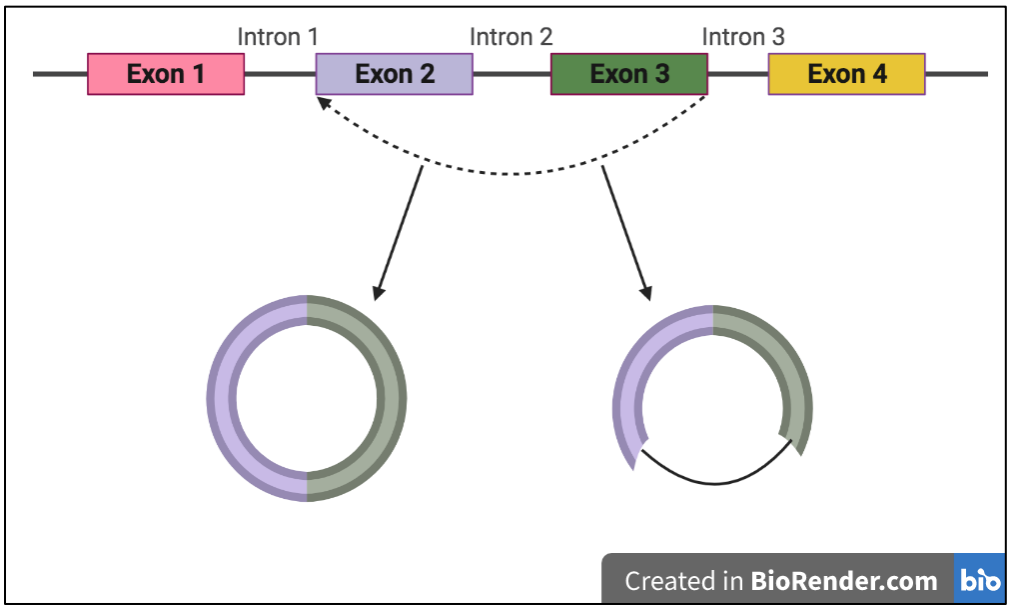


Figure 2: CircRNA Production via Back-Splicing
Back-splicing generates circRNAs by joining an upstream 3' splice site to a downstream 5' splice site.

CHAPTER TWO: BIOINFORMATICS

Abstract

Breast cancer is the second leading cause of cancer deaths in women. Not only is it expected to be diagnosed in over 280,000 U.S. women in 2021, it is also expected to affect over 2,600 men this same year (5). Patients with TNBC face hardships when it comes to the available treatment options, and due to its lack in identified targets and tenacity in progression, it has a poor prognosis. Therefore, developing biotechnological methods are commonly employed to identify molecular drivers that can serve as potential therapeutic targets. Bioinformatic analysis of total RNA sequencing of TNBC samples serves as a starting point for nucleic acid targeting. Rising attention on circRNAs presents bioinformatics as a major determinizing tool for unidentified circRNAs that might be playing a larger role in the drive of TNBC than realized.

Introduction

Over 30% of newly diagnosed cancers in women will be breast cancers in 2021, and for the same year 43,000 women in the U.S. alone are expected to pass away from it (5). While studies for the causes and predispositions for breast cancers have allowed for physicians to better recognize what permits these cancers to arise and progress, the treatment options are only so plentiful. The number of effective treatment options critically declines when the classification of the breast cancer happens to fall in the TNBC category, hence the emphasis for the identification of unknown molecular drivers.

In general, the field of bioinformatics involves the joining of molecular biology and genetics with computer sciences focusing on mathematics and statistics (11). Bioinformatics allows one to analyze large data sets that require intensive computational power to make the best possible predictions for such data. Overall, it can involve collecting computational data from biological data, building a computational model, solving problems with the modeling, and testing/evaluating the computational algorithm (11). Most useful for potential target identification is sequence analysis to gain an idea as to the ultimate function of genes and why they are expressed in certain patterns.

Sequence analysis for the goal of obtaining RNA targets involves a series of steps for optimal predictions. Transcriptomics focus on the complete set of RNA transcripts at a specific time under specific conditions (20). For transcriptome analysis, RNA is harvested and selected for in a manner that permits the non-polyadenylated transcripts to be present – in other words, selection of protein coding RNAs containing a poly(A) tail is not the route to take. Instead, rRNA is depleted from the total RNA samples, leaving behind ncRNAs and regulatory RNAs (12, 20). At this point, cDNA is generated from the available RNA to be sequenced, which then goes on to be analyzed bioinformatically. The sequenced material is aligned to a reference genome, then the present transcripts are assembled for desired analysis.

Analysis of transcriptome-level sequencing includes identification of differential gene expression and exon usage, alternative splicing, and even circular RNAs. Considering the canonical splicing methods, where introns are spliced out and exons are joined, only one protein per gene would be possible. However, there are alternative splicing methods that allow for thousands of proteins to be produced per gene, meaning there are thousands of RNAs made for one gene. Conclusively, alternative splicing creates a highly diverse set of RNAs, and they can be

characterized further when considering the different mechanisms of alternative splicing. One category of alternative splicing is back-splicing for the purpose of circRNA production. When an upstream 3' end is spliced to a downstream 5' end, this is called back-splicing. This kind of splicing allows for an exon that occurs later in the gene to be joined to an exon earlier in the gene, and often results in the retention of intronic sequence spanning such exons, as displayed in Figure 2. Exploitation of this alternative splicing pattern presents the potential identification of dysregulated expression of these RNA species. We hypothesize that circRNAs have dysregulated expression in TNBC cells when compared to immortalized breast epithelial cells, playing key roles in tumor development and progression.

Materials and Methods

Cell Lines and Media

There were three TNBC cell lines used: MDA231, MDA468, and BT549. The immortalized breast epithelium utilized was MCF10A. These cells were purchased from American Type Culture Collection (ATCC) and cultured in either RPMI (TNBC lines; Gibco) or DMEM/F12 (immortalized breast epithelium; Lonza) media. The RPMI media has additives of 10% fetal bovine serum (FBS: Gibco) and 1% penicillin/streptomysin (Gibco), while the DMEM/F12 media was supplemented with 5% Horse Serum (Gibco) and fresh add-ins of 1000X epidermal growth factor, hydrocortisone, and insulin. Cell lines were maintained in an incubator at 37°C with 5% CO_2 at less than 90% confluency. Any experiments were performed within the first 15 passages. Prior to harvesting cells for experiments using the pellets, they were rested in serum- and antibiotic-free RPMI for 24 hours, thus reducing the differential effects of serum growth factors on expression.

RNA Isolation and Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR)

Total RNA isolation was completed using the *Quick*-RNA Miniprep Kit (Zymo Research) and purified when necessary using the RNA Clean & Concentrator (Zymo Research) following manufacturer's protocol. RT-PCR was performed using a high-capacity cDNA reverse transcription kit following the manufacturer's protocol with indication for use of RNase Inhibitor (Applied Biosystems – Thermo Fisher Scientific).

Sequencing

MDA231 and BT549 TNBC RNA, as well as MCF10A breast epithelial RNA were sent to the DNA Sequencing center at Brigham Young University for deep RNA-Seq. Prior to sequencing, library preparation includes use of Ribo-Zero Gold (Illumina) to remove rRNA. Illumina HiSeq 2500 was used for 2 x 150 base pair paired-end sequencing to a depth of 100 million reads per sample at minimum.

Bioinformatic Analysis

Before beginning analysis of the sequencing, raw read files were checked for quality using FastQC (13). With Hg38 as the reference genome (GENCODE 27, Ensembl 91), sequencing was aligned using STAR (chosen for the chimeric junction output) and assembled with Stringtie (14, 15, 16, 17, 22). Following alignment and assembly, circtools was then employed to produce raw counts for detected circRNAs through exploitation of reads covering back-splice junctions with subsequent analysis to limit predictions. Targets were ultimately identified using a q-value of <.05.

Quantitative Real Time Polymerase Chain Reaction (qRT-PCR)

Custom primers designed via circtools, and in table 2, were used for qRT-PCR. PowerUp SYBR green master mix was utilized according to the guidelines in the user manual (ThermoFisher Scientific). QRT-PCR was run with cDNA generated as described in the RT-PCR section (after it was treated with RNase R to remove linear constructs), generally without the melting curve. Due to variations in β actin levels across the utilized cell lines, another housekeeping gene was selected, which was MRPL19 (Supplemental Table 1). Relative mRNA expression was normalized to MRPL19 using the ΔΔCT method.

Results

The workflow shown in Figure 3 was used to generate circRNAs targets from TNBC-derived RNA-Seq data. Two TNBC cell lines, MDA231 and BT549, and one immortalized basal epithelial cell line, MCF10A, were used in the bioinformatic prediction study. These three cell lines, in addition to another TNBC cell line, MDA468, were utilized for *in vitro* expression validation. Mammary basal epithelial cells, MCF10A, were used as the control non-cancerous cell line that express breast-specific antigens. MDA231, BT549, and MDA468 cell lines represent TNBC cell lines, each with their own specifications. MDA231 cells are adenocarcinoma-derived mammary epithelial cells with multiple mutations yielding them nearly triploid in chromosome count with two deletions and 11 stable rearrangements. BT549 cells are ductal carcinoma-derived mammary epithelial cells with multiple mutations yielding them a hypertriploid chromosome count with abnormal X chromosomes, 5 chromosomal under-representations, 3 chromosomal over-representations, and the presence of 4 marker chromosomes. MDA468 cells are adenocarcinoma-derived mammary epithelial cells with multiple mutations yielding them a

hypotriploid chromosome count with 7 under-representations due the presence of 19 marker chromosomes.

CircRNA Target Prediction via Circtools

Utilizing the general parameters suggested for circtools analysis and a q-value of <.05, there were 45 target predictions produced, as depicted in Table 1 (18). Of these 45 targets, seven were manually picked for further investigation and are represented in Figure 4A-G. In addition to literature on the host genes of the predicted targets, selection was based on a difference in $\log_2$ fold change (approximately 3-fold change or more) between the mammary basal epithelial cells and the two sequenced TNBC cell lines, MDA231 and BT549; if expression levels do not present this kind of $\log_2$ fold change, they were selected due to an oddity like having no expression at all in one line, or having heightened expression in one TNBC line and not the other while having similar expression to one or the other in the immortalized mammary basal epithelium. Circtools-generated graphs on the expression of these selected circRNA targets are represented in Figures 4A-G.

Validation of Selected CircRNAs

The predicted expression patterns of the seven hand-picked circRNA targets were validated prior to pursuing any cancer-related phenotypic effects. Circtools has an option to generate primers to directly detect the target in question. The primers utilized for qRT-PCR detection are depicted in Table 2. It is important to note that RNA utilized for cDNA products were first treated with RNase R to eliminate all linear constructs in order to ensure that any detection was due to only circular isoforms. Expression levels *in vitro* are presented in Figure 5A-G.

Discussion

Workflow & Target Prediction

Following the workflow presented in Figure 3, we were able to successfully utilize bioinformatics to identify circRNA targets in TNBC for further investigation. As Figure 3 depicts, we were able to take RNA-Seq data from TNBC and non-tumorigenic cell lines and align the sequencing to the Hg38 genome using STAR Aligner, which was able to generate an output with chimeric junction reads. RNA-Seq alignment output files were then be employed to assemble potential transcripts in a bioinformatic program called Stringtie. Assembled transcripts were analyzed in circtools to predict which circRNAs were present. This was possible due to the chimeric junction read output via STAR Aligner. Circtools generated a table of potential targets, which was further limited with desired specifications. We asserted a q-value of <.05, allowing us to cap the predictions to 45 targets presented in Table 1. Of these 45 targets, seven were selected, as depicted in Figure 4A-G, based on a difference in $\log_2$ fold change (approximately 3-fold change or more) between the immortalized mammary basal mammary epithelial cells and the two sequenced TNBC cell lines, MDA231 and BT549. Expression levels that did not present this kind of $\log_2$ fold change were selected for further exploration due to an oddity, such as having no expression in one line at all, or having heightened expression in one TNBC line and not the other while having similar expression to one or the other in the immortalized mammary basal epithelium.

Targets & Expected Expression

In pursuing the seven selected targets for further investigation, we begin with validating their expected expression levels prior to moving forward with experiments relative to them. Expected expression levels, generated via circtools analysis, are presented in Figure 4A-G. In order

to validate these expected expression levels, we utilize primers generated in circtools, listed in Table 2, via qRT-PCR using PowerUp SYBR green master mix. Expression levels in their respective cell lines are presented in Figure 5A-G.

Validation & Rejection of Target Expression

One circRNA was confirmed to express similarly to the circtools predictions, while the other six did not. CircREV1 expression levels were validated to agree with the expected expression levels from circtools, as seen in comparison of Figure 4A and Figure 5A. Expression levels for this circRNA were checked multiple times with numerous cell pellets. CircSETD3 qRT-PCR analysis did not agree with the circtools prediction, as seen in comparison between Figure 4B, where TNBC cell lines show upregulated expression, and Figure 5B, where expression levels in the TNBC lines show downregulation. CircZNF124 qRT-PCR analysis had large error bars in the mammary basal epithelium samples, however, it did show the expected downregulation of expression in the MDA231 TNBC cell line when compared to the non-cancerous line. Essentially, the first confrontation of expression levels for this circRNA did agree when comparing Figures 4C and 5C, however, they were not checked multiple times with numerous cell pellets and this target was not further pursued for experiments. CircPARD3 qRT-PCR analysis did not agree with the circtools prediction, as seen in comparison between Figure 4D, where TNBC cell lines show upregulated expression, and Figure 5D, where expression levels in the TNBC lines show downregulation. CircBARD1 qRT-PCR analysis did not agree with the circtools prediction, as seen in comparison between Figure 4E, where TNBC cell lines show upregulated expression, and Figure 5E, where expression levels in the TNBC lines show downregulation. CircLPAR1 qRT-PCR had large error bars and skewed expression levels for both the MDA231 TNBC cell line as

well as the non-cancerous MCF10A cell line. Taking these facts aside, when comparing the expected expression levels in Figure 4F to the *in vitro* expression levels in Figure 5F, they did not agree. The circtools expectations presented the non-cancerous MCF10A cell line as having absolutely no expression, while this was not the case in our *in vivo* exploration. CircMGA qRT-PCR analysis did not agree with the circtools prediction, as seen in comparison between Figure 4G, where TNBC cell lines show upregulated expression, and Figure 5G, where expression levels in the TNBC lines show downregulation.

Figures

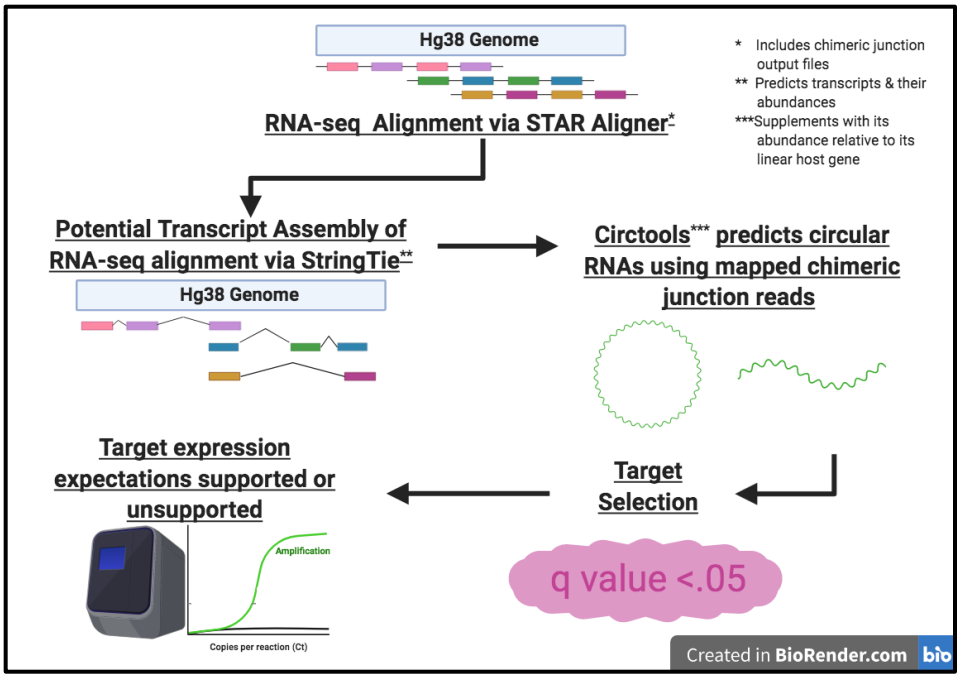


Figure 3: Target Identification Workflow via Bioinformatics
Bioinformatics-guided target selection of circRNAs detected within RNA-Seq of basal mammary epithelial cells (MCF10As) and Triple-Negative Breast Cancer cell lines (MDA231 & BT549).

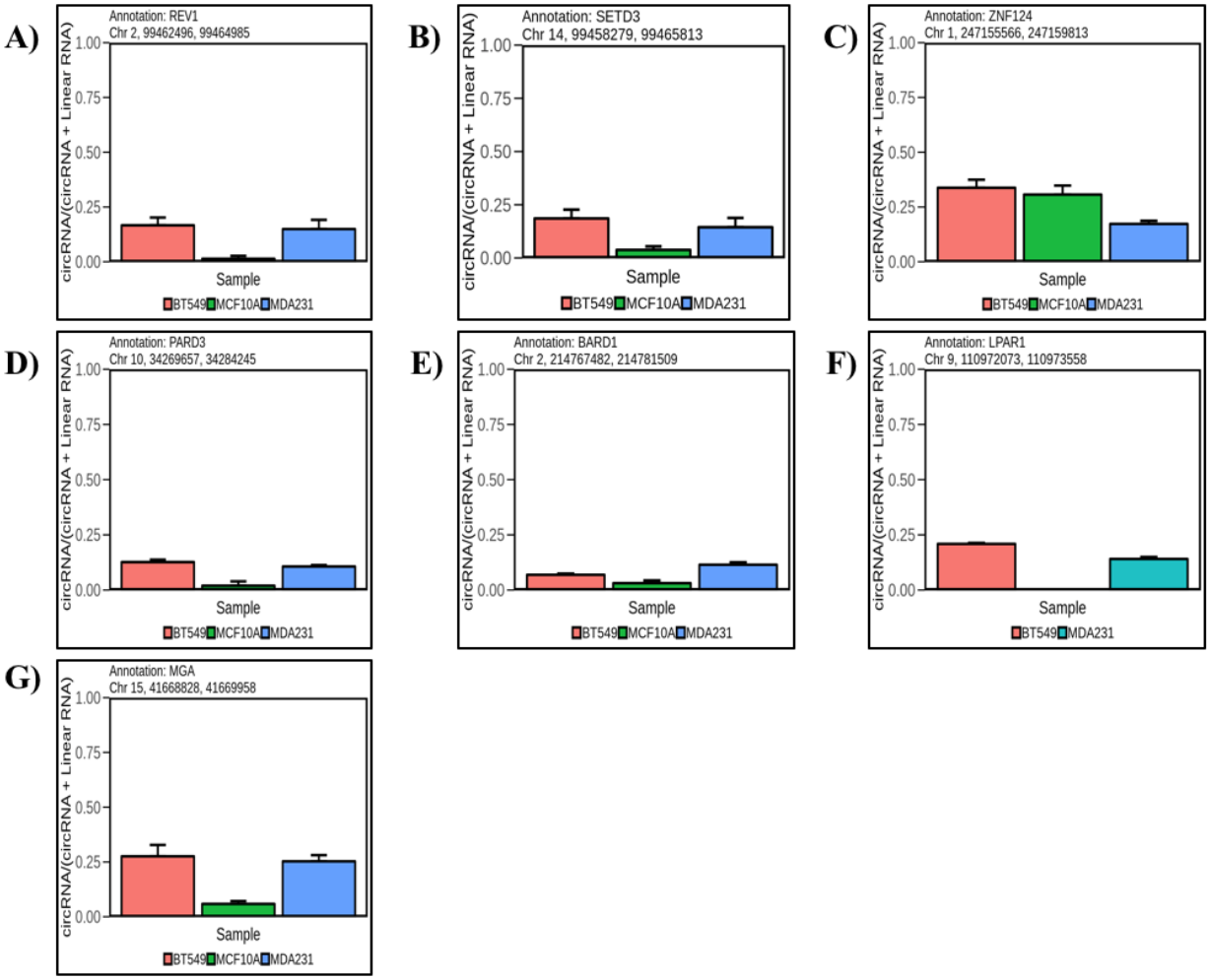


Figure 4: Circtools-generated Expression Expectations for Selected Targets
Circtools-generated graphs of the selected targets from predicted targets of circRNAs. Names of the linear host genes are as follows: A) REV1, B) SETD3, C) ZNF124, D) PARD3, E) BARD1, F) LPAR1, G) MGA.

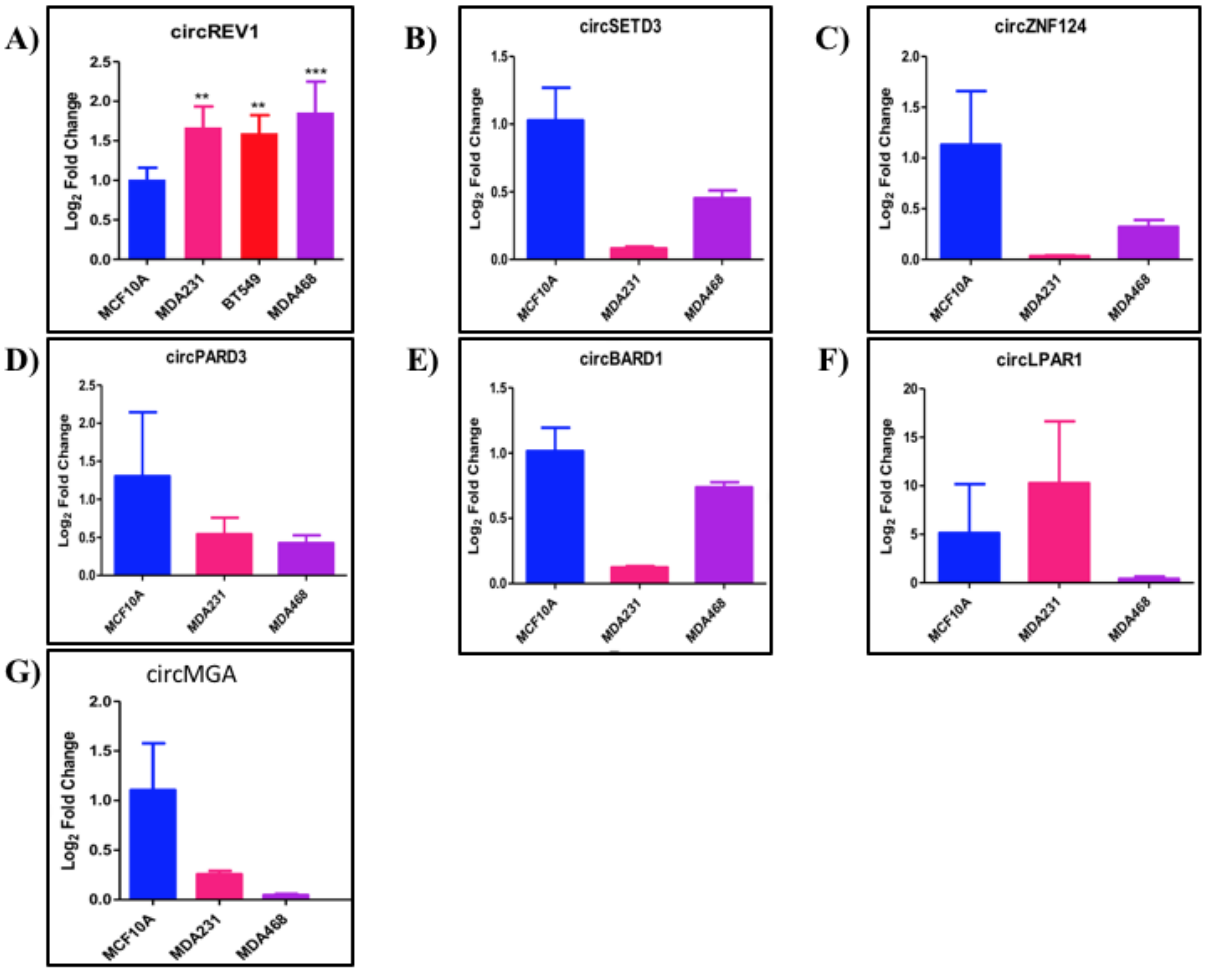


Figure 5: *In vitro* Expression Levels via qRT-PCR for Selected Targets
QRT-PCR analysis of *in vitro* expression levels for the selected circRNA targets. The represented circRNAs targeted are as follows: A) circREV1, B) circSETD3, C) circZNF124, D) circPARD3, E) circBARD1, F) circLPAR1, G) circMGA. Targets in B-G were primarily checked in cell lines MCF10A, MDA231 and MDA468, as BT549 cells were not cultured at that point in time, while target A was validated in all four cell lines, MCF10A, MDA231, BT549, and MDA468.

Tables

Table 1: Circtools-generated Targets
Circtools-generated targets using program-specified parameters and q-value of <.05 following workflow presented in Figure 3.

Host Gene Row 1	q-value Row 1	BT549 Row 1	MCF10A Row 1	MDA231 Row 1	Host Gene Row 2	q-value Row 2	BT549 Row 2	MCF10A Row 2	MDA231 Row 2
TNFRSF21	0.000190195	0.39844083	0.06786965	0.14481646	ZBTB44	0.007280343	0.30660417	0.1347398	0.14407814
PTK2	0.000190195	0.09479568	0.01267914	0.02094887	ARHGAP5	0.007280343	0.12751444	0.06092533	0.04769452
MAN1A2	0.000391381	0.29640272	0.12754065	0.17365152	ARHGAP10	0.007280343	0.21076321	0.07291667	0.07860499
PLOD2	0.000391381	0.10341904	0.01813626	0.04147724	MGA	0.007418638	0.27587383	0.05844156	0.25286174
RSRC1	0.000391381	0.14628255	0.03036737	0.04558425	ZNF91	0.011001426	0.13176582	0.06696231	0.28392952
FAT1	0.000391381	0.17145874	0.05302818	0.11215483	ERC1	0.012511379	0.22536748	0.0470696	0.19183502
PCMTD1	0.000391381	0.29327137	0.05674366	0.24311491	DNAH14	0.013697464	0.23330472	0.05910326	0.16604349
MTCL1	0.000785782	0.21736427	0.02777778	0.14751594	NAB1	0.013697464	0.1781874	0.04611249	0.15429621
RBM33	0.000785782	0.29703864	0.11400953	0.16431986	ASXL1	0.013697464	0.08470032	0.1069435	0.13920749
MAN1A2	0.003051733	0.10076663	0.02301934	0.08978462	CCDC66	0.013697464	0.26446721	0.03571429	0.14791351
BACH1	0.00463477	0.19083225	0.07855951	0.06859278	CAMSAP1	0.013845413	0.37851128	0.20574855	0.34341724
FNDC3B	0.00463477	0.10766478	0.02661114	0.02608805	LPAR1	0.016355587	0.20887646	0.5	0.14062875
MAN1A2	0.005147512	0.16713767	0.06355811	0.10398215	ATXN2	0.018079966	0.1827511	0.01162791	0.14650338
MED13L	0.005147512	0.3680761	0.09493088	0.11548034	CSNK1G3	0.019448543	0.25378761	0.09152971	0.15254019
CEP192	0.005147512	0.20491186	0.06887755	0.10375817	MTHFD2L	0.024817225	0.41026417	0.26293103	0.21800826
CDYL	0.005147512	0.38518623	0.17729041	0.29997794	BARD1	0.028746182	0.06882688	0.03099839	0.1145285
PTK2	0.005147512	0.08670798	0.03851365	0.04925596	PARD3	0.029853812	0.12644221	0.01960784	0.10646766
ASH1L	0.005558759	0.17285281	0.03670635	0.2215841	AFAP1	0.03017706	0.10556496	0.04416283	0.16708542
MIB1	0.006188092	0.11521758	0.02465278	0.11416822	REV1	0.030918158	0.16642755	0.01351351	0.14922003
USP25	0.00632079	0.20119512	0.05301902	0.07980899	KANSL1	0.033932062	0.22435897	0.05698854	0.17715054
ZNF124	0.036395105	0.33767806	0.3065661	0.17272727	LMBR1	0.03397651	0.10406	0.01932489	0.04507567
ZKSCAN1	0.040077292	0.14838207	0.06399195	0.15224084	SMARCA5	0.036116945	0.15018878	0.07753356	0.09543409
SETD3	0.04600997	0.18560119	0.03768939	0.14415712					

Table 2: Circtools-generated Primers for Selected Targets
Circtools-generated primers for detection of the specific target prediction. These were utilized in qRT-PCR for primary validation of expression patterns predicted in circtools.

Target	Forward Sequence	Reverse Sequence
circREV1	5'-CAT TTT CAG CTC GCT TCC TC-3'	5'-TCG ATC AGA TGC TGC TAT GC-3'
circSETD3	5'-CAC CAG TGC CAG ATT TCT GA-3'	5'-AAC ACA GCT CGA CAG TAC GC-3'
circZNF124	5'-TCA CAG CCA CAT CCT CAA AG-3'	5'-GAT GGG GTT TCA CCG TGT TA-3'
circPARD3	5'-CAT CTT TTC GAT GTT TGC CA-3'	5'-ACG AGA AGG GCA TAT GAT GG-3'
circBARD1	5'-TTC GAG GGC TAA ACC ACA TT-3'	5'-TAT GGA GCC TCC AGA AAT GC-3'
circLPAR1	5'-GTG GAT GGG GAG CTT CAT AA-3'	5'-TCT CGG CAT AGT TCT GGA CC-3'
circMGA	5'-AGA TAG GTG GAT GGG GAG CT-3'	5'-TTG TCT CCC GTA GTT CTG GG-3'

CHAPTER THREE: CIRCREV1 IN TNBC

Abstract

Different forms of alternative splicing yield various RNA species that may then be translated into protein, or may remain in RNA form and considered one of many ncRNAs. Back-splicing is an alternative splicing event where an upstream 3' splice site is joined to a downstream 5' splice site. This novel mechanism creates circRNAs where exons, or sometimes introns and exons, are joined in a reverse fashion where the nucleic acid backbone is circular with no free ends. CircRNAs have been found to play regulatory roles in human diseases, sometimes in manners allowing them to promote or inhibit cancer progression. We have found evidence that circREV1 is upregulated in TNBC cell lines when compared to immortalized mammary basal mammary epithelium. We further investigate the biological effects of circREV1 in order to determine if there is a cancer-related phenotype. Biological assays evaluating the phenotypic effects of knocking down the circREV1 transcript target we identified has revealed that inhibited expression of circREV1 in established TNBC lines does limit the proliferation and colony formation as well as its ability to close wounds. These findings suggest that circREV1 holds some kind of role in TNBC that allows it to grow more aggressively.

Introduction

Breast cancer is the most commonly diagnosed cancer amongst women in the U.S., as 1 in 8 women will develop an invasive form of it (5). Considering these cases, up to 15% of them will be TNBC (6). Because of TNBC's lack in targetable factors, such as HER2 and receptors for

progesterone and estrogen, this breast cancer subtype proves to be tenacious and difficult to treat. There is a large demand to determine new targets for treatment in this cancer, as the survival rates plummet with increasing distance in how far it spreads.

One field that presents a more promising outlook for targetable factors is that of ncRNA. These RNAs make up regulatory factors involved in many cellular functions and, in more recent years, have been discovered to act as major regulators of gene expression through a multitude of different pathways (8, 19). NcRNAs can be manipulated to characterize their effects on gene expression, and further offers aspects to better understand disease development. Considering they account for up to 95% of the total RNA transcribed from the genome, it is likely that ncRNAs hold significant roles yet to be defined (9).

Rising in ncRNA focus are circRNAs, especially in the aspect of human disease. Biotechnology allows for the identification of a large number of these molecules that tend to be expressed in a tissue- and developmental-specific manner (9). Little has been solidified in relation to their functions, but circRNAs that have been characterized act as miRNA sponges, preventing mRNA translation. They also present heavy influence on gene expression via transcription and RNA-binding protein (RBP) interactions (9). The suspected roles of these circRNAs make them a prospective target in disease prevention and progression interference.

CircRNAs are associated with autophagy, apoptosis, cell cycle, and proliferation, while other studies describe their regulatory functions in diseases such as Parkinson’s and Alzheimer’s (neurological), cardiovascular diseases, and a number of cancer typess (9). Through bioinformatic analysis, we identified a novel circRNA, circREV1, which is upregulated in TNBC cell lines when compared to immortalized mammary basal epithelium. CircREV1 is on the minus (-) strand of chromosome 2, with exon 3 back-spliced to exon 2 and contains the intron spanning the two. We

hypothesize that circREV1 may influence phenotypes which may affect cell proliferation and motility, two hallmarks of transformation.

Materials and Methods

Cancer Cell Lines and Media

There were three TNBC cell lines used: MDA231, MDA468, and BT549. The immortalized breast epithelium utilized was MCF10A. These cells were purchased from American Type Culture Collection (ATCC) and cultured in either RPMI (TNBC lines; Gibco) or DMEM/F12 (immortalized breast epithelium; Lonza) media. The RPMI media has additives of 10% fetal bovine serum (FBS: Gibco) and 1% penicillin/streptomysin (Gibco), while the DMEM/F12 media was supplemented with 5% Horse Serum (Gibco) and fresh add-ins of 1000X epidermal growth factor, hydrocortisone, and insulin. Cell lines were maintained in an incubator at 37°C with 5% CO_2 at less than 90% confluency. Any experiments were performed within the first 15 passages. Prior to harvesting cells for experiments using the pellets, they were rested in serum- and antibiotic-free RPMI for 24 hours, thus reducing the differential effects of serum growth factors on expression.

RNAi Design & Transfection

In order to silence the circREV1 transcript, siRNA was designed using Integrated DNA Technologies (IDT) to cover the back-splice junction. The sequence for this siRNA was: 3'GUUAAUGGAUACACAGAAGUU'5. For the negative control siRNA treatment, Negative Control No.1 from Ambion was used (AM4611). In the case of knocking down the linear construct of REV1, pre-designed and target-confirmed siRNA from Ambion was employed (AM16704).

CircREV1 was successfully knocked down without off-target effects at a concentration of 75 nm, while the negative control was generally administered at a concentration of 50nm. When trying to knockdown the linear construct, the cells do not handle this well, undergoing visual cell death at concentrations anywhere from 50-200nm with increasing effects at higher concentrations. Transfections were administered using Dharmafect 4 in MDA231 cells (and Dharmafect 1 in MCF10A when attempted) according to the manufacturer's protocol and suspended in complete medium. When harvesting cells, they were taken at 72 hours post-transfection.

RNA Isolation and Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR)

Total RNA isolation was completed using the *Quick*-RNA Miniprep Kit (Zymo Research) and purified when necessary using the RNA Clean & Concentrator (Zymo Research) following manufacturer's protocol. RT-PCR was performed using a high-capacity cDNA reverse transcription kit following the manufacturer's protocol with indication for use of RNase Inhibitor (Applied Biosystems – Thermo Fisher Scientific).

Quantitative Real Time Polymerase Chain Reaction (qRT-PCR)

Custom primers designed via circtools, and in Table 2 were used for qRT-PCR. PowerUp SYBR green master mix was utilized according to the guidelines in the user manual (ThermoFisher Scientific). QRT-PCR was run with cDNA generated as described in the RT-PCR section (after it was treated with RNase R to remove linear constructs), generally without the melting curve. Due to variations in β actin levels across the utilized cell lines, another housekeeping gene was selected, which was MRPL19. Relative mRNA expression was normalized to MRPL19 using the ΔΔCT method. Any time linear REV1 was assessed in qRT-PCR, IDT PrimeTime Primer

Hs.PT.58.14530085 was utilized, while the primer indicated in Supplemental Table 2 was employed for MRPL19.

Proliferation Assay

Cells were plated onto 96-well plates at $2.5x10^3$ cells per well. WST-1 assay (Sigma-Aldrich) was completed according to the manufacturer's protocol. WST-1 dye was added, incubated at 37°C with 5% CO_2 for 30-45 minutes, and absorbance was measured on a BioTek Synergy 2 plate reader on alternating days 0-6 at 440 nm.

Clonogenic Assay

Clonogenic assays were performed by plating 250 cells per well on a 6-well plate. Media was changed every 3-4 days and colonies were monitored to avoid overgrowth, while ensuring colonies had at least approximately 50 cells (as a cell colony is generally defined to consist of). Cells were fixed with 10% formalin 9-14 days after seeding and stained with crystal violet prior to counting of colonies. (23)

Scratch Assay

Scratch assays were seeded at a concentration of $3x10^5$ cells per well on a 24-well plate and settled to 100% confluency to create a monolayer of cells. Prior to inducing a scratch down the middle of the well, cells were serum-starved with media consisting of only 2% FBS for 3 hours. Upon serum-starvation, a scratch was carefully induced with a sterile P20 pipette tip, rinsed with PBS, and serum-starving media replaced. The plate was then placed in a Keyence Microscope

fitted for live cell imaging meeting the tissue culture conditions previously described. Images were taken every 5 minutes for 24 hours after setting points for images to be taken at for each well.

Actin Cytoskeleton Focal Adhesion Assay

Cells were plated on a 4-well chamber plate at a concentration of $1.25x10^5$ cells per well. Prior to fixation and staining, wells were scratched twice with a sterile P20 pipette tip, dividing the well into 3 sections, then rinsed with PBS. Fixation and staining were completed according to the manufacturer's protocol in the kit for the actin cytoskeleton / focal adhesion staining (Sigma-Aldrich).

Results

With validation of the circtools-generated predictions of expression levels in the relative cell lines, knockdown experiments were undertaken to determine how lack in expression of circREV1 affects TNBC cell lines. Cells were harvested 72 hours after the knockdowns to determine, via qRT-PCR, that knockdown was achieved; these qRT-PCR analyses are represented in Figure 6A-C.

Biological Assays

To determine the effects of circREV1 knockdown, biological assays were plated the morning following the afternoon of transfection. WST-1 proliferation assays to measure how quickly the cells replicate are represented in Figure 7A-C. We then evaluated the ability for colony formation on a plate via clonogenic assays, presented in Figure 8A-C. Further, we wished to assess cell motility via the scratch assay, represented in Figure 9. An attempt to view differences in cell

morphology, motility, and polarity in actin cytoskeleton / focal adhesion staining is portrayed in Figure 10. With approximately 4-fold change in circREV1 expression following circREV1 knockdown, there are significant changes in proliferation, clonogenics, and wound healing. Our results suggest that limited circREV1 expression lends to weakened proliferation and colony formation, as well as hindered ability to close a wound.

Discussion

CircREV1 Knockdown

After successful validation of the circtools-expected expression levels of circREV1*in vitro*, we began to determine the effects of knocking down its expression via siRNA. Figure 6A represents the qRT-PCR of the optimized knockdown in TNBC cell line, MDA231. It is shown that significant knockdown was achieved with limited off-target affects to the linear host gene. Using the same approach to accomplish knockdown in MDA231, circREV1 expression was limited in BT549, as seen in Figure 6B. It was also determined that nearly no off-target affects to the linear host gene were seen with circREV1 knockdown. With the same regime for circREV1 knockdown in MDA468, there was slightly more knockdown than seen in BT549, yet slightly more off-target effects than seen with BT549, depicted in Figure 6C. It was determined that 75 nm was the optimal concentration to knockdown circREV1 without significantly affecting the linear host gene.

Linear REV1 Knockdown

Linear knockdown was inconsistent and often lead to considerable cell death, making it hard to harvest cells for qRT-PCR analysis. Because of the complications seen with linear REV1

knockdown, it was decided that experiments involving their knockdown may be unreliable and were not further pursued. The minimal qPCR analysis that was able to be completed is represented in Unpursued Data Figure 7A-C. While these experiments were not further explored, their proliferation and clonogenic data collection for TNBC cell lines BT549 and MDA468 are included in their respective graphs (Figures 7B-C & 8B-C) due to being ran during same exact time as the relative experiments for the circREV1 knockdown in these cell lines. Linear knockdown was not initially taken into account when completing biological assays in the MDA231 TNBC cell line. It was determined that knocking down the linear host gene was too lethal for TNBC cells, as they may be relying on the REV1 protein to bypass abasic lesions during the cell cycle, as that has been suggested as its major role.

Proliferation After circREV1 Knockdown

Without circREV1 expression, proliferation of TNBC cells was significantly hindered. As seen in Figure 7A-C, absorbance at 440 nm was much less in cells with circREV1 knockdown for all TNBC cell lines, MDA231, BT549, and MDA468. It is likely that circREV1 is responsible for binding miRNAs or RBPs that are involved during the cell cycle by recognizing mishaps and mutations. With these suspected interactions, circREV1 binding to such would normally prohibit their interactions to step in and ensure the cell does not replicate and undergoes apoptosis, thus their knocked down expression would lend to lessened proliferation.

Linear REV1 knockdown was also read for TNBC cell lines BT549 and MDA468, shown in Figure 7B-C; it was not considered during the proliferation assay for MDA231. These figures do show that there was a significant reduction in proliferation for MDA468, however, the knockdown for linear REV1 appeared to be much more lethal for this cell line in particular.

However, the BT549 cell line indicated that the proliferation was able to bounce back. This could be due to the presence of different mutations in the different cell lines, making it possible for one cell to overcome the inconvenience and not the other, as cell signaling reliance seemed to be much heavier in the MDA468. In every attempt to knockdown linear REV1 for the MDA468 cell line, there was a visual abundance of cell death. Linear knockdown was inconsistent and often lead to considerable cell death, making it hard to harvest cells for qRT-PCR analysis. Because of the complications seen with linear REV1 knockdown, it was decided that experiments involving their knockdown may be unreliable and are therefore not further pursued.

Clonogenics After circREV1 Knockdown

Following knockdown of circREV1 expression, the ability of TNBC cells to form colonies from a single cell was significantly reduced. As seen in Figure 8A-B, the number of colony forming units was much less in cells with circREV1 knockdown for two TNBC cell lines, MDA231 and BT549. There was no significance found for the lesser number of colonies formed for the MDA468 cell line following circREV1 knockdown. While the knockdown was of significance in this cell line, this could be due to the nature of these cells, as they always present heavy reliance on signaling from nearby cells for growth in general tissue culture and possibly increase growth signaling due to the low abundance of cells present. The lessened colony formation following circREV1 knockdown in MDA231 and BT549 is likely due to the suspected responsibility of circREV1 binding miRNAs or RBPs involved during the cell cycle by recognizing mishaps and mutations. Suggesting these interactions means that circREV1 expression would normally promote binding to prohibit cell cycle precautions from stepping in and ensuring the cell does not

replicate and undergoes apoptosis, thus its knocked down expression would lend to limited colony formation

Linear REV1 knockdown was also considered for TNBC cell lines BT549 and MDA468, shown in Figure 8B-C; it was not considered during the clonogenic assay for MDA231. These figures show that there was not a significant difference in the number of colony forming units for these two cell lines. In general, these two cell lines do not handle the knockdown of linear REV1 very well, especially the MDA468 cell line as it displays a visual abundance of cell death. Because of this, it is likely that these cells only began forming colonies with the few cells that retained life toward the end of this assay once the effective knockdown of linear REV1 began to wear off. Linear knockdown was inconsistent and often lead to considerable cell death, making it hard to harvest cells for qRT-PCR analysis. Because of the complications seen with linear REV1 knockdown, it was decided that experiments involving their knockdown may be unreliable and are therefore not further pursued.

Scratch Assay After circREV1 Knockdown

When considering how well TNBC cells close a wound after circREV1 knockdown, we saw MDA231 TNBC cells close 27% less of the wound than the same cells having non-targeting siRNA treatments. While each set of cells worked to close the wound over the 24 hour time-lapse, much less closure was achieved with those enduring circREV1 knockdown, as seen in Figure 9. Cell migration is known to play important roles in physiological and pathological development and extends further to angiogenesis and tumor metastasis. The MDA231 TNBC line is said to be the most aggressive of the three TNBC cell lines utilized, therefore one of the major goals of this cell line is going to be inhabitance of “uncharted territory.” We likely see the reduced closure of

the wound with circREV1 knockdown due to the stalling in cell cycle processes and confirmed hindrance in proliferation. It has been noted in other studies that cell proliferation may bias this kind of experiment due to its influence on scratch closure (25). When attempting this assay with BT549 and MDA468 cells, neither cell line displayed a directed attempt to close the wound, they rather grew without direction and cells moved around nonsensically.

Actin Cytoskeleton / Focal Adhesion Assay After circREV1 Knockdown

Figure 10 displays the actin cytoskeleton / focal adhesion assay with and without circREV1 knockdown. Unfortunately, the anti-vinculin supplied with the Sigma-Aldrich kit did not stain these cells in a manner that emits visible fluorescence, no matter how high or low the exposure concentration was. For MDA231, BT549, and MCF10A cell lines, there were not obvious differences seen in cell morphology, as displayed in Figure 10 with the respective magnifications denoted. The lack in visible differences may be due to unsuccessful anti-vinculin staining, but more so could be due to the idea that circREV1 influences must not change cell morphology, motility, or polarity.

Conclusions

Overall, circREV1 has been seen to play roles in proliferation, ability to form colonies, and ability to close a wound in TNBC cells. Without circREV1 expression, all three TNBC cell lines see a significant reduction in their proliferation over six days. For clonogenic assays, two of the three TNBC cell lines display a significant decline in the number of colony forming units over a 9-14 day period. It is also seen that with muted circREV1 expression, MDA231 cells have limited ability to close a wound like they do with normal expression levels. Ultimately, circREV1 may be utilized as a means to promote TNBC cell cycle turnover or allowing the bypass of apoptotic

machinery in cell cycle checkpoints. It is shown that circREV1 plays an important role in TNBC's viability and ability to grow at the rates it does.

Figures

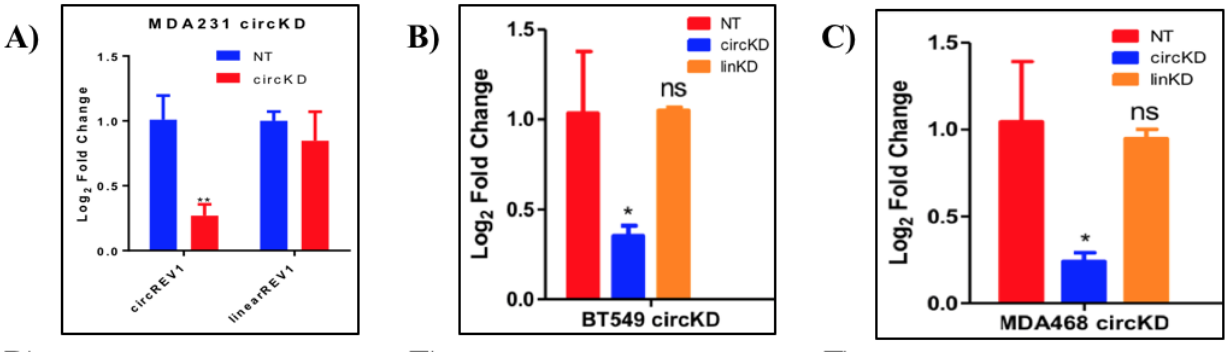


Figure 6: Knockdown of circREV1 via qRT-PCR
qRT-PCR representing knockdown of circREV1, optimized to limit off-target affects to the linear isoform, in: A) MDA231, B) BT549, and C) MDA468.

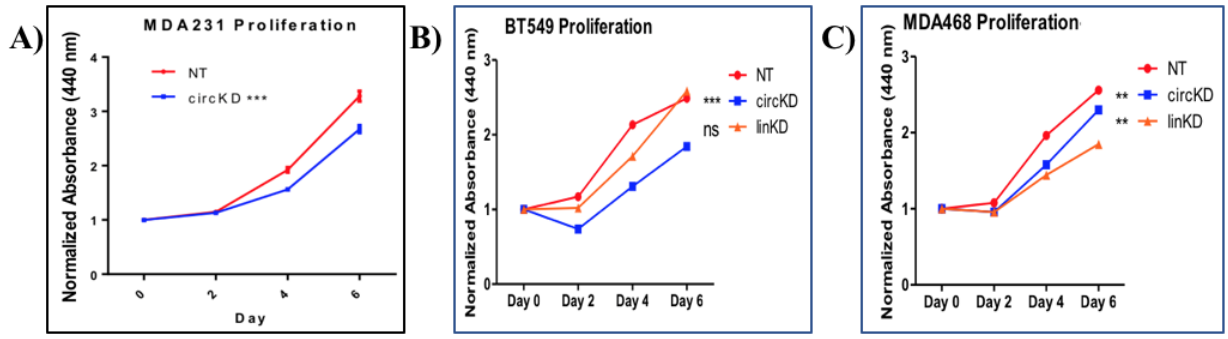


Figure 7: Proliferation Assays
WST-1 proliferation assay following circREV1 knockdown in: A) MDA231, B) BT549, and C) MDA468. No linear knockdown is represented in the graph for MDA231 due to it not being considered during the run for their biological assays.

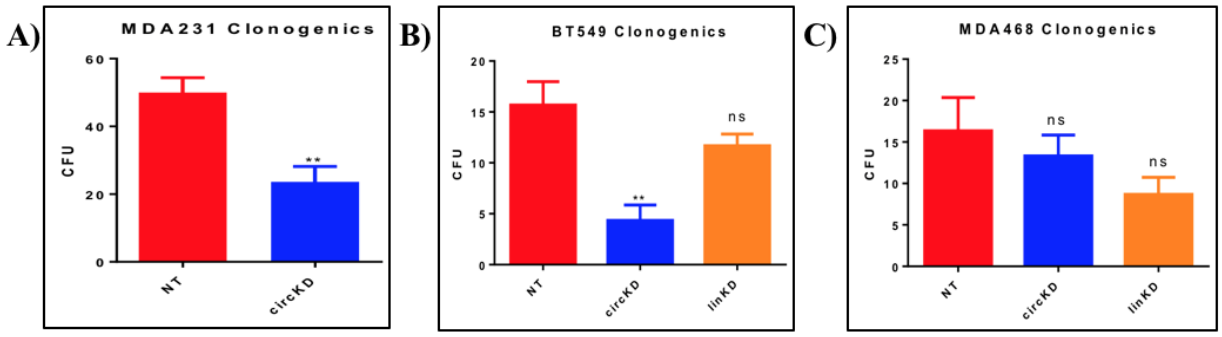


Figure 8: Clonogenic Assays
Clonogenic assays following circREV1 knockdown in: A) MDA231, B) BT549, and C) MDA468. No linear knockdown is represented in the graph for MDA231 due to it not being considered during the run for their biological assays.

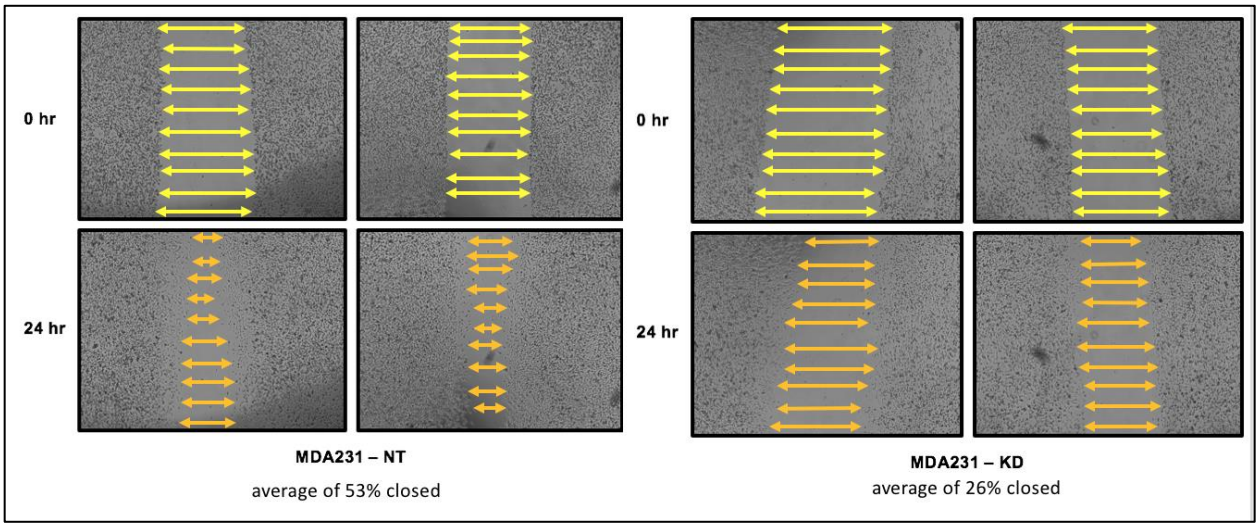


Figure 9: Scratch Assay
Scratch assay with time points of 0 hours and 24 hours in MDA231 following knockdown of circREV1. BT549 and MDA468 were not analyzed due to their lack in movement to close the wound, as the cells in their two assays moved around with no real direction.

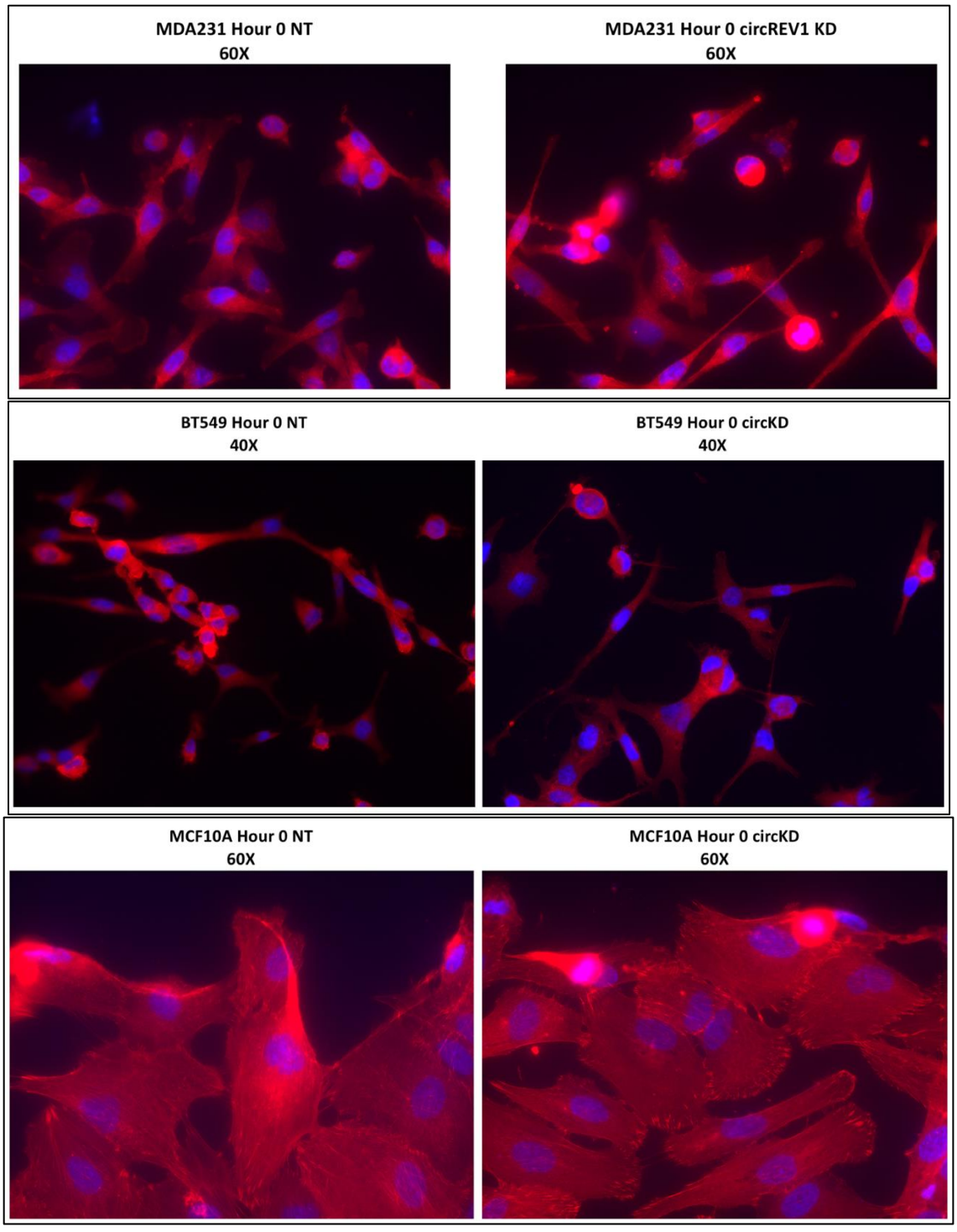


Figure 10: Actin Cytoskeleton / Focal Adhesion Assays
Actin cytoskeleton / focal adhesion assays following circREV1 knockdown in MDA231 (60X), BT549 (40X), and MCF10A (60X).

Tables

Table 3: siRNAs Used
siRNAs utilized to knockdown their respective target in each cell line.

Target	Manufacturer
Negative Control	Ambion – Cat#: AM4611
circREV1	Dharmacon – Oligo ID: HORMK-000001 (GUUAAUGGAUACACAGAAGUU)
Linear REV1	Ambion – Cat#: AM16704

REFERENCES

1. Schmadeka, R., Harmon, B. E., and Singh, M. (2014). Triple-negative breast carcinoma: current and emerging concepts. *Am J Pathol* **141**, 462-477. PMID: 24619745
2. Greene, J., Baird, A., Brady, L., Lim, M., Gray, S. G., McDermott, R., and Finn, S. P. (2017). Circular RNAs: Biogenesis, Function, and Role in Human Diseases. *Front Mol Biosci* **4:38**
3. Jakobi, T. (2018). circtools: a one-stop software solution for circular RNA research. *Dieterich Lab Revision* 25b02946
4. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher E, Hlavina W, Joardar VS, Kodali VK, Li W, Maglott D, Masterson P, McGarvey KM, Murphy MR, O'Neill K, Pujar S, Rangwala SH, Rausch D, Riddick LD, Schoch C, Shkeda A, Storz SS, Sun H, Thibaud-Nissen F, Tolstoy I, Tully RE, Vatsan AR, Wallin C, Webb D, Wu W, Landrum MJ, Kimchi A, Tatusova T, DiCuccio M, Kitts P, Murphy TD, Pruitt KD. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 2016 Jan 4;44(D1):D733-45 PubMed
5. Breastcancer.org (2021). U.S. Breast Cancer Statistics. *Breastcancer.org* revision February 2, 2021. Link.
6. American Cancer Society, Inc. (2021). Triple-negative Breast Cancer. *Cancer.org* revision January 27, 2021. Link.
7. Wang, Y., Liu, J., Huang, B., Xu, Y., Li, J., Huang, L., Lin, J., Zhang, J., Min, Q., Yang, W., Wang, X. (2015). Mechanism of alternative splicing and its regulation. *Biomed Rep.* 3(2): 152-158. PMID: 25798239
8. Hombach, S., Kretz, M. (2016). Non-coding RNAs: Classification, Biology and Functioning. *Adv Exp Med Biol.* 937:3-17. PMID: 27573892
9. Han, B., Chao, J., Yao, H. (2018). Circular RNA and its mechanisms in disease: From the bench to the clinic. *Pharmacol Ther.* 187:31-44. PMID: 29406246
10. Yang, W., Du, W. W., Li, X., Yee, A. J., Yang, B. B. (2016). Foxo3 activity promoted by non-coding effects of circular RNA and Foxo3 pseudogene in the inhibition of tumor growth and angiogenesis. *Oncogene* 35(30):3919-31. PMID: 26657152
11. Can, T. (2014). Introduction to bioinformatics. *Methods Mol Bio.* 1107:51-71. PMID: 24272431
12. Hrdlickova, R., Toloue, M., Tian, B. (2017). RNA-Seq methods for transcriptome analysis. *Wiley Interdiscip Rev RNA.* 8(1). PMID: 27198714
13. Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. *Babraham Institute*. Link.
14. Zerbino, D. R., et al. (2018). Ensembl 2018. *Nucleic Acids Res.* 46(D1):D754-D761
15. Birney, E., et al. (2004). An overview of Ensembl. *Genome Res.* 14(5):925-8

16. Dobin, A. (2019). STAR manual 2.7.0a. *Weill Cornell Medicine.* STAR manual.pdf. Link.
17. Akers, N. K., Schadt, E. E., Losic, B. (2018). STAR Chimeric Post for rapid detection of circular RNA and fusion transcripts. *Bioinformatics.* 34(14):2364-2370. PMID: 29474638
18. Jakobi, T., Uvarovskii, A., Dieterich, C. (2019). circtools – a one stop software solution for circular RNA research. *Bioinformatics.* 35(13):2326-2328. PMID: 30462173
19. Kaikkonen, M. U., Lam, M. T. Y., Glass, C. K. (2011). Non-coding RNAs as regulators of gene expression and epigenetics. *Cardiovasc Res.* **90(3)**: 430-40. PMID: 21558279
20. Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet.* **10(1)**: 57-63. PMID: 19015660
21. Eger, N., Schoppe, L., Schuster, S., Laufs, U., and Boeckel, J. (2018). Circular RNA Splicing. *AEMB* **1087**, 41-52. PMID: 30259356
22. Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., and Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc.* **11(9)**:1650-67. PMID: 27560171
23. Franken, N. A., Rodermond, H. M., Stap, J., Haveman, J., van Bree, C. (2006). Clonogenic assay of cells in vitro. *Nat Protoc.* **1(5)**:2315-9, PMID: 17406473
24. He, M. Y., Rancoule, C., Rehailia-Blanchard, A., Espenel, S., Trone, J., Bernichon, E., Guillaume, E., Vallard, A., Magne, N. (2018). Radiotherapy in triple-negative breast cancer: Current situation and upcoming strategies. *Crit Rev Oncol Hematol.* 131:96-101. PMID: 30293712
25. Gau, D. M., Roy, P. (2020). Single Cell Migration Assay Using Human Breast Cancer MDA-MB-231 Cell Line. *Bio Protoc.* 10(8):e3586. PMID: 32656296

APPENDICES

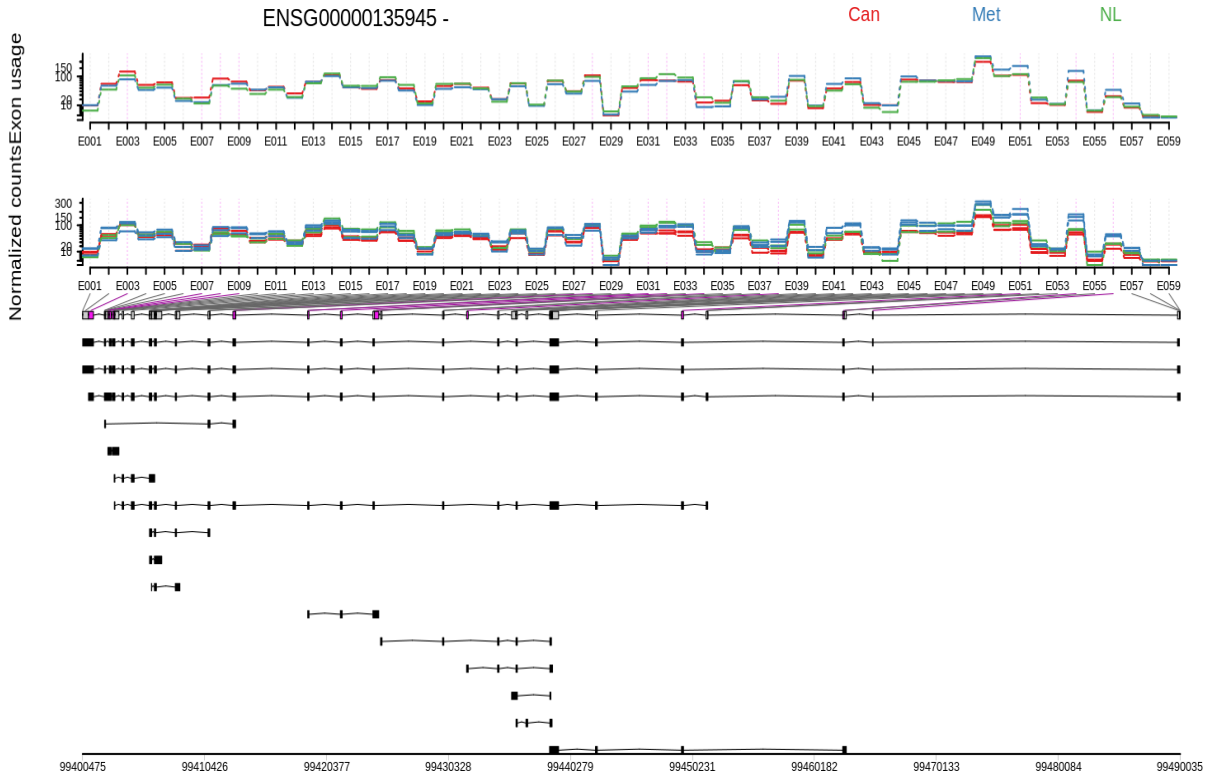
Appendix A: List of Cell Lines Used

Cell Line	Cancer Type	ATCC Description
MCF10A	N/A	non-tumorigenic epithelial cell line derived from mammary gland
BT549	TNBC	invasive ductal carcinoma from breast
MDA231	TNBC	adenocarcinoma from breast
MDA468	TNBC	metastatic adenocarcinoma from breast

Appendix B: List of siRNAs Used

Target	Manufacturer
Negative Control	Ambion – Cat#: AM4611
CircREV1	Dharmacon – Oligo ID: HORMK-000001 (GUUAAUGGAUACACAGAAGUU)
Linear REV1	Ambion – Cat#: AM16704

Appendix C: Supplemental Figures & Tables



Supplemental Figure A1: Differential Exon Usage of REV1 (ENSG00000135945)

Differential exon usage of REV1 in which “E056” represents Exon 2 and “E054” represents Exon 3. This is determined by the fact that our transcript is on the negative strand and corresponds to exons 2 and 3 in the reverse transcript assembly. “Can” in red represents BT549, “Met” in blue represents MDA231, and “NL” in green represents MCF10A.

Supplemental Table A1: Explored Reference Genes

Representing the reasoning as to why MRPL19 was chosen as the reference gene for qRT-PCR normalization. C_T values represent the cycle (out of 40) at which the nucleic acid adequately amplified. Reference genes need to amplify at approximately the same cycle in order to appropriately normalize across each cell line. (-R) represents the idea that these cell pellets were not treated with RNase R, and (cc) represents the idea that these cell pellets were run through an RNA Clean & Concentrator Kit following RNA extraction.

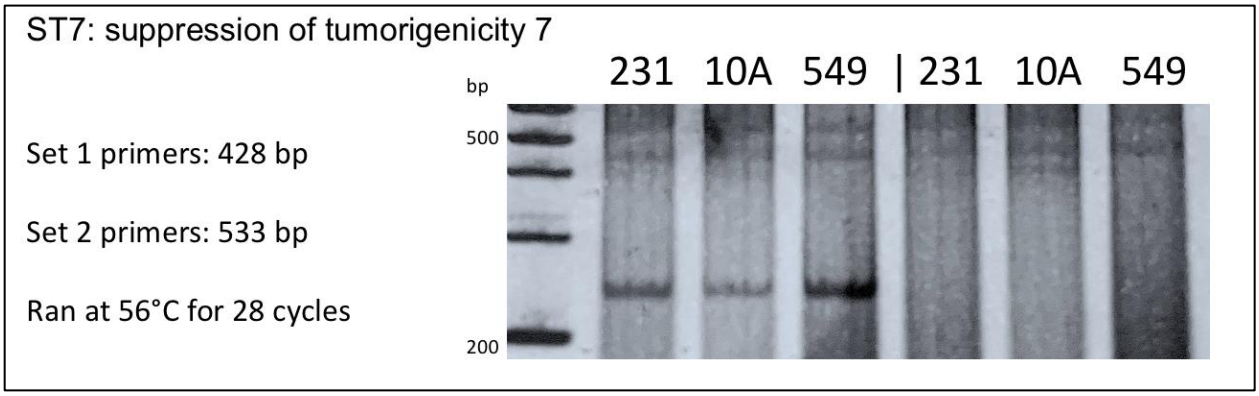
Sample Name	Target Name	C_T
MDA231-1-R (cc)	RPLP0	12.52194
MDA231-1-R (cc)	RPLP0	7.160653
BT549-1-R (cc)	RPLP0	7.031374
BT549-1-R (cc)	RPLP0	13.50618
MCF10A-1-R (cc)	RPLP0	14.47884
MCF10A-1-R (cc)	RPLP0	14.43928
MDA231-1-R (cc)	MRPL19	18.19035
MDA231-1-R (cc)	MRPL19	17.85708
BT549-1-R (cc)	MRPL19	17.79321
BT549-1-R (cc)	MRPL19	17.80665
MCF10A-1-R (cc)	MRPL19	18.98186
MCF10A-1-R (cc)	MRPL19	18.99398
MDA231-1-R (cc)	Actin B	20.83366
MDA231-1-R (cc)	Actin B	20.94716
BT549-1-R (cc)	Actin B	20.59416
BT549-1-R (cc)	Actin B	20.58715
MCF10A-1-R (cc)	Actin B	24.23779
MCF10A-1-R (cc)	Actin B	24.26465
MDA231-1-R (cc)	B2-MG	18.9381
MDA231-1-R (cc)	B2-MG	18.8308
BT549-1-R (cc)	B2-MG	18.56968
BT549-1-R (cc)	B2-MG	18.8966
MCF10A-1-R (cc)	B2-MG	20.63952
MCF10A-1-R (cc)	B2-MG	20.58218

Supplemental Table A2: Reference Gene Primers

qRT-PCR primers utilized to determine the best fit reference gene.

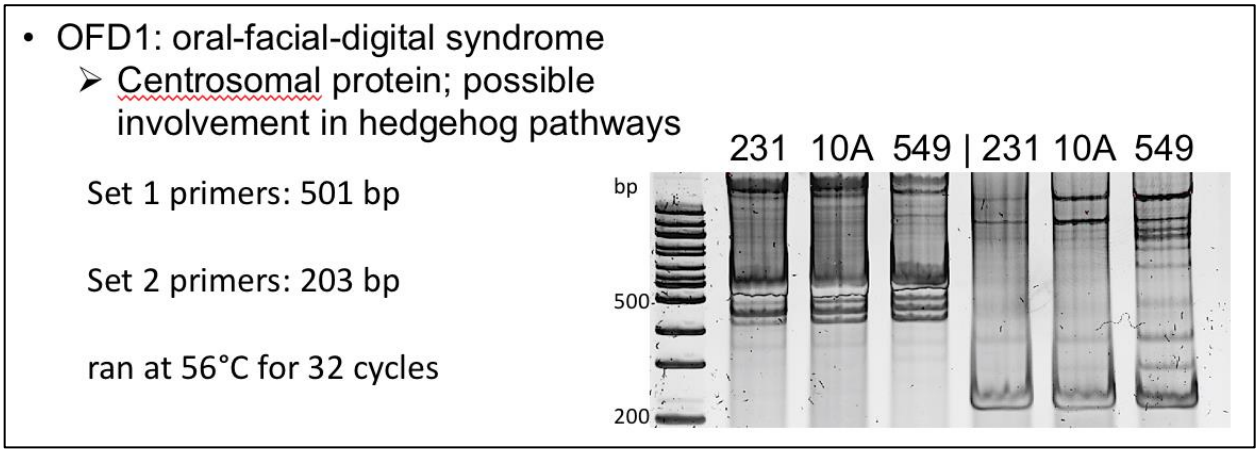
Target	**Manufacturer or Sequence**
RPLP0	IDT PrimeTime: Hs.PT.39a.22214824
MRPL19	IDT PrimeTime: Hs.PT.58.40629433
Actin B	IDT PrimeTime: Hs.PT.39a.22214847
B2-MG	F: 5'-AAC TTA GAG GTG GGG AGC AG-3' R: 5'-CAC AAG CAT GCC TTA CTT TAT C-3'

Appendix D: Unpursued Data



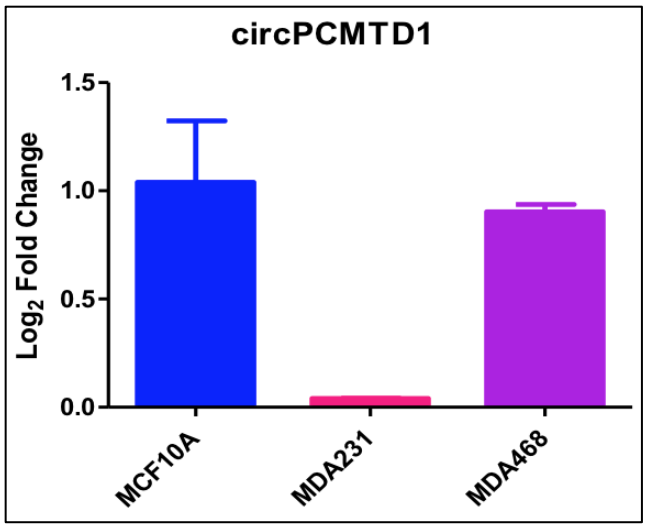
Unpursued Data Figure B1: Alternative Splicing PCR of ST7

Alternative splicing exploration of ST7 in the respective cell lines. The image represents a polyacrylamide DNA gel with PCR run at 56 degrees Celsius for 28 cycles. Set 1 primers were to target a product of 428 base pairs while Set 2 primers were to target a product of 533 base pairs.



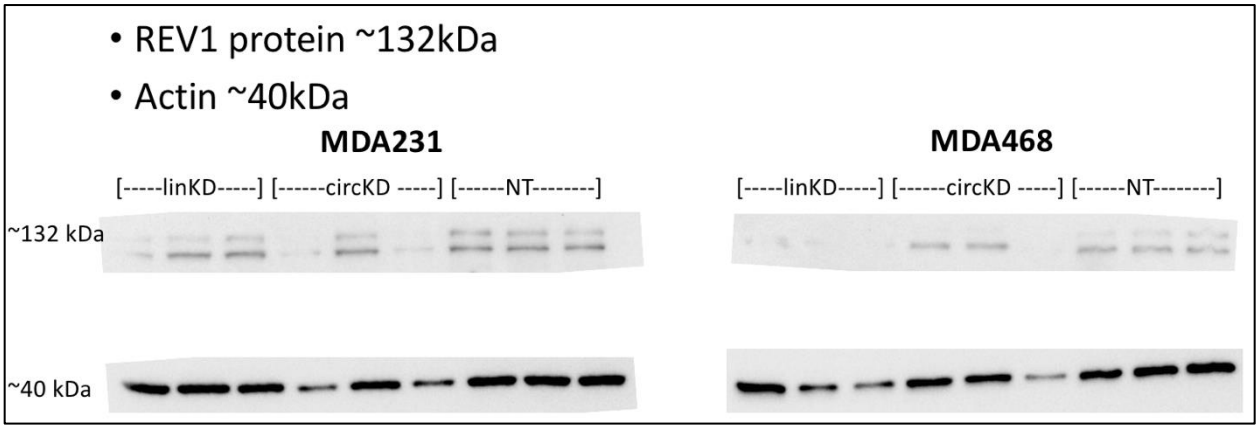
Unpursued Data Figure B2: Alternative Splicing PCR of OFD1

Alternative splicing exploration of OFD1 in the respective cell lines. The image represents a polyacrylamide DNA gel with PCR run at 56 degrees Celsius for 32 cycles. Set 1 primers were to target a product of 501 base pairs while Set 2 primers were to target a product of 203 base pairs.



Unpursued Data Figure B3: CircPCMTD1 qRT-PCR

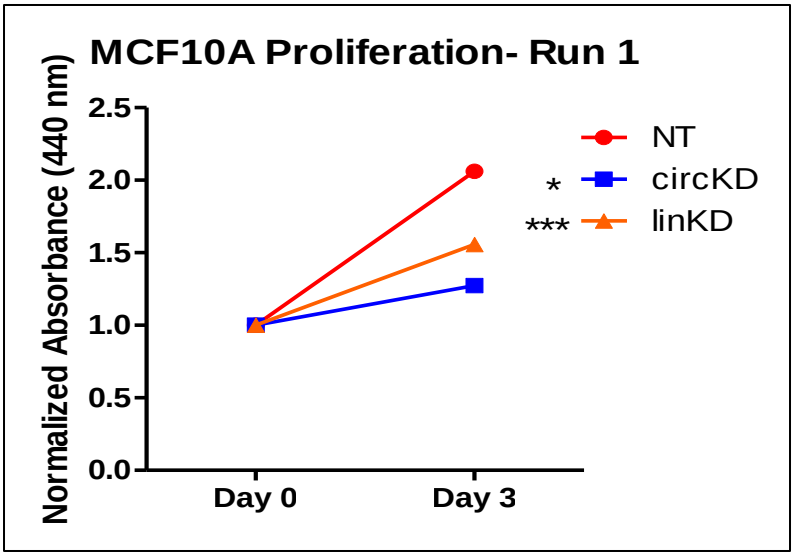
QRT-PCR analysis of *in vitro* gene expression for circPCMTD1. According to the targets in Table 1, expression should have been upregulated in MDA231 compared to MCF10A.



Target	Antibody	Selection
REV1	Santa Cruz Biotechnology: Anti-REV1 (A-11) sc-393022	Mouse, monoclonal
B-Actin	Cell Signaling Technology: B-Actin Mouse mAb #3700	Mouse, monoclonal

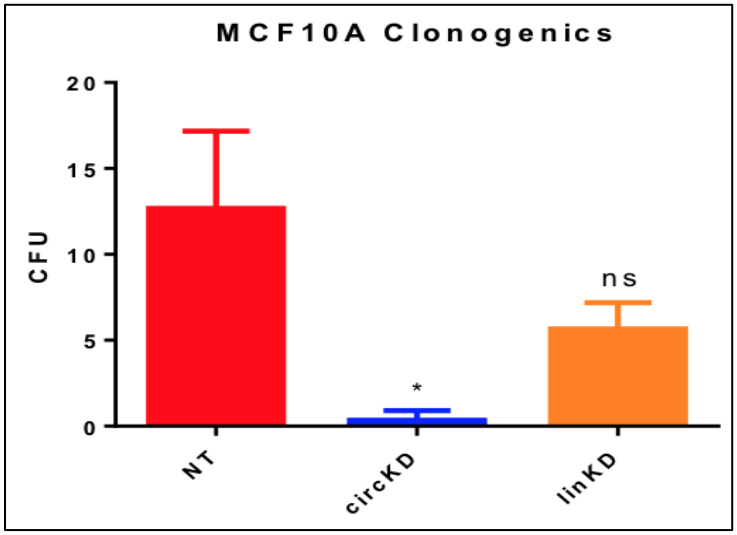
Unpursued Data Figure B4: Western Blots and Antibodies for circREV1

Western Blot of MDA231 and MDA468 with circREV1 knockdown and linear REV1 knockdown. REV1 protein is said to be approximately 132 kDa while actin is approximately 40 kDa. While protein levels were nanodropped to ensure even loading, the actin levels seem to indicate that there was still uneven loading concentrations. Experiments were not reproduced due to the lack in available cells to plate for protein and RNA harvesting when running biological assays. Antibodies utilized are represented in the table below the blots.



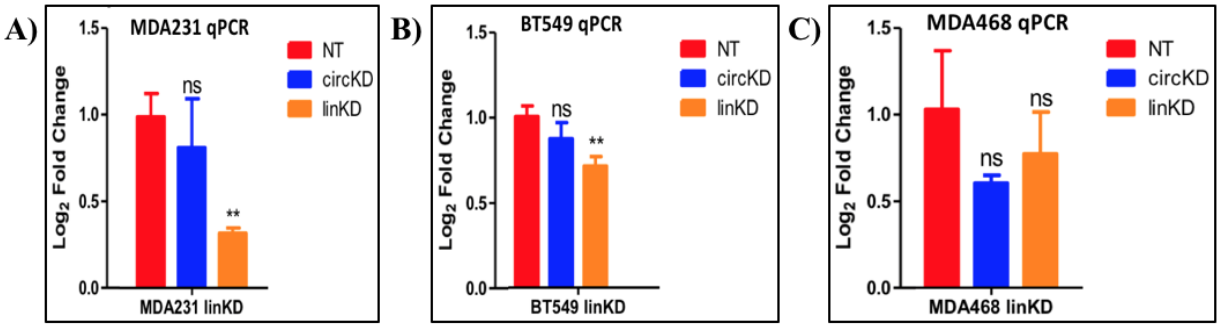
Unpursued Data Figure B5: MCF10A Proliferation Assay

Proliferation data for MCF10As following knockdown of circREV1 and linear REV1. Unfortunately, a lab mate had to complete the Day 3 reading due to illness of the original researcher, and they put the WST1 dye on all of the remaining plates, meaning the following days were not able to be read. Experiments were not reproduced due to the lack in urgency to emphasize the knockdown of circREV1 in the non-cancerous cell line.



Unpursued Data Figure B6: MCF10A Clonogenic Assay

Clonogenics data for MCF10As following knockdown of circREV1 and linear REV1. Colonies for this cell line took very long to form and were also sparse. Experiments were not pursued or reproduced due to the lack in urgency to emphasize the knockdown of circREV1 in the non-cancerous cell line.



Unpursued Data Figure B7: Linear REV1 Knockdown

Knockdown of linear REV1 in: A) MDA231, B) BT549, and C) MDA468. Linear knockdown was inconsistent and often lead to considerable cell death, making it hard to harvest cells for qRT-PCR analysis. Because of the complications seen with linear REV1 knockdown, it was decided that experiments involving their knockdown may be unreliable and were therefore not further pursued.