Effects of *TOX3* Regulation using CRISPR-Cas9 genome editing and the corresponding effects on susceptibility to Breast Cancer

An essay submitted in partial fulfillment of

the requirements for graduation from the

Honors College at the College of Charleston

with a Bachelor of Science in Biology

Cody Ashy

May 2017

Advisor: Melissa N. Scheiber, Ph.D.

Secondary Reader: Bartholomeus Smits, Ph.D.

Abstract

One in eight women will be diagnosed with breast cancer. Due to the prevalence, high morbidity, and high costs of treatment, new screening and prevention methods need to be developed. *TOX3* downregulation is correlated to increased breast cancer risk. This study investigated the *TOX3* locus, utilizing state of the art CRISPR/Cas9 technology to reduce gene expression in rats and develop a model of susceptibility. In addition, *TOX3* overexpression was achieved using a Cas9 mutant (dCas9-VP64) that is not able to cut the gene. It has been previously established in many known oncogenes that upregulation transformed healthy cells to a cancerous state. Using this technology, we theoretically could transform normal mammary epithelial cell line, MCF10a, cells into a cancerous state. This invention can serve as an efficient and inexpensive screening tool to study candidate genes and determine new oncogenes.

Introduction

Breast Cancer

Approximately one in eight women will develop breast cancer over the course of their lifetime (1). Breast cancer is the uncontrolled division of abnormal tissues within the breast, due to mutations within the DNA of cells. Congregation of mutated cells results in a tumor formation. When cells metastasize, or break off into the blood vessels or lymphatic vessels, they can spread to other tissues spreading the cancer. The breast tissue is made up of adipose tissue, or fat cells, from the collarbone to the underarm and across the rib cage (2). Within the adipose tissue there are structures that make up the breast epithelium: lobes, lobules and milk ducts. The lobes are collections of lobules that produce milk, and the milk ducts are the stems that carry milk to the nipples of the breast. Cancer is predominantly found within these breast structures (2).

Surrounding the Breast epithelium are a series of lymphatic vessels, lymph nodes, blood vessels, ligaments, nerves, and fibrous connective tissue. Where the cancer is located serves as a clinical indicator of the severity of the cancer and how far it has spread (2). Due to the disease's prevalence, mortality, and substantial cost associated with treatment, new screening sites and preventative measures need to be further investigated.

ER/PR/HER2 classification of Breast Cancer

In the clinic, breast cancer is often classified by expression of ER, PR, and HER2. Breast cancer cells usually have estrogen receptors (ER), Progesterone receptors (PR), and human epidermal growth factor 2/neu (HER2) receptors that contribute to their rapid growth and proliferation. These structures are common in healthy tissue as well as cancerous tissue. Binding

of specific ligands to these receptors initiate cellular responses often resulting in regulation of gene expression. However, these receptors can malfunction or be over expressed because of genetic mutations. Consequently, uncontrolled cell division due to increases in receptor responses can lead to tumor formation thereby explaining why ER, PR, and HER2 receptors are common in tumors. The presence of these structures in tumors are used for drug targeting (3). Approximately 30-70% of breast cancer cases are luminal A and 10-20% are Luminal B, which are typically ER+/PR+ (4). HER2 cases are characterized by a high expression of HER2, and account for roughly 15% of cases (4). Basal-like breast cancer is typically ER/PR/HER2 negative and typically is the most aggressively growing cancer. Therefore, ER/PR/HER2 cases typically have more negative prognoses. Luminal A (usually ER+) breast cancer has the best prognosis compared to the other subtypes since tumors can be treated with endocrine therapy. Current treatment methods target these estrogen receptors to selectively target tumors to prevent them from proliferating. Additionally, when tumors lack these receptors, such as those ER/HER2/PR negative cases, traditional targeting methods fail. Thus, there is a need for new methods of treating breast cancer or preventing it entirely. Early detection is also necessary in improving prognoses pointing to the necessity for new screening methods to identify increased risk. The focus of this study is on ER+ breast cancer because it is the most prevalent subtype of breast cancer.

Disparities in Breast Cancer

African American (AA) women experience higher rates of morbidity and mortality in breast cancer cases. Socioeconomic factors such as low income and lower education resulting in decreased access to healthcare and screening contribute to part of this disparity. However, AA

women are still more likely to develop triple negative breast cancer (TNBC), thereby having worse prognoses than Non-Hispanic white (NHW) women. NHW are more likely to develop breast cancer than AA women (5-9). In one study, AA women had increased mortality for breast cancer after prognostic factors, education, and income were controlled (10). Therefore, socioeconomic factors do not explain the ancestral disparities between AA women and NHW women indicating biological factors such as genetics are likely involved.

Risk Factors

Breast cancer has been found to have a strong genetic component, however known genes can only account for 25% of familial risk (11). Genome wide association studies (GWAS) were conducted to analyze the genome for mutations that could lead to increased risk. Besides the well-known high penetrance mutations such as BRCA1/2, these studies found that most susceptibility loci were in noncoding regions of the genome. Therefore, these loci are likely involved in regulatory elements of gene expression. Most loci identified in the GWAS were also low penetrance, meaning that their presence does not guarantee a substantial increase in breast cancer risk. The BRCA1/2 mutations that are high penetrance contrarily indicate a more substantial likelihood of breast cancer development.

GWAS identify single nucleotide polymorphisms (SNPs), which are variations of a single base pair in DNA. These variants are common within the population. The SNPs investigated in this study are associated with increased risk of breast cancer. Areas within the genome can be divided into distinct linkage disequilibrium (LD) blocks. LD is a measure of how likely genetic variants are linked to one another, with higher LD corresponding to increasing likelihood of linkage. The *16q12-TOX3/LOC643714* locus (*16q12.1*) on human chromosomal band *16q12.1* had multiple variants that were found to have some of the strongest association with breast

cancer through the GWAS, although these samples were mainly from women with NHW ancestry (11). The variants in this locus were linked to increased ER+ breast cancer and had a weaker association to ER/HER2/PR negative breast cancer in NHW ancestry (12). Since this locus was identified as having variants strongly associated with breast cancer, it was decided to be the central focus of this present research study. One of these variants, rs3803662, was identified in the GWAS. This variant was composed of 21 different SNPs that had LD correlation coefficients of 0.8 or higher (r^2>0.8). The high LD along with the associated risk of breast cancer in NHW made this variant a satisfactory selection for further investigation. The variant rs3803662 is in the promoter region of *TOX3*. Consequently, it is believed that this site and variant is associated with binding to a transcription factor.

The *16q.12.1-TOX3/LOC643714* Locus

Within the locus of interest in this study, there are two genes of interest. One is *TOX3*, which is a transcription factor. The other is *LOC643714* which is a hypothetical gene that is non-coding. However, *LOC643714* may encode for a transcript with an unknown function. The NHW variants in this study are located closer to the 5' end of the *TOX3* locus. Since the variants are in the noncoding region, they likely affect transcription levels. *TOX3* has been a focus of various breast cancer studies. There is a correlation between the risk allele rs3803662 and lower *TOX3* expression (13,14). Additionally, poorer outcomes and increased mortality of the Luminal-A (mostly ER+) subtype have been found to be correlated with lower *TOX3* expression (15). Knockdown of *TOX3* in cell models has also been associated with increased cellular proliferation (13).

The function of *TOX3* is not entirely understood presently. *TOX3* is expressed in ER+ luminal cells[16]. *TOX3* function in Luminal-A breast cancer cells has been shown to upregulate ER target genes. This interaction could prevent ER-targeting therapy possibly explaining the poorer outcomes and increased mortality in Luminal-A breast cancer cases. It is hypothesized that *TOX3* binds to ER-alpha, one of the main subtypes of ER. The combination of incomplete understanding of *TOX3* function along with the correlations of decreased *TOX3* expression with increased breast cancer risk suggests *TOX3* requires further investigation.

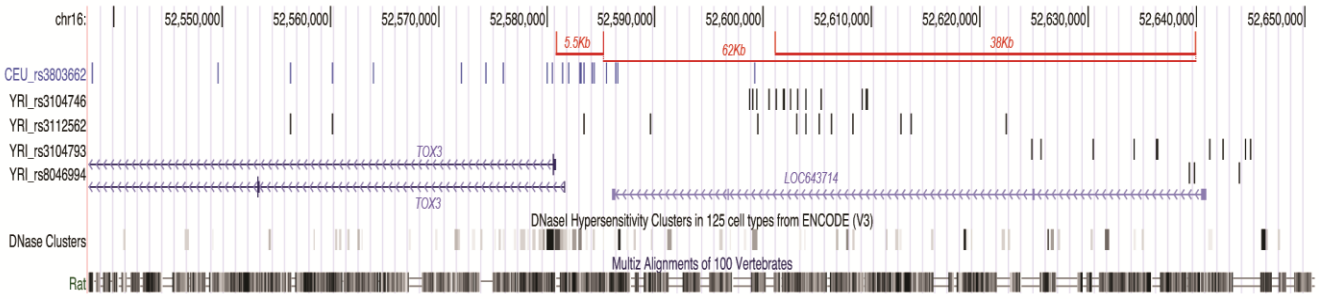

Figure 1: UCSC genome browser view of the *16q12.1* Breast cancer-associated locus. The 21 SNPs correlated (r^2>0.8) to GWAS-identified variant rs3803662 that is associated with Breast cancer in the NHW population (indicated as CEU_rs3803662), are shown as blue vertical bars. Also shown are the positions of transcripts *TOX3* and *LOC643714*, DNAseI hypersensitivity sites (indicative of gene regulatory elements), and sequence conservation (MultiZ Alignment) to the rat genome (indicating co-linear alignment and good conservation). Indicated in red are the human orthologs for the rat sgRNA sites for CRISPRCri/Cas9-mediated targeting of the rat orthologs of the genetic elements associated with breast cancer in NHW women (5.5Kb deletion) and AA/NHW women (62Kb deletion).

Crispr-Cas9 Technology

Crispr-Cas9 is a new method for editing the genome that is relatively cheap, efficient, and limits off-target interactions. The technology stems from type II bacterial clustered regularly interspaced short palindromic repeats (CRISPR)-Cas immune system. Cas9 potentially cuts DNA wherever it is guided, which can be controlled by expression of small guide RNAs (sgRNAs). Therefore, the CRISPR-Cas9 system can be guided to specific locations of the genome and can effectively knockout segments of the genome, or allow for insertions at places within the genome as demonstrated in previous studies (17,18). The Crispr-Cas9 complex can also be altered for other purposes. The complex can be deactivated to be guided to certain areas of the genome without cutting DNA. This specific Cas9 is known as dCas9. Pairing this complex with the VP64 transcription activation domain has been found to promote transcription of genes surrounding the areas where the dCas9-VP64 domain is guided (19). Consequently, gene expression can be increased differentially depending on where specifically this complex is guided. MCF-10 A cells are a non-cancerous breast epithelium cell line. Therefore, these cells serve as a good model for observing the relationship between gene expression and breast cancer development. Engineering the dCas9-VP64 complex to be used in MCF-10A cells, it is possible to target the 21 variants within rs3803662 to enhance *TOX3* expression. Observing the fold-change in expression associated with guiding of the dCas9-VP64 domain to each variant location would provide evidence for which control *TOX3* expression the most. Furthermore, molecular gene therapy could potentially utilize this technology to correct *TOX3* expression levels thereby lowering the risk of developing breast cancer.

If this technology is capable of upregulating *TOX3* expression when guided to the 21 variants within rs3803662, theoretically the dcas9-VP64 complex could be used to upregulate

other genes. Therefore, other sgRNAs can be designed to guide the complex to other gene promoter regions. Consequently, these genes could also be upregulated. The sgRNAs associated with the complex would be designed to target the promoter region of each gene because the promoter region is where transcription factors typically bind. The guiding of the dcas9-VP64 complex to the promoter region will likely lead to the largest increase in expression in comparison to the complex being guided towards other regions around each gene. If multiple genes can be regulated with the dCas-VP64 complex, then the entire genome could potentially be targeted. Using the guided dCas-VP64 complex, other genes could theoretically be targeted and upregulated. By targeting the entire genome, this technology can be used as a novel screening tool to discover novel oncogenes. Genes that are upregulated and transform healthy breast epithelial cells (MCF-10A) into cancerous cells would be classified as oncogenes. In conclusion, this technology could provide a faster and more cost-effective method for discovering new oncogenes.

Rat models for breast cancer

The *16q.12.1-TOX3/LOC643714* Locus is highly conserved between the rat and human genome. Rats are the preferred model for ER+ breast cancer due to similar morphology of rat mammary glands in comparison to human mammary glands. The Wistar Furth (WF) inbred rat is bred specifically to be susceptible to ER+ breast cancer, and this was the rat breed used in this present study. Furthermore, mice can develop ER+ tumors, but these tumors are not entirely ER dependent. Therefore, the rat model serves as the best model for understanding the phenotypic effects of the specific variants within this locus. Carcinogenic studies can be administered in the future to indicate the susceptibility of rats with specific genetic variants to developing breast

cancer. The susceptibility of the rats can be applied to understanding the likelihood that humans with similar genetic variants will develop breast cancer. To simulate the genetic variants discovered by the GWAS, several knockout breeding lines of rats were created to remove large portions of the genome corresponding to increased risk in AA and NHW ancestry. The genetic knockouts were created by pro-nuclear injection of sgRNA and Cas9 mRNA into rat embryos. The specific sgRNA guided the Crispr-Cas9 complex to specific sites to remove specific segments of the locus. Em1, em2, em4, em5, and em6 are the five knockout rat strains within this study targeting various SNPs within the *16q.12.1-TOX3/LOC643714* Locus (Figure 2). Since the rat genome is orthologous to the human genome, these animal models will provide greater understanding of the effects of *TOX3* in ER+ breast cancer. Additionally, which regions of the *16q.12.1-TOX3/LOC643714* Locus affect regulation of *TOX3* will be understood by observing the phenotypes from the knockouts created. Furthermore, *TOX3* function will be further investigated.

Hypothesis

After creation of the five breeding lines of knockouts in rats, it is hypothesized that *TOX3* expression will decrease due to these knockouts being present within the regulatory region of this locus. Selected rats will be injected with 7,12-dimethylbenz[a]anthracene (DMBA), a known carcinogen, to stimulate tumorigenesis. Thus, carcinogenic administration will likely lead to increased susceptibility of rats within these five knockout lines due to decreased *TOX3* expression. The em1 and em2 knockouts are hypothesized to have the most severe effect on *TOX3* expression due to their proximity to the promoter of *TOX3.* Understanding the effects of each specific knockout will provide insight into which regions of the *16q.12.1-*

TOX3/LOC643714 correspond to the most severe phenotypes, the mechanisms behind why decreased *TOX3* expression leads to increased breast cancer risk, and how *TOX3* is regulated uniquely by each specific targeted region.

Using the dCas9-VP64 mechanism, it is hypothesized that if the mechanism is guided to the variants within rs3803662, *TOX3* transcription and expression will increase. Additionally, it is expected that each variant will correspond to a unique fold-change in *TOX3* expression in MCF-10A cells. Ergo, how *TOX3* expression is affected by each variant will be understood. These findings can later lead to development of molecular gene therapy to increase *TOX3* expression to normal levels and lowering the risk of developing breast cancer. Furthermore, if the dCas9-VP64 complex is guided to other genes, then these genes will be upregulated similarly to *TOX3.* The area around each gene that will result in the highest fold-change in expression level will likely be near the promoter region. Lastly, if genes are upregulated and cause breast epithelial cells to transform into cancerous cells, these respective genes can be deemed oncogenic. The utilization of this tool could then be used as an efficient screening tool to discover new oncogenic genes.

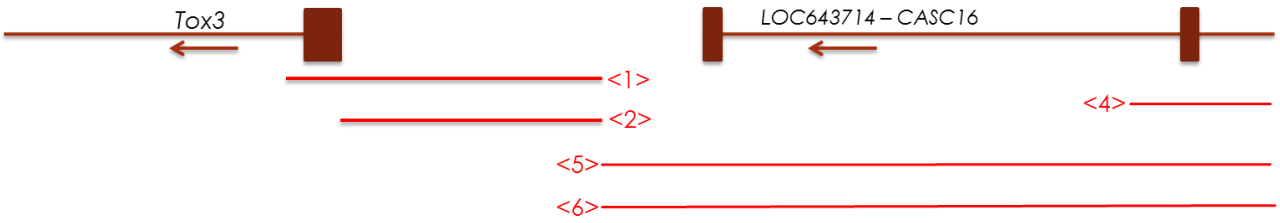

Figure 2: The five knockout rat mutant breeding lines locations within the *16q.12.1-TOX3/LOC643714* locus are shown above: <em1>, <em2>,<em4>, <em5>, and <em6>. The <em1> mutant is hypothesized to be the most severe since it knocks out the promoter of *TOX3.* Similarly, <em2> is missing only a methionine and will likely also have a severe phenotype. The <em4>, <em5>, and <em6> mutants affect regulatory regions.

Methods

RNA Extraction

Sprague Dawley outbred rats were ordered from Envigo. Embryonic injections were performed using a microinjector to generate the unique em1, em2, em4, em5, and em6 rat strains. Rats produced from embryonic injection were then backcrossed to ER+ breast cancer susceptible Wistar Furth inbred rats, also ordered from Envigo.

Rats were euthanized following approved IACUC protocols. Endogenous lower mammary glands were excised from the animals and immediately frozen in liquid nitrogen and stored at -80°C. For extraction, a 150mg piece of the tissue from the frozen section was lysed in 1ml of RNeasy Mini Kit Lysis buffer. RNA was isolated following the manufacturer's protocol (Qiagen #74104).

For cellular RNA extractions. Cultured cells were pelleted and resuspended in 1ml of RNeasy Mini Kit Lysis buffer. Following suspension, the RNA was then isolated following the RNeasy Mini Kit protocol (Qiagen #74104). All RNA was stored at -80°C until further use.

RT-QPCR

For reverse transcription (RT), 800ng of RNA was placed into a single well of a 96-well thermocycler plate. The following reagents were added to each sample: 0.5µl Oligo dt 18 (1mM, ThermoFisher Scientific), 1µl oligo dNTP's (10mM, ThermoFisher Scientific), 1µl random hexamer (100µM, ThermoFisher Scientific), and 0.4µl RNA SC (ThermoFisher Scientific) and Milli-Q H_2O for a 10µl reaction. The samples were incubated at 90°C for 3 minutes. After incubation, 4µl 5x FSS (ThermoFisher Scientific), 1µl 0.1 mM DTT (ThermoFisher Scientific), and 1µl Reverse Transcriptase (invitrogen) were added to each sample. Following addition of said reagents, the samples were incubated at 42°C for two hours and then 70°C for 15 minutes.

This cDNA was diluted 1:4 for use. One randomly chosen cDNA sample was serially diluted two-fold from 1:2 to 1:256. 1.5µl of each diluted cDNA sample was used in each reaction. 13.5µl of master mixes for each gene being investigated were added to every 1.5µl of cDNA. The master mixes were made up of: 7.5µl of SYBR Green (2x), 0.75µl forward/reverse qPCR primer (100µM) and 4.5µl Milli-Q water. *TOX3,* CR2F2R, MYC, FAM, LSM1, 18S primers (See supplemental figure 1 for primer sequences) were used for respective master mixes. Each cDNA sample was run as a triplicate with each master mix. A serial dilution was run with each master mix using the serially diluted cDNA. qPCR was run on a Roche LightCycler 480 using the following settings: preincubation 1 cycle (5minutes at 95°C), amplification 45 cycles (95°Cfor 10 seconds, 60°C for 10 seconds, 72°C for 10 seconds), melting curve 1 cycle (95°C for 5 seconds, 65°C for one minute, 97°C), and cooling 1 cycle (40°C for 30 seconds).

Upon analysis, all reactions were normalized to the serial dilutions for each master mix. If CT values within a triplicate were >0.5 apart, those readings were omitted. Average CT values were calculated for each triplicate. All gene expression was normalized to *18s* expression for comparison between samples.

Cell-culturing

MCF-10A cells were cultured in DMEM/F12 media (Invitrogen #11330-032) with Horse-serum (Invitrogen #16050-122, 5%), EGF (20ng/ml), Hydrocortisone (0.5mg/ml), cholera toxin (100ng/ml), Insulin (10µg/ml), and Pen/Strep (100x solution Invitrogen #15070-063, 1%) added. HEK293T cells media consisted of DMEM/High Modified media supplemented with: Pen-Strep(1%), +4.500 mg/L glucose, +110 mg/L Sodium Pyruvate, and -L- glutamine).

Single Cell Cloning

For creation of single cell clones, MCF-10A dCas-VP64_Blast cells were plated at a density of 4,000 cells/well of a 96-well plate. Cells were serially diluted 1:2 down the first column wand across the plate to obtain single cells in the final wells. The plate was monitored daily for confluency and single cells were selected for expansion and harvested when the plate reached 90% confluency. Frozen populations were stored in 90% complete media, 10% DMSO at -80°C until further use.

Plasmid Preparation

For lentiviral production, psPAX2 (plasmid #12260), VSV.G (plasmid #14888), lenti-dCas-VP64_Blast (plasmid #61425), and lentiGuide-Puro (plasmid #52963) were ordered as bacterial stabs from Addgene. The bacterial stabs were streaked onto a LB+Ampiciliin (100µg/ml) agar plates 12-18 hours at 37°C. Single colonies were isolated from the plates and cultured overnight in an LB+ampicillin solution (100µg/ml) placed in a shaking incubator at 37°C, 220 RPM. Plasmids were isolated using QIAprep Spin Miniprep Kit (#27106) and verified by gel electrophoresis.

Cloning of sgRNAs into lentiGuide-Puro Vector

The isolated lentiGuide-Puro Vector was digested with BsmbI (Esp3I) in a reaction containing: 5µg lentiGuide-Puro, 3µl FastDigest Esp3I, 3µl FastAP, 6µl 10x FastDigest Green Buffer, 0.6µl 100mM DTT, and a calculated amount of Milli-Q water to bring the total reaction volume to 60µl. The reaction was incubated at 37°C for 30 minutes. Following incubation, the sample was run on a 0.8% agarose gel at 140V for 45 minutes. The gel was stained in Ethidium

Bromide for ten minutes, rinsed with water for ten minutes, and then imaged using UV light. The cut vector was excised from the gel and isolated using the Thermo Scientific Gene Jet Gel Extraction Kit (K0691). Quality of the isolated cut lentiGuide-Puro vector was analyzed by gel electrophoresis.

Oligos for the sgRNA were hybridized to one another in a reaction consisting of: 1µl Oligo 1 (100µM), 1µl Oligo 2 (100µM), 1µl 10X T4 Ligation Buffer, 6.5µl ddH_20, and 0.5µl T4 PNK. Oligo sequences used in this experiment can be found in supplemental figure 2. The hybridization reaction was placed in a thermocycler and then incubated at 37°C for 30 min. Following incubation, the samples were heated to 95°C for 5 minutes, and then were placed on the bench top to cool to room temperature (25°C). The hybridized oligos were then diluted 1:200 into sterile water.

The hybridized oligos were ligated into the BsmbI digested lentiGuide-Puro Vector in a reaction consisting of: 50ng of digested lentiGuide-Puro, 1 µl diluted hybridized oligos, 5µl 2x Quick Ligase Buffer, 1µl Qucik Ligase, and a calculated volume of ddH_2O to bring the total reaction volume to 11µl. The reaction was incubated for ten minutes at room temperature, and then incubated for 12 hours at 16°C.

The ligated lentiGuide-Puro vector with the sgRNA inserted was transformed in Thermo Fisher Stbl3 cells (C737303) by adding 1µl of ligated vector to 16µl of cells. The cells and ligated vector mixture was placed-on ice for 30 minutes and occasionally mixed. The vial was then incubated at 42°C for 45 seconds, and then placed back on ice for two minutes. 250µL of SOC recovery medium was added subsequently, and the samples were placed in a shaking incubator set to 37°C and 220 RPM. After 1 hour incubation, the samples were plated onto LB agar plates with added Ampicillin (100µg/ml). The plates were incubated for 12 hours at 37°C. Colonies were

picked and analyzed by Colony PCR for quality analysis. If colonies were deemed of sufficient quality, they were grown in LB-Ampicillin solution (100µg/ml) for 12 hours in a shaking incubator set at 37°C and 220 RPM. The transformed vectors were then isolated using were isolated using QIAprep Spin Miniprep Kit (#27106).

Transfection of HEK293T cells for lentiviral production

24 hours before transfection, HEK293T cells were plated at a density of $3.5x10^6$ cells/well. The day of transfection, plates were at approximately 90% confluency. Antibiotic-free complete media was warmed to 37°C. Simultaneously, 5µg of each plasmid (VSV.G plasmid, psPAX2 plasmid, and dCas-VP64_Blast) were separately added to 1 ml aliquot of Opti-MEM and alongside separate tubes containing Lipofectamine 3000 in 1 ml of Opti-MEM. Equal amounts of Lipofectamine solution was added to the plasmid mixtures and incubated for 20 minutes at room temperature. 8ml of antibiotic free media was added each mixture and the total volume was used to feed each plate of HEKT293 cells. The media was replaced after 24 hours and at the 48 hours timepoint, media with virus was harvested and filtered through a 0.2µm filter. Filtered virus was stored at 4°C. One plate of HEK293T cells was prepared for each virus. The same protocol was used to produce the lentiGuide-Puro viruses, except dCas9-VP64_Blast plasmid was substituted for the unique lentiGuide-Puro sgRNA plasmids including 13+, 17+, 18+, 19-, 27-, 31+, NC2+, LGP+, *MYC*, *LSM1*, *Fam 2.1*, *Fam 2.2*, and 10-.

Infection of MCF-10A Cells with lentiviruses

MCF-10A cells were grown to 90% confluency. Cells were trypsinized and resuspended in 8.5 ml of media, 500µl of dCas-VP64_Blast virus, and 10µl of polybrene. 48 hours later the

media was removed. Cells were selected using media containing (10µg/ml) Blasticidin, including proper controls. Visualization of control wells allowed selection to end and antibiotic medium was replaced with complete medium.

For lentiGuide-Puro viral infections, cells were plated at a density of 5x10^5 cells/well. 13+, 17+, 18+, 19-, 27-, 31+, NC2+, LGP, *MYC, LSM1, FAM 2.1, FAM 2.2,* and 10-, each were plated into three wells with 500µl, 100µl, and 50µl of said virus respectively added to MCF-10A media. Appropriate amounts of polybrene were added to each well to yield a final concentration of 1µg/ml. After 48 hours, selection with Puromycin was initiated by adding Puromycin to cell media for a final concentration of 20µg/ml. Once control wells died off, selection was ceased and antibiotic medium was replaced with normal MCF-10A medium.

In addition to selection for the lentiGuide-Puro infection, a transient assay was also completed. Infections were completed using 13+, 17+, 18+, LGP, *MYC, LSM1, FAM 2.1, FAM 2.2,* and 10- viruses. MCF10-dCAS_VP64 cells were plated a a density of $1.0x10^6$cells/well. Polybrene (1µg/ml) and 100µl of each respective virus was added to each well. Each virus was used to infect two wells. One well was harvested after 24 hours after infection, while the other was harvested 48 hours after infection. When cells were harvested, free-floating cells as well as adhered cells were harvested except for the 48-hour harvest on February 22nd, 2017.

Results

TOX3 RNA expression levels were decreased in em1, em2, em5, and em6 rats

At the beginning of this study, it was hypothesized that em1, em2, em5, and em6 knockouts would have lower *TOX3* expression than WT rats. Additionally, em4 expression was hypothesized to be insignificantly different from WT samples because this strain knocked out a region further away from the first exon of *TOX3*. Furthermore, em1 and em2 were hypothesized to have the most severe mutation because em1 knocked out the promoter and first exon of *TOX3*. Similarly, em2 deleted the first methionine of the *TOX3* protein as well as the promoter. The hypothesis that em1 would have the lowest expression was supported by it having the lowest fold change in expression (7.7x10-5) of all strains. Additionally, em1 expression was found to be significantly lower than WT expression ($p=0.0097$, $\alpha=0.1$). However, em2 did not support the hypothesis of having reduced expression. Instead, em2 had a 1.713-fold-change in expression in comparison to the WT strain, and did not have a significant difference in comparison to WT expression ($p=0.2902$, $\alpha=0.1$). The hypothesis that em5 and em6 would have reduced expression was supported by their relative fold-changes in expression being 0.3231 and 0.3748 respectively. Additionally, both em5 ($p=0.0499$) and em6 ($p=0.0995$) were found to have significantly lower expression of *TOX3* in comparison to WT rats ($\alpha=0.10$). Lastly, the hypothesis that em4, would have similar *TOX3* expression to WT rats was supported with em4 rats having a 1.4074-fold-change in expression compared to WT rats. Additionally, no significant difference was found between em4 expression and WT *TOX3* expression ($p=0.1738$, $\alpha=0.10$)

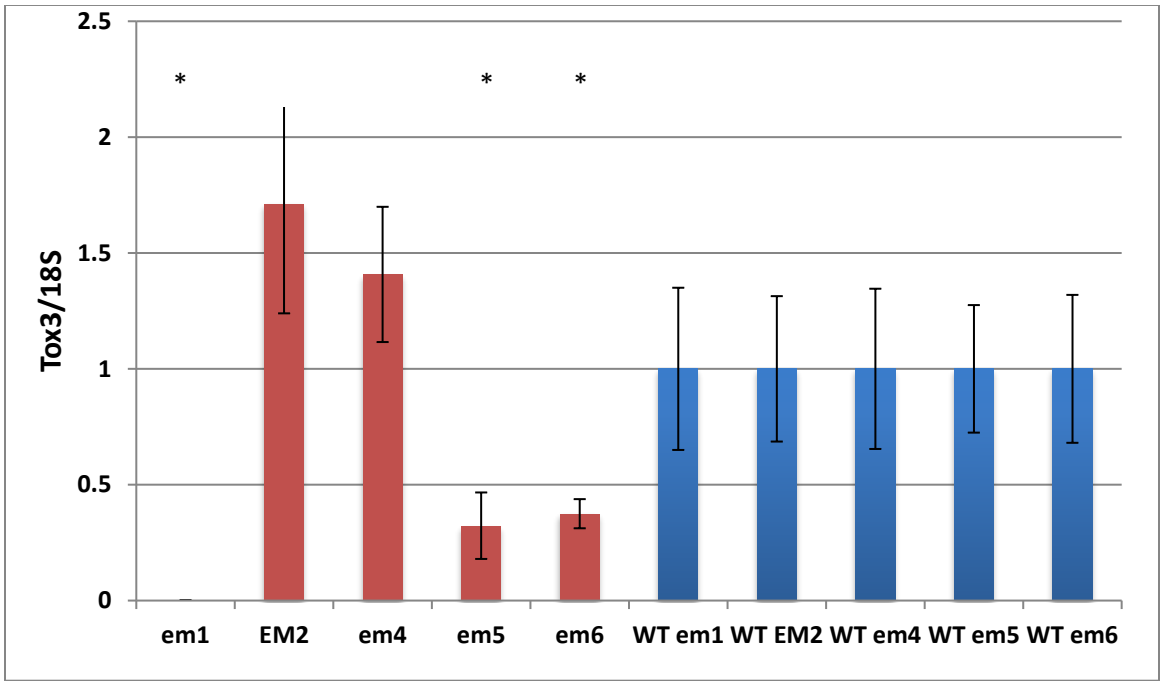

Figure 3: *TOX3/18s* expression is shown above for the five knockout strains (em1, em2, em4, em5, and em6) and their corresponding WT samples (WT em1, WT em2, WT em4, WT em5, and WT em6). WT sample groups were normalized to have *TOX3/18s* expression levels of 1. Standard error values are as follows: em1=2.4×10^{-5}, em2=0.4733, em4=0.29177, em5=0.14348, em6=0.06288, WT em1=0.3500, WT em2=0.3138, WT em4=0.3457, WT em5=0.27521, and WT em6=0.3191). Asterisks indicate significant differences in expression compared to respective WT strains.

Substantially decreased *TOX3* expression is correlated to lower body weight

Rat body weights were recorded after noticeable differences in fat contents were qualitatively observed between knockout and wild type rats. Total body weights for selected wild type (WT), <em1>, <em2>, <em4>, <em5>, and <em6> male rats were averaged and were 249.96g, 177.02g, 200.00g, 243.03g, 244.042g, and 239.38g respectively (Figure 4). A t-test with a confidence interval of 0.05 was completed to analyze if there was a significant difference between the knockout strains and WT strain average body weights. When compared to average wild type male body weight, both <em1> and <em2> rat weights were found to be significantly lower with p-values of 6.90×10^{-5} and 0.011 respectively. However, <em4>, <em5>, and <em6> weights were found to have no significant difference compared to WT average body weight with p-values of 0.339, 0.419, and 0.583 correspondingly.

Additionally, total body weight for female rats was recorded for the WT, <em1>, <em2>, <em4>, <em5>, and <em6> strains (Figure 5). The average body weights were 179.74g, 153.17g, 165.85g, 184.33g, 187.50g, and 188.31g for WT, <em1>, <em2>, <em4>, <em5>, and <em6> accordingly. A t-test with a confidence interval of 0.05 was completed to analyze if there was a significant difference between the knockout strains and WT strain average body weights. Consequently, <em1> rats were found to have significantly lower body weights than the WT strain with a p-value of 0.0038. Meanwhile, <em2> (p=0.617), <em4> (p=0.230), <em5> (p=0.525), and <em6> (p=0.528) all did not have significant differences in comparison to the WT strain.

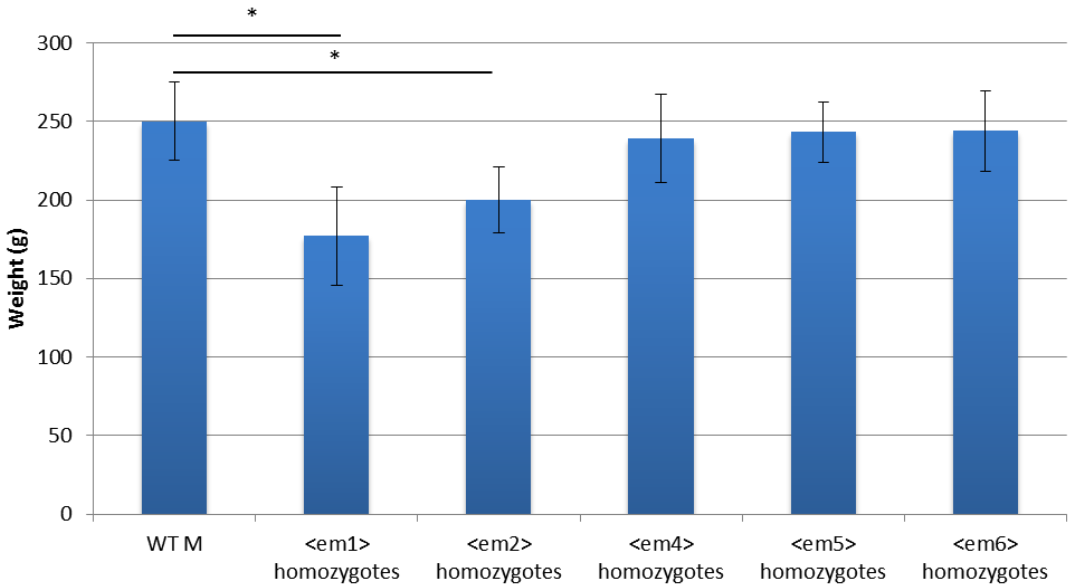

Figure 4: Averaged male body weights (g) of euthanized rats from each strain are shown above. Standard deviation values for each genotype population are as follows: WT=24.94, em1=31.38, em2=21.20, em4=28.12, em5=19.40, em6=25.54. Asterisks indicate significant differences between averaged male body weights when compared to the WT group (α=0.05)

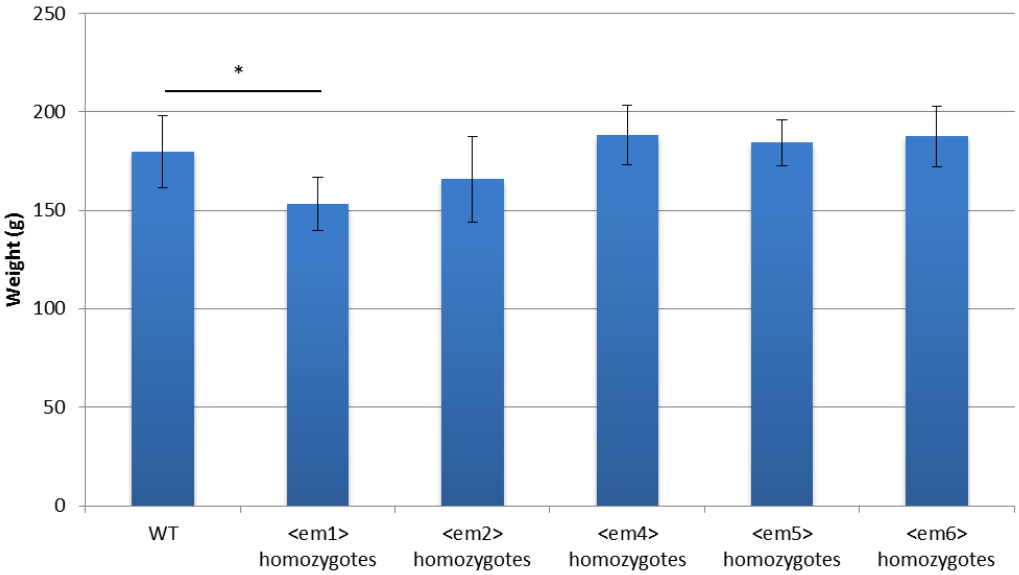

Figure 5: Averaged female body weights (g) of euthanized rats from each strain are shown above. Standard error values for each genotype population are as follows: WT=18.26, em1=13.47, em2=21.84, em4=14.92, em5=11.62, and em6=15.52 Asterisks indicate significant differences between averaged female body weights in comparison to the WT strain. (α=0.05).

Em1 and em2 rats have significantly higher adiposity

After observing bodyweight differences between strains, organ weights and adipose tissue weights were taken to further investigate why there were differences between rats. Heart, kidneys, fat pads, brain, D mammary glands, and total mammary glands were weighed for euthanized female rats. Tissue weights were normalized to total body weight for each rat. Perirenal fat averages were 0.0193g, 0.0184g, 0.011g, 0.00936g, 0.0087g, and 0.010g for em1, em2, em4, em5, em6, and WT strains subsequently. Significant differences in perirenal fat were found between em1 ($p=8.82x10^{-6}$) and em2 ($p=0.0001$) compared to the WT strain ($\alpha.=0.05$). Mammary gland average weights were 0.0247g (em1), 0.0226g (em2), 0.0108g (em4), 0.0094g (em5), 0.0087g (em6), and 0.0106g (WT). Mammary gland weights were significantly higher in em1 ($p=8.18x10^{-12}$) and em2 ($p=6.14x10^{-11}$) rats in comparison to the WT strain ($\alpha=0.01$). D-mammary gland weights were 0.0195G (em1), 0.0182G (em2), 0.0070g (em5), 0.0074g (em6), 0.00786g (em4) and 0.0086g (WT). D-mammary gland weights were significantly higher in em1 ($p=5.03x10^{-12}$) and em2 ($p=1.53x10^{-12}$) rats when compared to the WT strain glands ($\alpha=0.01$). There was no observable difference between heart, kidney and brain weights for each strain.

Additionally, tissue weights for males were recorded for the heart, kidneys, fat pads, and brain. Tissue weights were normalized to body weight. Fat pad weights were 0.0182g, 0.0155g, 0.0122g, 0.0096g, 0.0102g, and 0.0096g for WT, em1, em2, em4, em5, and em6 respectively. Both em1 ($p=0.0002$) and em2 ($p=0.0062$) were found to have significantly higher fat pad weights in comparison to the WT strain ($\alpha=0.01$). Heart, kidneys, and brain weights were found to not be significantly different between strains.

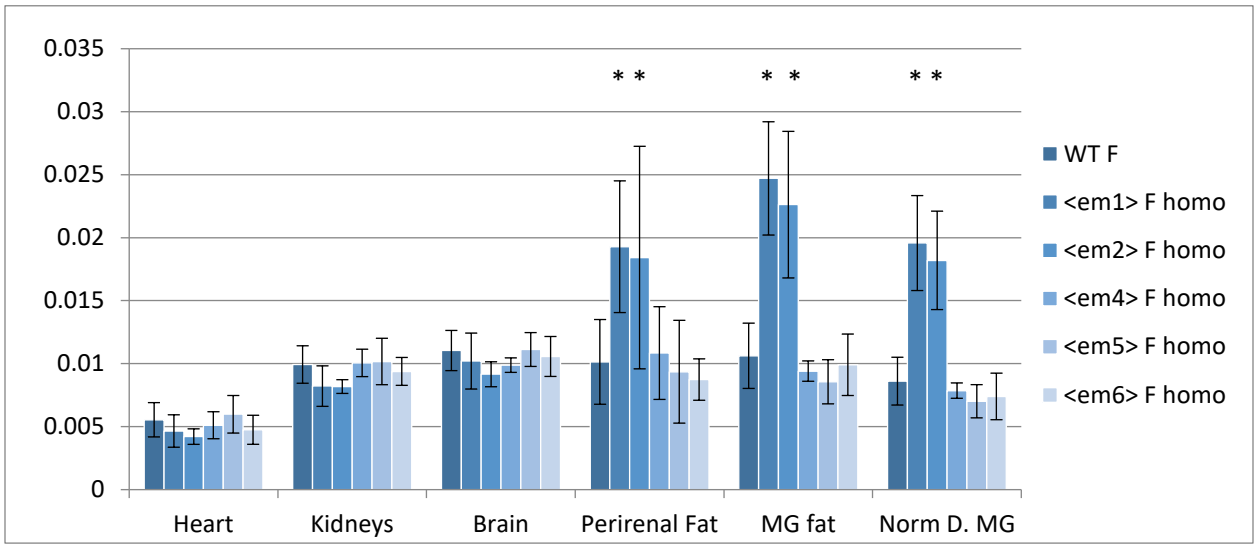

Figure 6: Tissue weights(g) for heart, kidneys, brain, perirenal fat, MG fat, and D-mammary glands are shown from left to right in the figure above for WT, em1, em2, em4, em5, and em6 strains. Asterisks indicate significant differences in tissue weight compared to WT rats. Tissue weights were normalized to body weight.

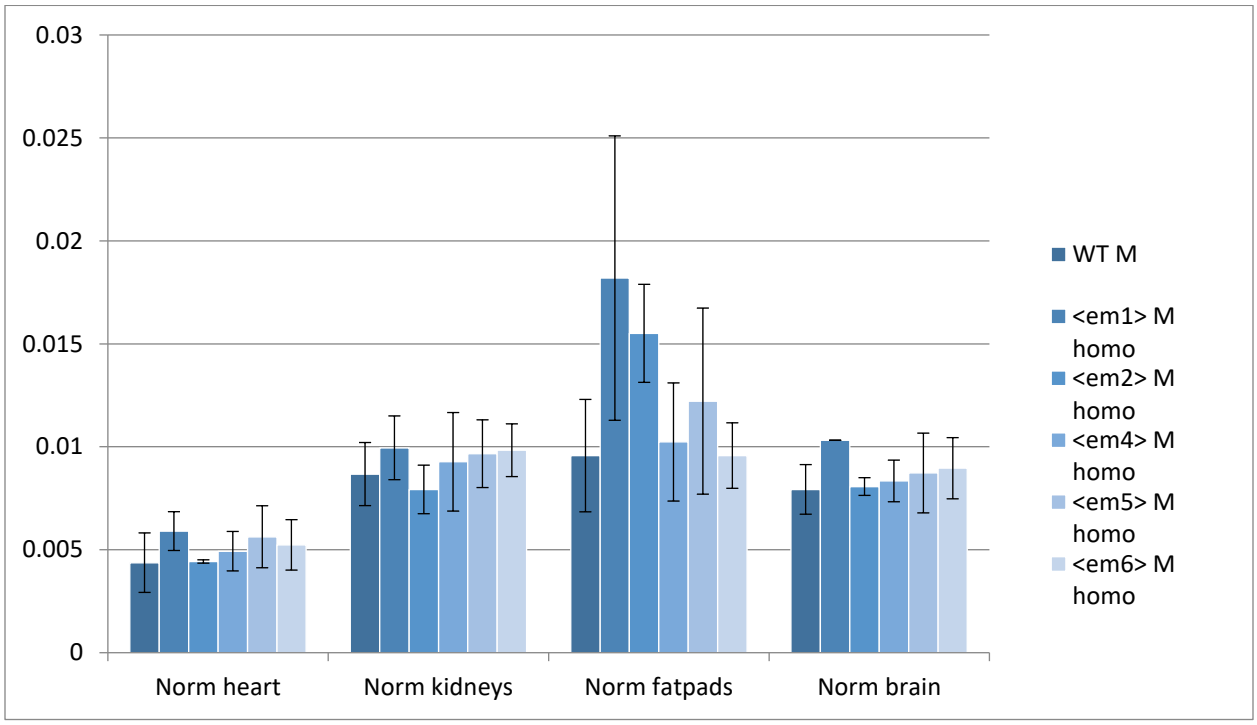

Figure 7: Tissue weights for males were taken for heart, kidneys, fat pads, and the brain. Tissue weights were normalized to total body weight. Asterisks indicate significant differences between tissue weights in comparison to the WT strain.

Successful expression of dCas-VP64 in single cell clones

Before infection with sgRNA viruses could be carried out, it needed to be determined if lentiviral delivery of dCas-VP64 and subsequent expression of dCas9 was successful in MCF-10A cells. Consequently, qPCR with Cas9RT2F2R primers and 18s was completed on H5 single cell clone, A7 single cell clone, HEK293 PX462 (+ control), MCF-10A PLIX (-) control and the MCF-10A dCas9-VP64 population of cells. CR2F2R/18s expression data was 0.305, 0.334, 1.157, 0.0093, and 4.080 for H5, A7, HEK293 PX462, MCF-10A PLIX and the MCF-10A dcas9-VP64 population of cells respectively (Figure 9). This ensured that dCas9 was present in cells to be infected with sgRNA viruses, and this finding allowed experimentation to continue as planned.

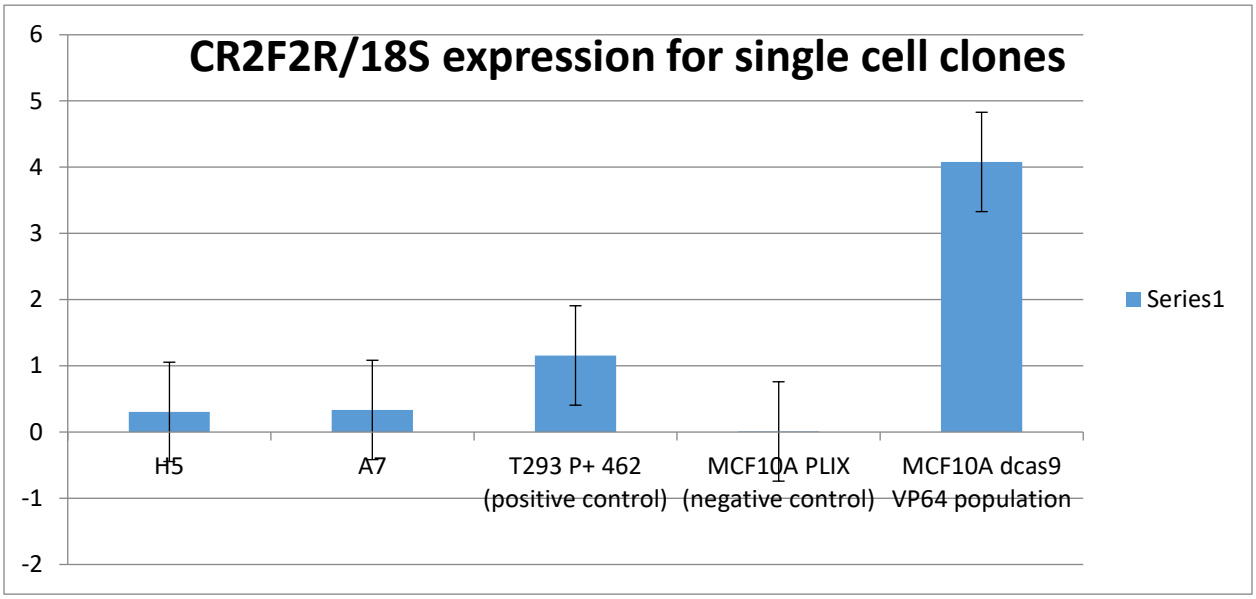

Figure 8: H5, A7, T293 PX462, MCF-10A PLIX, and MCF-10A dCas9 VP64 cell population CR2F2R/18s expression is shown above. The standard error for each sample is 0.750365.

Gene upregulation using sgRNA viruses.

Lentiviral delivery of the dCas-VP64 complex and subsequent infection with viruses expressing designed sgRNAs were hypothesized to increase *TOX3* expression levels. On 2/17/17, the first round of qPCR for 17+, 27-, 31+, NC2 (- control), and LentiGuide Puro without an sgRNA insert (- control) was completed to measure *TOX3* expression levels. These samples had RNA extracted after Puromycin selection was completed. None of the samples amplified *TOX3*.

On 2/22/17, another qPCR was run on 13+, 17+, 18+, and LentiGuide Puro using RNA extracted 24 and 72 hours after infection with each respective virus. *TOX3* was amplified for 13+ 24 hour (CT value=37.33), 18+ 24 hour (CT value=38.04), 17+ 48 hour (CT value= 37.20) and 18+ 48 hour (CT value=37.35).

On 3/14/17, qPCR was completed on 13+, 17+, 18+, with LentiGuide Puro using RNA from 48 and 72 hours after infection. Additionally, HEK293T RNA was included as a positive control for *TOX3* expression. The *TOX3/18s* expression data is shown below in figure 10. 13+ 48 hour ($3.38X10^{-5}$), 18+ 48 hour ($7.29X10^{-6}$), LGP 48 hour ($1.35X10^{-5}$), 18+72 hour ($4.62X10^{-6}$), LGP 72 hour ($1.83X10^{-7}$) and HEK293T RNA (0.916) were the only samples with *TOX3* amplified. Consequently, successful gene activation was completed using this lentiviral delivery system.

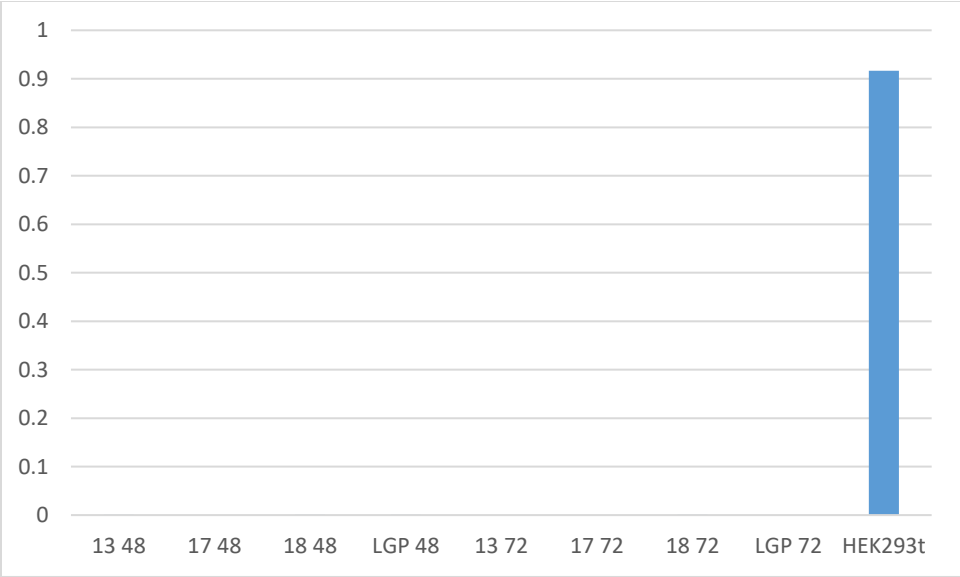

Figure 10: RNA from 48 hour and 72 hours after infection and corresponding *TOX3/18s* expression for 13+, 17+, 18+, and LGP above is shown above. HEK293T RNA was run as a positive control for *TOX3* expression.

Discussion

TOX3 Expression in rat knockouts

Analysis of *TOX3* mammary gland RNA expression found that em1, em5, and em6 had significantly decreased *TOX3* expression. Additionally, em2 was surprisingly found to have nonsignificant fold changes in expression in comparison to WT rats. Lastly, em4 expression was not significantly different from WT rats. Despite these findings, the sample size for expression analysis was rather small. Thus, more rats from each strain will need to be analyzed to increase power of detecting potential differences between rat mutants' *TOX3* expression levels, especially for em2 and em4 versus WT. The deviation between samples and small sample size could indicate why em2 did not support the hypothesis of having lower *TOX3* expression than WT rats. In contrast to em1, however, the em2 mutation does not delete the portion of the first intron important for *TOX3* expression. As em2 deletes the first Methionine essential for translational start, we think that despite the absence of transcriptional downregulation, the em2 mutation still renders the *TOX3* protein non-functional. This needs to be investigated further by testing Tox3 functionality using em2 knockout samples.

Presently, the expression effects of each strain provide a glimpse of what role each respective knockout region plays. For example, em1 and em2 knocked out a region containing SNPs that increased breast cancer risk in women with European ancestry. Since decreased *TOX3* expression increases risk for ER+ breast cancer, and *TOX3* expression was reduced in these strains; it is likely the SNPs within this region decrease *TOX3* expression. Similarly, em4, em5 and em6 knocked out regions with SNPs found to increase risk for breast cancer in women of African American ancestry. Consequently, it can be concluded that SNPs within this region likely diminish *TOX3* expression as well.

Moving forward with this study, carcinogenesis will be induced on selected rats harboring one of the mutations. Carcinogenesis will provide insight into how *TOX3* expression influences tumor initiation, tumor progression and longevity of rats within each strain. Since decreased *TOX3* expression increases risk for ER+ breast cancer and leads to poorer case outcomes, strains with lower *TOX*3 should display similar trends. While continuing the breeding for this study, it has been found that there is a fatality phenotype within the em1 and em2 strains. Thus, the expression indicates em5 and em6 are the best model for the suggested carcinogenesis study. Theoretically, em5 and em6 rats should have increased tumor progression and poorer outcomes than rats within the em4 and WT strain. Carcinogenesis will confirm if lower *TOX3* expression leads to increased risk for ER+ breast cancer.

TOX3 expression and its effect on bodyweight and adiposity

The body weights of em1 ($p=6.90x10^{-5}$) and em2 ($p=0.011$) were found to be significantly lower than WT body weights. Intriguingly, em1 and em2 rats had significantly more fat within their fat pads, mammary glands, and D-mammary glands. Therefore, em1 and em2 rats weighed less than the other rat strains, but had more body fat. Since em1, was found to have significantly reduced *TOX3* expression, it appears that *TOX3* expression and fat content are related. However, em2 rats had nonsignificant differences in levels of *TOX3* expression compared to WT rats. Thus, the idea that lower *TOX3* expression leads to more fat content is not supported by em2 *TOX3* mRNA expression levels. It is possible that the obesity phenotype is due to the em1 and em2 knockouts altering other genes' expression levels. *FTO*, *IRX3*, and *IRX5* are all genes tied to obesity located downstream of the 16q.12.1-TOX3/LOC643714 Locus. These genes are found to interact with one another through long range chromatin interactions[20]. Similar

long range chromatin interactions associated with the em1 and em2 knockouts could alter expression of said obesity genes. Accordingly, the potential alteration in expression levels of genes like *FTO, IRX3,* and *IRX5* could explain the obesity phenotype observed within this study. However, if it is found gene expression levels of *FTO, IRX3,* and *IRX5* are not significantly different from the WT rat strain, other explanations would need to be explored.

TOX3 encodes a transcription factor, and it is found in ER+ breast epithelial cells[24]. It is currently unknown how downregulation of this gene could lead to increased breast cancer risk, but one hypothesis could be that *TOX3* binds to ER-alpha and thereby regulates estrogen mediated gene expression. Lower *TOX3* would lead to increased estrogen receptor-mediated gene expression potentially leading to breast cancer development. Our rat studies add a different view to the story. Since the em2 rat strain did not follow the lower *TOX3* expression trend of em1, but exhibits the obesity phenotype like em1; the functionality of the *TOX3* protein in the mammary gland needs to be explored. Arguably, *TOX3* could be a novel obesity regulatory gene. Another explanation for the obesity phenotype in em2 is that *TOX3* protein is not transported to the nucleus to fulfill its function. Since em2 knocks out the first amino acid of *TOX3*, it is possible that this knockout affects the target peptide responsible for transporting proteins to their intended locations. Therefore, if *TOX3* is not transported into the nucleus, *TOX3* may not fulfill its intended function thereby causing the obesity phenotype seen in em1 rats. Protein quantification via Western Blot should be completed in the future to determine if mRNA levels correspond to protein expression levels for each strain.

Obesity and its relation to breast cancer risk is still not fully understood. Research thus far suggests that higher body mass index (BMI) is related to increased breast cancer risk in post-menopausal women[21]. However, the reverse is true for pre-menopausal women[21]. In pre-

menopausal women, increased BMI and risk for breast cancer are inversely correlated[21]. If *TOX3* is indeed a novel obesity gene, then *TOX3* expression and its effect on obesity could explain why decreased *TOX3* expression leads to increased risk of breast cancer. In this present study, em1 possibly explains the increase in risk for post-menopausal women with higher BMI. Since lower *TOX3* within this strain appears to be correlated to increased adiposity, resulting obesity from diminished *TOX3* expression could explain why decreased *TOX3* has been associated with increased ER+ breast cancer risk. However, this trend is not upheld in the em2 strain that is also correlated with increased adiposity. Consequently, further analysis of localization of *TOX3*, histological analysis of white adipocytes within the rat strains, and *TOX3* protein quantification will provide more insight into the complex relationship between *TOX3* expression, obesity, and breast cancer risk.

Lentiviral delivery of dCas9-VP64 and its effects on gene expression

While the rat study indicated the broad genomic location of the TOX3 regulatory elements, the purpose for the dCas9-VP64 tool in this study was to determine which of the breast cancer risk associated SNPs within the *TOX3* regulatory region have the largest influence on *TOX3* expression. Larger fold-changes in expression would indicate the relative importance of respective regulatory sites within the region. Understanding which regions of the *TOX3* regulatory region correspond to the largest increase in expression can determine which SNPs likely increase breast cancer risk the most. These SNPs can be used in future breast cancer risk screening panels to accurately determine women's genetic predisposition.

In the future, this tool can potentially be used as a therapeutic, provided the development of a safe drug-delivery system. The risk-increasing alleles SNPs can be targeted using this

dCas9-VP64 complex to adjust *TOX3* expression to normal levels. Specific regions will result in specific fold-changes in *TOX3* expression, therefore *TOX3* expression could be fine-tuned to meet levels corresponding with minimal breast cancer risk.

TOX3 expression is undetectable in MCF-10A cells[23]. Consequently, any observed indication of *TOX3* expression within the cells infected in this study would indicate the dCas9-VP64 tool was successful. In this first round of infection within this study, *TOX3* was not detected in RNA extractions from 17+, 18+, 27-, 31+, NC2+, and LGP after Puromycin selection. An infection using the 13+ sgRNA insert in the lentiGuide-Puro plasmid died off during selection, and consequently no RNA was extracted from cells infected with this construct. After obtaining this result, we hypothesized that *TOX3* was not expressed in MCF-10A cells that were selected for because *TOX3* has a cytotoxic effect on this cell line. Thus, subsequent infections were harvested after 24, 48, and 72 hours without antibiotic selection. Free-floating cells as well as adhered cells were taken for RNA extraction. Since *TOX3* was hypothesized to have a cytotoxic effect, the free-floating cells, dying cells were expected to be the cells most likely expressing *TOX3.*

We used 13+, 17+, 18+, and LGP sgRNA constructs for this infection. 13+ 24 hour (CT value=37.33), 18+ 24 hour (CT value=38.04), 17+ 48 hour (CT value= 37.20) and 18+ 48 hour (CT value=37.35) all had *TOX3* amplification after qPCR. There was no positive control for *TOX3* expression in this run. Therefore, it was not possible to normalize the data to *18s* expression and represent the values as fold-changes in expression. However, this round of qPCR and infection displayed that this technology can activate gene expression for genes that were not expressed before.

Another round of infection was completed for 13+, 17+, 18+, and LGP for RNA extracted 48 and 72 hours after infection. In addition to these samples, HEK293T RNA was included for comparison of *TOX3* expression levels. 13+ 48 hour ($3.38X10^{-5}$), 18+ 48 hour ($7.29X10^{-6}$), LGP 48 hour ($1.35X10^{-5}$), 18+ 72 hour ($4.62X10^{-6}$), LGP 72 hour ($1.83X10^{-7}$) and HEK293T RNA (0.916) were the only samples with *TOX3*/*18s* readings. The values were much lower in this round of infection than the first infection with successful *TOX3* expression. Even though *TOX3* is expressed in MCF-10A cells after infection, these levels are much lower than cells who express *TOX3* naturally and of those in the previous infection. This is displayed by the much higher *TOX3* expression levels in HEK293T cells in comparison to the MCF-10A cells used for infection. Additionally, *TOX3* readings were not expected for LGP because this lentivirus did not have a sgRNA inserted to guide the dCas-VP64 complex to a region within the genome. Thus, no increase in expression was expected for viruses using LGP. A possible explanation for why there was amplification for LGP could be that there was contamination from other cDNA samples during qPCR. The readings were also late in the qPCR cycle, indicating that these readings could in fact be false readings or products of contamination.

Furthermore, this latest round of infection was done using a population from a single cell clone (H5). As shown in figure 9, H5 had much lower expression of the dCas-VP64 construct than the population of MCF-10A dCas-VP64 cells used in previous experiments. It is possible that higher amounts of integration of the dCas-VP64 construct increase likelihood for its association with each lentiGuide-Puro sgRNA construct. Ergo, higher dCas-VP64 expression would likely lead to increased likelihood for *TOX3* expression. Therefore, if infection was repeated with a population of cells with higher dCas-VP64 expression, the likelihood of observing *TOX3* expression during qPCR would also increase.

Taking these results into account, more trials need to be completed to ensure that this gene activation tool is reliable and functioning as expected. The pre-emptive results suggest that this technology is functioning. However, activation of genes in cells where these genes are not natively expressed seems to require more stringent conditions. Further trials altering variables such as number of cells used for infection, amount of dCas-VP64 expressed, the amount of virus, and where the sgRNA guides the complex can be tweaked to increase efficiency. Additionally, of the sgRNA targets tested so far, those located closest to the promoter of *TOX3* appeared to be more likely of inducing *TOX3* expression (13+, 17+, 18+). The sgRNA targets that are located further away from the promoter, 27-, NC2+ and 31+, did not induce *TOX3* expression. Creating more viruses with sgRNA targets located closer to the promoter increase the likelihood for successful gene activation. Lastly, the cytotoxic effect of *TOX3* needs to be further investigated. When doing future infections with viruses that had been successful in activating *TOX3* previously, free-floating cells and adhered cells could have RNA extracted separately. If the free-floating, presumably dying cells, are found to have *TOX3* expressed, then it would appear that *TOX3* expression is responsible for cell death in MCF-10A cells. This would be a novel biological finding. Consequently, more research would need to be conducted to determine why *TOX3* leads to cell death in MCF-10A cells and if cell death occurs via apoptosis or another mechanism.

Expanding the use of this lentiviral gene activation is the next step to be taken from research on this project. First, a single cell clone based cell line would control for variation within the population of cells. In future infections, it would be certain that each cell would be equally likely to express *TOX3* after infection with lentiGuide-Puro derived viruses. Meanwhile, in the population of cells, it is undiscernible to tell whether the entire population of cells is

responsible for *TOX3* expression readings, or if a few outlier cells are responsible for the readings that are observed. It is quite possible that both the dCas-VP64 construct and the lentiGuide-Puro viruses both are successfully integrated in target cells, but the two constructs do not interact. Furthermore, it is possible that cells display both Blasticidin and Puromycin resistance, but have faulty plasmids incapable of increasing gene expression. Both possibilities need to be controlled for.

In any case, single cell clones with high quantifications of dCas-VP64 using qPCR outlined in this experiment would display that dCas-VP64 is present in cells awaiting infection. Preferably, single cell clones would have a dCas-VP64 expression level equivalent to the population of MCF-10A dCas-VP64 cells. Additionally, primers can be designed to target the sgRNA within each lentiGuide-Puro lentivirus and the tracrRNA that recruits the dCas-VP64 complex. Successful amplification of this sequence between these primers would indicate successful interaction between the dCas-VP64 complex and each sgRNA lentivirus. Quantification of the amount of interaction between dCas-VP64 and the sgRNA would show the infection efficiency for each virus. Theoretically, if substantial amounts of the sgRNA and the dCas-VP64 complex are associating with one another, activation of gene expression should follow.

Ultimately, this technology can be expanded to the whole genome. Anywhere the dCas-VP64 complex can be guided could result in activation of genes in the surrounding area. Targeting promoters will likely lead to the largest increase in expression as demonstrated by this study. Therefore, the next step in expanding the usefulness of this tool is to target genes that are known to be oncogenic. Previously, a collaborating lab, found that overexpression of *WHSC1L1* transforms MCF-10A cells.[22] If a virus was created to guide the dCas-VP64 complex to the

promoter of *WHSC1L1*, then gene expression would be upregulated and transformation should follow. Similarly, other genes can be targeted, and gene expression can be activated. Infected cells could then be observed to see if cells are transformed. If cells are transformed, then the genes responsible for the transformation will be coined oncogenic. Therefore, this tool can be expanded as a fast and efficient screening tool to discover new genes that play a role in breast cancer risk and initiation. On the other hand, this technology can be used as drug therapy to activate genes that are under expressed and increase likelihood of breast cancer risk. The ramifications of this technology could be critical in understanding which genes directly contribute to breast cancer development. Additionally, this technology could serve as the best way to prevent breast cancer through correcting gene expression. Furthermore, this technology can be expanded to other cell lines to address cancer risks specific to tissue type.

Supplemental Figures

Primer	Sequence 5'--->3'
TOX3 Forward	cggtactgccgacacttgt
TOX3 Reverse	tcttccttccatcacagtctcaa
18s Forward	gtaacccgttgaaccccatt
18s Reverse	ccatccaatcggtagtagcg
Lsm1 Forward	tgtggagcgtattcatgtggc
Lsm1 Reverse	agcttggtctgctgttccaccc
Cas9 RT Forward	cctgagcgcctctatgatca
Cas9 RT Reverse	cagcaggtcctctgtgttca
MYC Forward	cttcctcatcttcttgctcttct
MYC Reverse	ttctctccttcctcggactc
Fam84B Forward	atcctacctgctccctaagtc
Fam84B Reverse	ccatttagtattcctgggcctt

Supplemental Figure 1: Primer sequences for qPCR used within this study are shown above.

Name	A OR B	sgRNA_min	sgRNA_plus
hTOX3_17	A	CACCGCCCATTTGCGAAGAAAGTAC	CACCGTCGCTAAGAGACAGCTATAG
	B	AAACGTACTTTCTTCGCAAATGGG	AAACCTATAGCTGTCTCTTAGCGA
hTOX3_18	A	CACCGTATATAGATCTGTCATAGA	CACCGTGATAAACATCCATGTTTCT
	B	AAACTCTATGACAGATCTATATAC	AAACAGAAACATGGATGTTTATCA
hTOX3_31	A	CACCGTTCCCATTCTTGCAGCTAGC	CACCGTACTGCATTCAGCTTTGGG
	B	AAACGCTAGCTGCAAGAATGGGAAC	AAACCCCAAAGCTGAATGCAGTAC
hTOX3_10	A	CACCGCGGGGAAGCTGTGGTCGCGC	CACCGTGACCCCCTTCCTTCTTCAT
	B	AAACGCGCGACCACAGCTTCCCCG	AAACATGAAGAAGGAAGGGGGTCA
hTOX3_13	A	CACCGCTAACTCCTGAGAGCCTTAG	CACCGCTAACTCCTGGCCATTGTTT
	B	AAACCTAAGGCTCTCAGGAGTTAG	AAACAAACAATGGCCAGGAGTTAG
hTOX3_27	A	CACCGATAATCCATGCGATATTTCT	CACCGTAAAGTCCCCAGCAGACACC
	B	AAACAGAAATATCGCATGGATTATC	AAACGGTGTCTGCTGGGGACTTTAC

Supplemental Figure 2: Sequences for *TOX3* sgRNA inserts into lentiGuide-Puro

References

1 Altekruse, S. F. *et al.* SEER Cancer Statistics Review, 1975-2007. (2010).

2 National Breast Cancer Organization. About Breast Cancer. (2015).

3 Schnitt, S. J. Classification and prognosis of invasive breast cancer: from morphology to moleular taxonomy. (2010) *Mod Pathol* **23**, 60-64, doi:10.1038/modpathol.2010.33 (2010).

4 Howlader N, Altekruse SF, Li CI, et al. US incidence of breast cancer subtypes defined by joint hormone receptor and HER2 status. J Natl Cancer Inst. 106(5), 2014.

5 Chen, F. *et al.* Fine-mapping of breast cancer susceptibility loci characterizes genetic risk in African Americans. *Hum Mol Genet* **20**, 4491-4503, doi:10.1093/hmg/ddr367 (2011).

6 Huo, D. *et al.* Evaluation of 19 susceptibility loci of breast cancer in women of African ancestry. *Carcinogenesis* **33**, 835-840, doi:10.1093/carcin/bgs093 (2012).

7 Ruiz-Narvaez, E. A. *et al.* Polymorphisms in the TOX3/LOC643714 locus and risk of breast cancer in African-American women. *Cancer Epidemiol Biomarkers Prev* **19**, 1320-1327, doi:10.1158/1055-9965.EPI-09-1250 (2010).

8 Ruiz-Narvaez, E. A. *et al.* Fine-mapping of the 6q25 locus identifies a novel SNP associated with breast cancer risk in African-American women. *Carcinogenesis* **34**, 287-291, doi:10.1093/carcin/bgs334 (2013).

9 Zheng, W. *et al.* Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat Genet* **41**, 324-328 (2009).

10 Albain, K. S., Unger, J. M., Crowley, J. J., Coltman, C. A., Jr. & Hershman, D. L. Racial disparities in cancer survival among randomized clinical trials patients of the Southwest Oncology Group. *J Natl Cancer Inst* **101**, 984-992, doi:10.1093/jnci/djp175 (2009).

11 Easton, D. F. *et al.* Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**, 1087-1093 (2007).

12 Gudmundsdottir, E. T. et al. The risk allele of SNP rs3803662 and the mRNA level of its closest genes TOX3 and LOC643714 predict adverse outcome for breast cancer patients. BMC cancer 12, 621, doi:10.1186/1471-2407-12-621 (2012).

13 Cowper-Sal lari, R. et al. Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. Nat Genet 44, 1191-1198, doi:10.1038/ng.2416 (2012).

14 Riaz, M. et al. Correlation of breast cancer susceptibility loci with patient characteristics, metastasis-free survival, and mRNA expression of the nearest genes. Breast Cancer Res Treat 133, 843-851, doi:10.1007/s10549-011-1663-3 (2012).

15 Gudmundsdottir, E. T. et al. The risk allele of SNP rs3803662 and the mRNA level of its closest genes TOX3 and LOC643714 predict adverse outcome for breast cancer patients. BMC cancer 12, 621, doi:10.1186/1471-2407-12-621 (2012).

16 Seksenyan, A. et al. TOX3 is expressed in mammary ER(+) epithelial cells and regulates ER target genes in luminal breast cancer. BMC cancer 15, 22, doi:10.1186/s12885-015-1018-2 (2015).

17 Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819-823, doi:10.1126/science.1231143 (2013).
18 Mali, P. et al. RNA-guided human genome engineering via Cas9. Science 339, 823-826, doi:10.1126/science.1232033 (2013).
19 Maeder M.L., S.J. Linder, V.M. Cascio, Y. Fu, Q.H. Ho, J.K. Joung. CRISPR RNA-guided activation of endogenous human genes. 2013. Nat Methods. 10(10): 977-979.
20 Claussnitzer, M., et al. (2015). "FTO Obesity Variant Circuitry and Adipocyte Browning in Humans." New England Journal of Medicine **373**(10): 895-907.
21 Harvie, M., et al. (2003). "Central obesity and breast cancer risk: a systematic review." Obesity Reviews **4**(3): 157-173.
22 Irish, J. C., et al. (2016). "Amplification of WHSC1L1 regulates expression and estrogen-independent activation of ERα in SUM-44 breast cancer cells and is associated with ERα over-expression in breast cancer." Molecular Oncology 10(6): 850-865.
23 Kiselev, V. Y., et al. (2015). "Perturbations of PIP3 signalling trigger a global remodelling of mRNA landscape and reveal a transcriptional feedback loop." Nucleic Acids Research 43(20): 9663-9679.
24 Seksenyan, A., et al. (2015). "TOX3 is expressed in mammary ER+ epithelial cells and regulates ER target genes in luminal breast cancer." BMC Cancer **15**(1): 22.