Characterization of the function and therapeutic potential of the non-coding mammary carcinoma susceptibility loci *Mcs5c* and *Mcs1a*

By

Amanda N. Henning

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Genetics)

at the

UNIVERSITY OF WISCONSIN-MADISON

2016

Date of final oral examination: 3/1/2016

The dissertation is approved by the following members of the Final Oral Committee:

Michael Gould, Professor, Oncology
Amy Moser, Associate Professor, Human Oncology
Akihiro Ikeda, Associate Professor, Medical Genetics
Mark Burkard, Associate Professor, Medicine
William Dove, Professor Emeritus, Oncology and Medical Genetics

Abstract

A growing number of genetic breast cancer risk factors have been identified through the use of familial breast cancer case studies, candidate gene approaches, and genome-wide association studies (GWAS). The disease variants these approaches have identified are the basis for our current risk assessment strategies, and they have played a pivotal role in our initial understanding of breast cancer pathogenesis. Increasingly, the importance of common, low-penetrant variants to breast cancer risk has been recognized. However, these approaches are limited in their ability to identify and characterize such risk alleles. The Gould lab has undertaken a comparative genomics approach, utilizing the laboratory rat for the identification and characterization of breast cancer risk alleles. The ultimate goal of these studies being to translate laboratory findings back to human patients. Quantitative trait loci analysis in the rat has identified a number of Mammary carcinoma susceptibility, *Mcs*, loci capable of affecting carcinoma multiplicity in a carcinogen-induced model. My thesis work has focus on furthering the characterization of two of these loci – *Mcs5c* and *Mcs1a*. Work on the *Mcs5c* locus has focused on the functional characterization of its method of action, and has revealed a novel association between its activity and the windows of susceptibility (WOS) phenomenon observed in humans and rats. *Mcs5c* acts to regulate nearby gene expression through tissue-specific chromatin folding and allele-dependent DNA methylation that is regulated in a WOS-specific manner. This is the first molecular mechanism associated with the WOS phenomenon, and represents a significant finding in the field. Work on the *Mcs1a* locus has focused on validating the therapeutic relevance of its target gene *Nr2f1/NR2F1* for the treatment of triple negative breast cancer (TNBC). Indeed, we have shown that *NR2F1*-overexpression promotes a reduction in TNBC tumor growth, and appears to do so through the promotion of a luminal molecular profile. Luminal-like breast cancers tend to be less aggressive and more treatable than their TNBC counterparts, and this “luminalization” effect represents a novel therapeutic strategy. Together, my work on the *Mcs5c* and *Mcs1a* loci shows the potential of comparative genomic studies to deliver

translational findings that can be used in a prevention and therapeutic capacity, and provides useful models with which to test future, clinically-relevant hypotheses.

Acknowledgements

I would like to thank my committee members, Dr. Akihiro Ikeda, Dr. Bill Dove, Dr. Amy Moser, and Dr. Mark Burkard for their helpful comments and suggestions throughout the course of this project.

Thank you to the entire Gould lab, both past and present members, for their assistance with projects, troubleshooting advice, willingness to teach me new techniques, reviews of grant and manuscript drafts, and for their friendship over these past years. Coming to work every day has been a pleasure because of all of you.

A big thanks to Dr. Michael Gould for his guidance, support, and encouragement throughout the entirety of this project. Thank you for always seeing things in a positive light, especially when I found it challenging to do so.

Thank you to all of my graduate school friends for their friendship, advice, and support. Thanks for always being up for celebrating when the results were good, and for commiserating with me when the results were bad. A huge thank you to my family for their unwavering support and encouragement. I wouldn't be where I am today without it. Lastly, I'd like to thank my husband Matt for his support, and for being a great partner in life.

Table of Contents

Figures and Tables

Chapter 1 – Introduction

Chapter 2 – The non-coding Mammary carcinoma susceptibility locus, *Mcs5c*, functions in a mammary cell-autonomous manner to regulate *Pappa* expression via age-specific chromatin folding and allele-dependent DNA methylation

Chapter 4 – Conclusions and Future Directions

Chapter 1 – Introduction

In the United States, breast cancer is the most frequently diagnosed cancer and second leading cause of cancer death among women [1]. Its etiology is complex, consisting of both genetic and environmental risk factors. Early breakthroughs in understanding genetic contributions to breast cancer risk came from the analysis of patients with an extensive family history of breast cancer, so called familial breast cancer, leading to the discovery of the *BRCA1* and *BRCA2* susceptibility genes [2,3]. Women carrying a mutation in *BRCA1* or *BRCA2* have an average cumulative breast cancer risk by age 70 of 65% and 45%, respectively [4], and represent a high-risk population compared to the 12.3% average lifetime risk of the general public [1]. While *BRCA1/2* are therefore classified as high-penetrance alleles, their frequency in the population is rare (<1%) [5] and they therefore only account for an estimated 25% of familial and 1-2% of all breast cancer cases [6,7]. Additional familial studies led to the discovery of other high-penetrant, low-frequency mutations in genes for *TP53* [8] and *STK11* [9]. Candidate gene association studies have resulted in the identification of additional low-frequency risk alleles of moderate penetrance in *ATM* [10], *CHEK2* [11], *BRIP1* [12], *PALB2* [13], and *CASP8* [14]. However, candidate gene approaches have had limited results overall [15]. The rare nature of these alleles can lead to underpowered studies, and the requirement of *a priori* knowledge can be limiting. Altogether, risk alleles found through family-based and candidate gene approaches are only able to explain a small percentage of breast cancer cases [6]. It was therefore hypothesized that the missing heritability for breast cancer and other common diseases could be attributable to common variants, generally single nucleotide polymorphisms (SNPs), with low penetrance, the so called 'common disease, common variant' hypothesis [16,17]. Modeling studies on breast cancer support this hypothesis, and suggest that the presence of multiple common, low-penetrant variants may contribute in an additive manner to overall breast cancer risk [18,19]. The advent of genome wide association studies (GWAS) has made it possible to detect these types of risk-associated SNPs in a global, unbiased approach, although large

case/control cohorts are still ideal. To date, GWAS have been successful in identifying 83 risk loci, however, the functional relevance of many of the identified SNPs remains elusive [20]. While many of the SNPs identified by GWAS may be functionally relevant, others may simply be markers for causal SNPs in linkage disequilibrium. Additionally, for SNPs located in non-coding regions of the genome, their function is not readily apparent. The identification and characterization of casual risk-associated SNPs will be crucial for increasing our risk assessment capabilities, and in understanding the genetic pathways associated with breast cancer pathogenesis for therapeutic purposes. Clearly, while GWAS have been crucial in the identification of high-frequency, low-penetrant risk variants, they have a limited ability to discern causality, and additional methods are needed for functional characterization.

In an effort to identify and rigorously characterize common, low-penetrant breast cancer risk alleles, the Gould lab has taken a comparative genomics approach using the laboratory rat as a model. The rat makes for an excellent model as it mimics human breast cancer in terms of carcinoma morphology, histopathological progression, and hormone responsiveness [21–23]. Additionally, different inbred rat strains vary considerably in their susceptibility to spontaneous and induced mammary cancer [24], making for an ideal model in which to study the genetics of breast cancer susceptibility. This variation in susceptibility to chemically induced mammary cancer was exploited by our lab in a number of quantitative trait loci (QTL) analysis studies. The susceptible Wistar-Furth (WF) and resistant Copenhagen (COP) and Wistar-Kyoto (WKy) rat strains were used in linkage mapping to identify QTL associated with mammary carcinoma susceptibility in a 7,12-dimethylbenz-[*a*]anthracene (DMBA)-induced model. The first QTL analysis performed on (WF x COP)F1 x WF backcrossed rats identified 4 mammary carcinoma susceptibility (*Mcs*) loci, with the COP alleles of *Mcs1-3* acting to decrease and *Mcs4* acting to increase carcinoma number [25,26]. QTL analysis on (WF x WKy)F1 x WF

backcrossed rats identified 4 additional *Mcs* loci, with the WKy alleles of *Mcs5*, *-6*, and *-8* decreasing and *Mcs7* increasing carcinoma number [27].

Subsequent efforts in the lab have focused on the characterization of the *Mcs1* and *Mcs5* loci. Fine-mapping studies have revealed the existence of subloci within both *Mcs1* and *Mcs5*. The *Mcs1* locus is comprised of *Mcs1a*, *-1b*, and *-1c* which each act independently to reduce carcinoma number by 60% in congenic WF rats homozygous for the COP *Mcs* allele [28]. Fine mapping of the *Mcs5* locus identified *Mcs5a*, *-5b*, and *-5c*, the interactions of which proved more complex than the *Mcs1* subloci. On their own, *Mcs5a* and *Mcs5c* were found to reduce carcinoma number in congenic WF rats homozygous for the WKy allele, while *Mcs5b* increased carcinoma number. *Mcs5a* was found to act in an epistatic manner to mask the effects of *Mcs5b,* and *Mcs5a* acts in a synergistic manner with *Mcs5c* to decrease carcinoma number above what would be predicted by an additive model [29]. Clearly, a comparative genomics approach is capable of identifying complex genetic architecture akin to the types of interactions observed in humans. My thesis work has focused on the further characterization of the *Mcs1a* and *Mcs5c* loci and, as will be outlined below, both demonstrate the ability of a comparative genomics approach to be useful in the discernment of susceptibility loci relevant to human breast cancer. More broadly, they illustrate the potential for 'bench to bedside' experimentation in a mammalian model organism.

The *Mcs5c* locus: A study on the impact of age on risk locus functionality

As previously mentioned, breast cancer risk is influenced by both genetic and environmental factors. Interestingly, the contribution of these factors to overall risk can vary depending on the developmental context of the individual. In this way, there are multiple windows during a woman's life where she is more susceptible to genetic susceptibilities and environmental insults and behaviors that

will affect her long-term breast cancer risk. The time periods of birth to young adulthood, post-pregnancy, and menopause represent three such windows of susceptibility [30–32]. The most well documented of these windows is the birth-to-young-adult window of susceptibility (byaWOS), where the effect of age-specific radiation exposure is readily apparent. In studies of women exposed to radiation during the atomic bombings of Hiroshima and Nagasaki, women 20 years or younger at time of exposure had an increased risk of developing breast cancer later in life compared to women exposed at older ages [33]. Studies on women exposed to radiation for the treatment of Hodgkin's lymphoma found a similar result. Women 30 years or younger at time of treatment showed an increased relative risk of developing breast cancer, with risk being highest for those exposed before age 15 [34]. While the byaWOS encompasses a broad age range, studies seem to suggest that risk may be variable even within this time period, with possible sub-windows of childhood, adolescence, and young adulthood. Additional studies point to the complex relationship that exists between environmental risk factors and various windows of susceptibility. For example, studies on BMI and weight gain suggest that higher BMI in adolescence may be protective against future breast cancer diagnosis, while weight gain in postmenopausal women is associated with increased risk [35–37]. Pregnancy itself represents a variable risk factor, as early pregnancy is more protective compared to late-aged pregnancy, while the period following pregnancy represents a general WOS [31,38]. Additionally, the windows of susceptibility themselves likely interact with genetic factors that affect, among other things, hormone and growth factor levels that play crucial roles during breast development, pregnancy, post-pregnancy involution, and menopause [32,38,39]. While numerous associations between window-specific exposure and breast cancer risk have been observed, the cellular mechanisms governing their interactions with environmental and genetic risk factors are poorly understood. This is particularly true for the byaWOS, as the affected age group represents a vulnerable population not generally included in research studies

or amenable to tissue sampling. Also, most studies involving the byaWOS are retrospective, relying on recall of past behaviors and events, which may be subject to error.

In an effort to understand the mechanisms underlying the byaWOS, comparative genomic studies in the rat make for a highly relevant model as it too exhibits the WOS phenomenon. This was first demonstrated by Huggins *et al* [40], where a single dose of the carcinogen 3-Methylcholanthrene (3-MC) was fed to outbred Sprague-Dawley (SD) rats at ages ranging from 23 to 100 days old. Researchers found that carcinoma incidence increased with greater age at feeding, peaking for rats fed 3-MC between 50 and 65 days old, and dropping off considerably for 100 day old rats. A similar result was seen in a study of DMBA-induced mammary cancer, although carcinoma incidence remained high for all rats treated between 30 to 55 days of age, with carcinomas/rat peaking at 46 days at treatment [41]. Data from our lab suggests that rat windows of susceptibility may be influenced by strain specific differences and carcinogen class. Studies using the inbred Fischer 344 (F344) rat strain, showing an intermediate susceptibility to mammary cancer, found differences in carcinoma incidence and carcinomas/rat depending on the age of carcinoma treatment and the class of carcinogen used. Specifically, when the alkylating agent *N*-nitroso-*N*-methylurea (NMU) was administered at 21 days of age, carcinoma incidence and carcinomas/rat were significantly higher than when administration occurred at 56 days. The results were reversed for the polycyclic aromatic hydrocarbon DMBA, as 56 days represented the most susceptible induction period [42]. Additional studies in the lab looking at the susceptible WF rat strain have replicated these DMBA findings, showing a significant reduction in carcinomas/rat when administration occurred at 32 or 93 days versus 56 days (Fig 1; unpublished data). Clearly, while windows of susceptibility exist in the rat, it is important to interpret results in a strain- and carcinogen-specific manner.

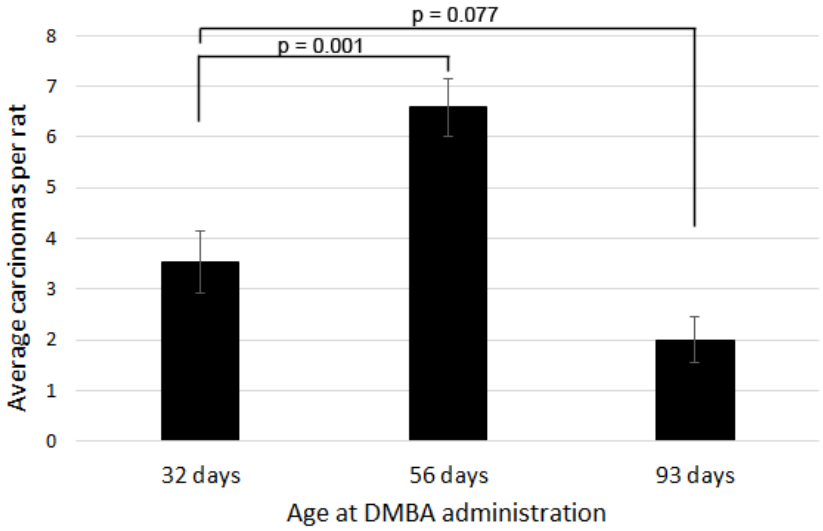


Fig 1: The susceptible WF rat strain displays the WOS phenotype. Susceptible Wistar-Furth (WF) rats were administered DMBA as a single oral dose dissolved in sesame oil at 65mg/kg of body weight to induce mammary carcinoma formation. DMBA was administered when animals were 32 (n = 13), 56 (n = 36), or 93 (n = 8) days of age to gage the boundaries of the WOS phenotype. Carcinomas greater than 3x3mm were counted at necropsy, 15 weeks post-DMBA administration. P-values were obtained using the non-parametric Mann-Whitney U test.

It has been speculated that the reason for such age-specific susceptibilities is due to the proliferative index and differentiation of terminal end buds (TEBs) within the mammary gland. In SD rats, TEBs reach their maximum density at 21 days, with peak DNA synthesis occurring at 40 days, roughly corresponding to the initiation of estrous cycling. From 21 days on, the TEBs are extending into the fat pad and differentiating into the functional structures of the mammary gland, alveolar buds (ABs) and terminal ducts (TDs), and this process is generally finished by 84 days of age [41,43]. This time period would roughly capture the entirety of the human byaWOS, and the data presented above further suggests that there may exist two distinct windows within this time period, namely, the immature (iWOS) and adolescent window of susceptibility (aWOS). In the immature mammary glands of young rats prior to estrous cycling, TEBs are at a maximum and are poised for future mitosis and differentiation. This may be representative of pre-pubescent human breast tissue, where the epithelial tissue is similarly poised for hormonally-induced growth at puberty. In adolescent rat mammary glands, beginning with the onset of estrous cycling, growth and differentiation intensifies and is impacted by each successive

estrous cycle. This is similar to the extensive human breast development that occurs in a hormone-dependent manner beginning with puberty and continuing until the first full-term pregnancy [21,23,44,45]. Thus, not only does the rat recapitulate human breast development and the byaWOS phenotype, but it provides an opportunity to study the possibility of distinct sub-windows within the broad byaWOS.

The importance of the byaWOS became apparent during the course of *Mcs5c* characterization. Previous work on *Mcs5c* had fine-mapped the locus to a 170kb region of rat chromosome *5* (Fig 2), lying in a 1Mb gene desert that shares homology with both mice and humans (see Chapter 2, Fig 1). Rats homozygous for the resistant WKy allele introgressed on a WF background had a 50% and 40% reduction in carcinoma number for DMBA- or *HER2/neu*-induced mammary cancer, respectively [46]. Initial data from Veillet *et al* [46] identified the gene Tenascin C (*Tnc*) as a possible target of *Mcs5c* activity, with differential expression observed in non-mammary gland tissues following carcinogen exposure. As will be outlined in Chapter 2, our mammary gland transplant experiments show that *Mcs5c* acts within the mammary gland to affect carcinoma number, and therefore shed doubt on the relevance of these previously observed expression differences. A reevaluation of *Mcs5c* genotype-dependent gene expression was conducted, with a focus on expression in the mammary gland. Of the genes nearest to the *Mcs5c* locus, Pregnancy-associated plasma protein A (*Pappa*) stood out as being of potential interest. Pappa/PAPP-A is a zinc metalloproteinase originally isolated from the plasma of pregnant women and later identified as a component of the Insulin-like growth factor 1 and 2 (IGF-I/II) signaling pathways [47,48]. Specifically, PAPP-A cleaves IGF binding proteins 2, 4 and 5 (IGFBP2/4/5), which are responsible for binding IGF in the extracellular space and thereby preventing signaling through the IGF receptor (IGF1R) [49–51]. Cleavage of the binding proteins by PAPP-A releases bound IGF, and therefore acts to positively regulate IGF bioavailability and signaling [48,52]. While the specific role of PAPP-A in

normal breast development has not been studied, the IGF-I pathway, in general, is an essential component of breast/mammary gland development, as evidence by the severe mammary gland defects of *Igf-I* and *Igf1r* knockout mice [53–55]. Given the extreme relevance of *Igf-I* to mammary gland development, we focused our expression analysis on young rats within the iWOS and aWOS, speculating that if *Pappa* was the target gene, it may be important during this developmental period. Indeed, we have identified two mechanisms by which *Mcs5c* regulates the expression of *Pappa* in the mammary gland during the aWOS, and have evidence to suggest that a different mechanism is at work during the iWOS. Our results therefore not only support *Pappa* as an important mediator of mammary/breast cancer susceptibility and suggest its potential as a therapeutic target, but also support the delineation of the immature and adolescent windows of susceptibility.

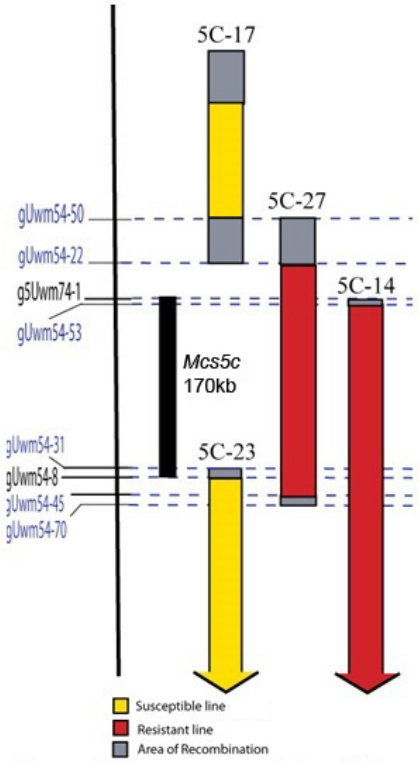


Fig 2: Chromosome *5* congenic lines defining the *Mcs5c* locus. Congenic lines containing a section of the resistant WKy genome introgressed on the susceptible WF background are shown. Susceptible and resistant lines are indicated by yellow and red bars, respectively. The *Mcs5c* locus spans 170kb and is defined by markers g5Uwm74-1 and gUwm54-8. Congenic line 5C-27 is the smallest resistant line that contains the *Mcs5c* locus (275kb).

The *Mcs1a* locus: Establishing the therapeutic potential of a susceptibility locus target gene

The ultimate goal of comparative genomic studies is to translate laboratory findings back to human patients in the so-called bench-to-bedside model. In the case of the *Mcs* loci, our aim is to identify prevention and therapeutic strategies through the functional characterization of their method of action. The *Mcs1a* locus epitomizes this bench-to-bedside strategy, as a logical next step in therapeutic potential became apparent following its characterization. Previously published work from the lab by Smits *et al* [56] had fine-mapped the *Mcs1a* locus to a 277kb region located in a 3Mb gene desert on rat chromosome *2* (Fig 3). Congenic rat lines homozygous for the resistant COP allele introgressed on a WF background showed an approximately 50% reduction in carcinoma number in DMBA-, NMU-, and *HER2/neu*-induced mammary cancer models. *Mcs1a* was found to function in the mammary gland, where it regulates expression of the target gene *Nr2f1/Coup-TFI*. The resistant *Mcs1a* allele contains a variant which disrupts a conserved Nr2f1 binding site involved in self-repression of the *Nr2f1* gene, resulting in increased *Nr2f1* expression in resistant rats. Resistant rats also showed an increase in the luminal population of mammary epithelial cells (MECs) compared to susceptible rats, and these cells showed a decrease in proliferative capacity, suggesting that increased *Nr2f1* expression in the mammary gland may affect cell proliferation and differentiation, shifting cells to a luminal fate. RNA-seq analysis on mammary glands from a mouse megadeletion model, resulting in a severe knockdown of *Nr2f1* expression, revealed that *Nr2f1* was anti-correlated with genes involved in the cell cycle, proliferation, and DNA-damage response. These data were recapitulated in publically available breast cancer microarrays, where not only was *NR2F1* expression anti-correlated with similar classes of genes, but also with histological grade and carcinoma sub-type. Specifically, *NR2F1* expression was lower in the triple negative clinical class of tumors. Altogether, characterization of the *Mcs1a* locus found that it regulates *Nr2f1* expression in the rat mammary gland, with increased expression in resistant rats leading to an increased luminal cell population that has a decreased proliferative capacity. Global expression

analyses in mice and humans suggested that increased *Nr2f1/NR2F1* expression decreases the expression of proliferation-associated genes, consistent with the observations in the rat.

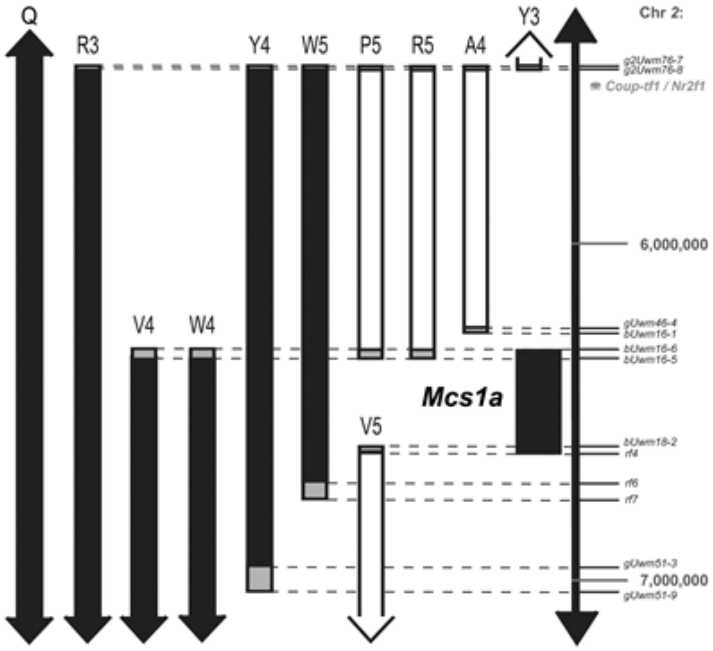


Fig 3. Chromosome 2 congenic lines defining the *Mcs1a* locus. Shown is a genetic map of the congenic lines contributing to the positional identification of the *Mcs1a* locus on rat chromosome 2. Each congenic line, as defined by genotyping the genetic markers indicated along the vertical scale bar, represents a segment from the resistant Cop inbred strain introgressed onto the susceptible WF genetic background. Susceptible and resistant congenic lines are indicated by the white and black bars, respectively. These lines were used to define the critical interval for the *Mcs1a* resistance allele to the 277kb region indicated. The coordinates (in bp) along the vertical axis are from the 2004 version of the rat genome (UCSC Genome Browser, rn4) (figure adapted from Smits *et al*, 2013).

The association of *NR2F1* with the clinical subtype of triple negative breast cancer (TNBC) made it an attractive candidate for potential therapeutic purposes. TNBCs account for approximately 15-20% of all breast cancer cases, and are so-called as they lack expression of estrogen and progesterone receptors (ER and PR) and human epidermal growth factor receptor 2 (HER2) amplification [57]. While these tumors are molecularly heterogeneous, a majority of them fall into the basal subtype, with

expression profiles that mimic basal epithelial and myoepithelial breast cells [58]. TNBCs occur more frequently in younger women and African American patients. They tend to be more aggressive, with a higher histological grade and worse prognosis when compared to other clinical subtypes [57,59–61]. The general therapeutic course of action for TNBC is systemic cytotoxic chemotherapy, and while initial response rates are comparable to non-TNBC, 3-year overall survival rates were lower and recurrence rates were higher in TNBC [62]. To achieve the success that hormone receptor and HER2 antagonists have had in the treatment of ER/PR+ and HER2+ breast cancer, similarly targeted therapies for TNBC need to be developed [63].

Our characterization of the *Mcs1a* locus led to the hypothesis that its target gene, *Nr2f1*, may be a good candidate for TNBC-specific drug development. NR2F1 is an orphan nuclear receptor, with sequence analysis placing it into the steroid/thyroid hormone receptor superfamily of proteins. NR2F1 has the highest binding affinity for an AGGTCA direct repeat with a 1 nucleotide spacer, however, binding to imperfect repeats with variable spacer lengths also occurs. This consensus site overlaps with the response elements for the retinoid X receptor (RXR), retinoic acid receptor (RAR), thyroid hormone receptor (TR), vitamin D receptor (VDR), peroxisome proliferator-activated nuclear receptor (PPAR), and ER, resulting in competition for binding and thereby affecting the expression of target genes in these pathways. Additionally, NR2F1 frequently dimerizes with RXR, RAR, or TR, and such an association tends to disrupt their function [64,65]. Although generally thought of as a negative regulator of transcription, NR2F1 can also act as an activator, and its regulatory action seems to be dependent on interactions with other transcription factors or coactivators present in the cell [65–68]. *Nr2f1* is highly expressed in a distinct pattern within the central nervous system during mouse embryonic development, and is also expressed in other developing organs, specifically those that require epithelial proliferation and differentiation [64,69,70]. Knock-out mice die perinatally with severe neuronal developmental defects

[71]. Together, these data suggest that *Nr2f1* plays an important role in organ development and differentiation, particularly for neuronal tissues. Indeed, *Nr2f1* was found to promote cell cycle exit and differentiation when ectopically expressed in the dorsal cortex of transgenic mice [72].

Our work on the *Mcs1a* locus, described above, has shown that *Nr2f1* may also play a role in the development and differentiation of the mammary gland. *Mcs1a*-mediated increases in *Nr2f1* mammary gland expression appear to promote a mammary carcinoma-resistant phenotype through reduced epithelial cell proliferation and promotion of the luminal cell fate. Breast cancers whose expression profiles overlap with those for luminal breast cells tend to have a more favorable clinical prognosis [73]. We therefore hypothesized that increasing *NR2F1* expression in basal-like, TNBC may have therapeutic potential through a reduction in tumor proliferation and/or promotion of the less-aggressive luminal phenotype. In the work to be discussed in Chapter 3, we found that *NR2F1*-overexpression does indeed slow the growth of a TNBC cell line in a xenograft model. Although RNA-sequencing analysis of the carcinomas does not show a strong reduction in proliferative-associated genes, it does appear that these tumors have increased expression of genes associated with luminal breast cancers, and has provided interesting targets for future therapeutic intervention. This work shows the potential of translating comparative genomic studies into clinically relevant results, and represents the next step in taking the *Mcs1a* findings from 'bench to bedside'.

References

1. American Cancer Society. Breast Cancer Facts & Figures 2015-2016. Atlanta: 2015. http://www.cancer.org/acs/groups/content/@research/documents/document/acsp

2. Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science 1994;266:66–71.

3. Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, et al. Identification of the breast cancer susceptibility gene BRCA2. Nature 1995;378:789–92. doi:10.1038/378789a0.

4. Antoniou A, P Pharoah PD, Narod S, Risch HA, Eyfjord JE, Hopper JL, et al. Average Risks of Breast and Ovarian Cancer Associated with BRCA1 or BRCA2 Mutations Detected in Case Series Unselected for Family History: A Combined Analysis of 22 Studies. Am J Hum Genet 2003;72:1117–30.

5. Peto J, Collins N, Barfoot R, Seal S, Warren W, Rahman N, et al. Prevalence of BRCA1 and BRCA2 gene mutations in patients with early-onset breast cancer. J Natl Cancer Inst 1999;91:943–9.
6. Melchor L, Benítez J. The complex genetic landscape of familial breast cancer. Hum Genet 2013;132:845–63. doi:10.1007/s00439-013-1299-y.

7. Easton DF. How many more breast cancer predisposition genes are there? Breast Cancer Res 1999;1:14–7.

8. Malkin D, Li FP, Strong LC, Fraumeni JF, Nelson CE, Kim DH, et al. Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. Science 1990;250:1233–8.

9. Hemminki A, Markie D, Tomlinson I, Avizienyte E, Roth S, Loukola A, et al. A serine/threonine kinase gene defective in Peutz-Jeghers syndrome. Nature 1998;391:184–7. doi:10.1038/34432.

10. Renwick A, Thompson D, Seal S, Kelly P, Chagtai T, Ahmed M, et al. ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. Nat Genet 2006;38:873–5. doi:10.1038/ng1837.

11. Meijers-Heijboer H, van den Ouweland A, Klijn J, Wasielewski M, de Snoo A, Oldenburg R, et al. Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. Nat Genet 2002;31:55–9. doi:10.1038/ng879.

12. Seal S, Thompson D, Renwick A, Elliott A, Kelly P, Barfoot R, et al. Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. Nat Genet 2006;38:1239–41. doi:10.1038/ng1902.

13. Rahman N, Seal S, Thompson D, Kelly P, Renwick A, Elliott A, et al. PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. Nat Genet 2007;39:165–7. doi:10.1038/ng1959.

14. Cox A, Dunning AM, Garcia-Closas M, Balasubramanian S, Reed MWR, Pooley KA, et al. A

common coding variant in CASP8 is associated with breast cancer risk. Nat Genet 2007;39:352–8. doi:10.1038/ng1981.

15. Zhang B, Beeghly-Fadiel A, Long J, Zheng W. Genetic variants associated with breast-cancer risk: comprehensive research synopsis, meta-analysis, and epidemiological evidence. Lancet Oncol 2011;12:477–88. doi:10.1016/S1470-2045(11)70076-6.

16. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature 2009;461:747–53. doi:10.1038/nature08494.

17. Lander ES. The new genomics: global views of biology. Science 1996;274:536–9.

18. Antoniou AC, Pharoah PDP, Mcmullan G, Day NE, Ponder BAJ, Easton D, et al. Evidence for Further Breast Cancer Susceptibility Genes in Addition to BRCA1 and BRCA2 in a Population-Based Study. Genet Epidemiol 2001;21:1–18.

19. Dudbridge F, Fletcher O, Walker K, Johnson N, Orr N, Dos Santos Silva I, et al. Estimating causal effects of genetic risk variants for breast cancer using marker data from bilateral and familial cases. Cancer Epidemiol Biomarkers Prev 2012;21:262–72. doi:10.1158/1055-9965.EPI-11-0719.

20. Fachal L, Dunning AM. From candidate gene studies to GWAS and post-GWAS analyses in breast cancer. Curr Opin Genet Dev 2015:32–41. doi:10.1016/j.gde.2015.01.004.

21. Russo IH, Russo J. Mammary gland neoplasia in long-term rodent studies. Environ Health Perspect 1996;104:938–67.

22. Russo J, Russo IH. Experimentally induced mammary tumors in rats. Breast Cancer Res Treat 1996;39:7–20.

23. Russo J. Significance of Rat Mammary Tumors for Human Risk Assessment. Toxicol Pathol 2015;43:145–70. doi:10.1177/0192623314532036.

24. Isaacs JT. Genetic control of resistance to chemically induced mammary adenocarcinogenesis in the rat. Cancer Res 1986;46:3958–63.

25. Hsu LC, Kennan WS, Shepel LA, Jacob HJ, Szpirer C, Szpirer J, et al. Genetic identification of Mcs-1, a rat mammary carcinoma suppressor gene. Cancer Res 1994;54:2765–70.

26. Shepel LA, Lan H, Haag JD, Brasic GM, Gheen ME, Simon JS, et al. Genetic identification of multiple loci that control breast cancer susceptibility in the rat. Genetics 1998;149:289–99.

27. Lan H, Kendziorski CM, Haag JD, Shepel LA, Newton MA, Gould MN. Genetic loci controlling breast cancer susceptibility in the Wistar-Kyoto rat. Genetics 2001;157:331–9.

28. Haag JD, Shepel LA, Kolman BD, Monson DM, Benton ME, Watts KT, et al. Congenic rats reveal three independent Copenhagen alleles within the Mcs1 quantitative trait locus that confer resistance to mammary cancer. Cancer Res 2003;63:5808–12.

29. Samuelson DJ, Aperavich BA, Haag JD, Gould MN. Fine mapping reveals multiple loci and a possible epistatic interaction within the mammary carcinoma susceptibility quantitative trait locus, Mcs5. Cancer Res 2005;65:9637–42. doi:10.1158/0008-5472.CAN-05-1498.

30. Biro FM, Deardorff J. Identifying opportunities for cancer prevention during preadolescence and adolescence: Puberty as a window of susceptibility. J Adolesc Heal 2013;52:S15–20. doi:10.1016/j.jadohealth.2012.09.019.

31. Schedin P. Pregnancy-associated breast cancer and metastasis. Nat Rev Cancer 2006;6:281–91. doi:10.1038/nrc1839.

32. Hankinson SE. Endogenous hormones and risk of breast cancer in postmenopausal women. Breast Dis 2005;24:3–15.

33. Land CE, Tokunaga M, Koyama K, Soda M, Preston DL, Nishimori I, et al. Incidence of female breast cancer among atomic bomb survivors, Hiroshima and Nagasaki, 1950-1990. Radiat Res 2003;160:707–17. doi:10.1667/RR3082.

34. Hancock SL, Tucker MA, Hoppe RT. Breast cancer after treatment of Hodgkin's disease. J Natl Cancer Inst 1993;85:25–31.

35. Berkey CS, Willett WC, Frazier AL, Rosner B, Tamimi RM, Colditz GA. Prospective study of growth and development in older girls and risk of benign breast disease in young women. Cancer 2011;117:1612–20. doi:10.1002/cncr.25692.

36. Han D, Nie J, Bonner MR, McCann SE, Muti P, Trevisan M, et al. Lifetime adult weight gain, central adiposity, and the risk of pre- and postmenopausal breast cancer in the Western New York exposures and breast cancer study. Int J Cancer 2006;119:2931–7. doi:10.1002/ijc.22236.

37. Magnusson C, Baron J, Persson I, Wolk A, Bergström R, Trichopoulos D, et al. Body size in different periods of life and breast cancer risk in post-menopausal women. Int J Cancer 1998;76:29–34.

38. Lyons TR, Schedin PJ, Borges VF. Pregnancy and breast cancer: when they collide. J Mammary Gland Biol Neoplasia 2009;14:87–98. doi:10.1007/s10911-009-9119-7.

39. Kleinberg DL, Barcellos-Hoff MH. The Pivotal Role of Insulin-Like Growth Factor I in Normal Mammary Development. Endocrinol Metab Clin North Am 2011;40:461–71. doi:10.1016/j.ecl.2011.06.001.

40. HUGGINS C, GRAND LC, BRILLANTES FP. Mammary cancer induced by a single feeding of polymucular hydrocarbons, and its suppression. Nature 1961;189:204–7.

41. Russo J, Wilgus G, Russo IH. Susceptibility of the Mammary Gland to Carcinogenesis. Am J Pathol 1979;96:721–33.

42. Ariazi JL, Haag JD, Lindstrom MJ, Gould MN. Mammary glands of sexually immature rats are more susceptible than those of mature rats to the carcinogenic, lethal, and mutagenic effects of N-

nitroso-N-methylurea. Mol Carcinog 2005;43:155–64. doi:10.1002/mc.20104.

43. Russo IH, Russo J. Developmental stage of the rat mammary gland as determinant of its susceptibility to 7,12-dimethylbenz[a]anthracene. J Natl Cancer Inst 1978;61:1439–49.

44. Russo J, Gusterson BA, Rogers AE, Russo IH, Wellings SR, van Zwieten MJ. Comparative study of human and rat mammary tumorigenesis. Lab Investig 1990;62:244–78.

45. Howard BA, Gusterson BA. Human breast development. J Mammary Gland Biol Neoplasia 2000;5:119–37. doi:10.1016/j.semcdb.2012.03.013.

46. Veillet AL, Haag JD, Remfert JL, Meilahn AL, Samuelson DJ, Gould MN. Mcs5c: A mammary carcinoma susceptibility locus located in a gene desert that associates with tenascin c expression. Cancer Prev Res 2011;4:97–106. doi:10.1158/1940-6207.CAPR-10-0187.

47. Lin TM, Galbert SP, Kiefer D, Spellacy WN, Gall S. Characterization of four human pregnancy-associated plasma proteins. Am J Obstet Gynecol 1974;118:223–36.

48. Conover CA, Bale LK, Overgaard MT, Johnstone EW, Laursen UH, Füchtbauer E-M, et al. Metalloproteinase pregnancy-associated plasma protein A is a critical growth regulatory factor during fetal development. Development 2004;131:1187–94. doi:10.1242/dev.00997.

49. Monget P. Pregnancy-Associated Plasma Protein-A Is Involved in Insulin-Like Growth Factor Binding Protein-2 (IGFBP-2) Proteolytic Degradation in Bovine and Porcine Preovulatory Follicles: Identification of Cleavage Site and Characterization of IGFBP-2 Degradation. Biol Reprod 2003;68:77–86. doi:10.1095/biolreprod.102.007609.

50. Lawrence JB, Oxvig C, Overgaard MT, Sottrup-Jensen L, Gleich GJ, Hays LG, et al. The insulin-like growth factor (IGF)-dependent IGF binding protein-4 protease secreted by human fibroblasts is pregnancy-associated plasma protein-A. Proc Natl Acad Sci U S A 1999;96:3149–53. doi:10.1073/pnas.96.6.3149.

51. Laursen LS, Overgaard MT, Soe R, Boldt HB, Sottrup-Jensen L, Giudice LC, et al. Pregnancy-associated plasma protein-A (PAPP-A) cleaves insulin-like growth factor binding protein (IGFBP)-5 independent of IGF: Implications for the mechanism of IGFBP-4 proteolysis by PAPP-A. FEBS Lett 2001;504:36–40. doi:10.1016/S0014-5793(01)02760-0.

52. Laursen LS, Kjaer-Sorensen K, Andersen MH, Oxvig C. Regulation of Insulin-Like Growth Factor (IGF) Bioactivity by Sequential Proteolytic Cleavage of IGF Binding Protein-4 and -5. Mol Endocrinol 2007;21:1246–57. doi:10.1210/me.2006-0522.

53. Kleinberg DL, Wood TL, Furth PA, Lee A V. Growth hormone and insulin-like growth factor-I in the transition from normal mammary development to preneoplastic mammary lesions. Endocr Rev 2009;30:51–74. doi:10.1210/er.2008-0022.

54. Ruan W, Kleinberg DL. Insulin-like growth factor I is essential for terminal end bud formation and ductal morphogenesis during mammary development. Endocrinology 1999;140:5075–81. doi:10.1210/endo.140.11.7095.

55. Bonnette SG, Hadsell DL. Targeted Disruption of the IGF-I Receptor Gene Terminal End Buds. Endocrinology 2001;142:4937–45.

56. Smits BMG, Haag JD, Rissman AI, Sharma D, Tran A, Schoenborn AA, et al. The Gene Desert Mammary Carcinoma Susceptibility Locus Mcs1a Regulates Nr2f1 Modifying Mammary Epithelial Cell Differentiation and Proliferation. PLoS Genet 2013;9. doi:10.1371/journal.pgen.1003549.

57. Toss A, Cristofanilli M. Molecular characterization and targeted therapeutic approaches in breast cancer. Breast Cancer Res 2015;17:60. doi:10.1186/s13058-015-0560-9.

58. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest 2011;121:2750–67. doi:10.1172/JCI45014.

59. Dent R, Trudeau M, Pritchard KI, Hanna WM, Kahn HK, Sawka CA, et al. Triple-negative breast cancer: clinical features and patterns of recurrence. Clin Cancer Res 2007;13:4429–34. doi:10.1158/1078-0432.CCR-06-3045.

60. Morris GJ, Naidu S, Topham AK, Guiles F, Xu Y, McCue P, et al. Differences in breast carcinoma characteristics in newly diagnosed African-American and Caucasian patients: a single-institution compilation compared with the National Cancer Institute's Surveillance, Epidemiology, and End Results database. Cancer 2007;110:876–84. doi:10.1002/cncr.22836.

61. Dai X, Li T, Bai Z, Yang Y, Liu X, Zhan J, et al. Breast cancer intrinsic subtype classification, clinical use and future trends. Am J Cancer Res 2015;5:2929–43.

62. Liedtke C, Mazouni C, Hess KR, André F, Tordai A, Mejia JA, et al. Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. J Clin Oncol 2008;26:1275–81. doi:10.1200/JCO.2007.14.4147.

63. Saurel CA, Patel TA, Perez EA. Changes to adjuvant systemic therapy in breast cancer: a decade in review. Clin Breast Cancer 2010;10:196–208. doi:10.3816/CBC.2010.n.027.

64. Pereira FA, Qiu Y, Tsai MJ, Tsai SY. Chicken ovalbumin upstream promoter transcription factor (COUP-TF): Expression during mouse embryogenesis. J Steroid Biochem Mol Biol 1995;53:503–8. doi:10.1016/0960-0760(95)00097-J.

65. Tsai SY, Tsai M-J. Chick Ovalbumin Upstream Promoter-Transcription Factors (COUP-TFs): Coming of Age. Endocr Rev 1997;18:229–40.

66. Pipaón C, Tsai SY, Tsai MJ. COUP-TF upregulates NGFI-A gene expression through an Sp1 binding site. Mol Cell Biol 1999;19:2734–45.

67. Hall RK, Sladek FM, Granner DK. The orphan receptors COUP-TF and HNF-4 serve as accessory factors required for induction of phosphoenolpyruvate carboxykinase gene transcription by glucocorticoids. Proc Natl Acad Sci U S A 1995;92:412–6.

68. Power SC, Cereghini S. Positive regulation of the vHNF1 promoter by the orphan receptors COUP-TF1/Ear3 and COUP-TFII/Arp1. Mol Cell Biol 1996;16:778–91.

69. Qiu Y, Cooney AJ, Kuratani S, DeMayo FJ, Tsai SY, Tsai MJ. Spatiotemporal expression patterns of chicken ovalbumin upstream promoter-transcription factors in the developing mouse central nervous system: evidence for a role in segmental patterning of the diencephalon. Proc Natl Acad Sci U S A 1994;91:4451–5.

70. Pereira F a, Tsai MJ, Tsai SY. COUP-TF orphan nuclear receptors in development and differentiation. Cell Mol Life Sci 2000;57:1388–98.

71. Qiu Y, Pereira FA, DeMayo FJ, Lydon JP, Tsai SY, Tsai MJ. Null mutation of mCOUP-TFI results in defects in morphogenesis of the glossopharyngeal ganglion, axonal projection, and arborization. Genes Dev 1997;11:1925–37. doi:10.1101/gad.11.15.1925.

72. Faedo A, Tomassy GS, Ruan Y, Teichmann H, Krauss S, Pleasure SJ, et al. COUP-TFI coordinates cortical patterning, neurogenesis, and laminar fate and modulates MAPK/ERK, AKT, and β-catenin signaling. Cereb Cortex 2008;18:2117–31. doi:10.1093/cercor/bhm238.

73. Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001;98:10869–74. doi:10.1073/pnas.191367098.

Chapter 2

The non-coding Mammary carcinoma susceptibility locus, *Mcs5c*, functions in a mammary cell-autonomous manner to regulate *Pappa* expression via age-specific chromatin folding and allele-dependent DNA methylation

Amanda N. Henning[1], Jill D. Haag[1], Bart M. G. Smits[2], Michael N. Gould[1]*

[1]Department of Oncology, McArdle Laboratory for Cancer Research, University of Wisconsin – Madison School of Medicine and Public Health, Madison, WI, USA

[2]Department of Pathology and Laboratory Medicine, Hollings Cancer Center, Medical University of South Carolina, Charleston, SC, USA

*Corresponding author: gould@oncology.wisc.edu

This work was supported by the NIH R01-CA077494, NIEHS U01-ES019466, and DOD W81XWH-11-0161. Support for the University of Wisconsin Carbone Cancer Center (UWCCC) Shared Services used to complete this research is provided by NIH/NCI P30 CA014520. Trainee support for ANH was provided by the University of Wisconsin Predoctoral Training Program in Genetics NIH 5 T32 GM007133-39. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

In understanding the etiology of breast cancer, the contributions of both genetic and environmental risk factors are further complicated by the impact of breast developmental stage. Specifically, the time period ranging from childhood to young adulthood represents a critical developmental window in a woman's life when she is more susceptible to environmental hazards that may affect future breast cancer risk. Although the effects of environmental exposures during particular developmental Windows of Susceptibility (WOS) are well documented, the genetic mechanisms governing these inherent susceptibilities are largely unknown. To begin to understand the function of genetic risk factors in the context of WOS, we utilized the Mammary Carcinoma Susceptibility 5c, *Mcs5c*, congenic rat model of breast cancer. In this model, rats harboring the resistant Wistar-Kyoto *Mcs5c* allele exhibit a 55% reduction in DMBA-induced carcinomas compared to rats with the susceptible Wistar-Furth allele. We have functionally characterized the *Mcs5c* locus *in vivo* at various stages of mammary gland development using quantitative real-time PCR, chromosome conformation capture, and bisulfite pyrosequencing. *Mcs5c* acts within the mammary gland to regulate expression of the neighboring gene *Pappa* during a critical mammary developmental time period in the rat which corresponds to the human young adult WOS. The *Pappa* gene has been shown to positively regulate the *IGF* signaling pathway, which is both required for proper mammary gland/breast development and is of increasing interest in breast cancer pathogenesis. *Mcs5c* mediated regulation of *Pappa* appears to occur through age-dependent and mammary gland-specific chromatin looping, as well as genotype-dependent CpG island shore methylation. This represents, to our knowledge, the first insight into cellular mechanisms underlying the WOS phenomenon and represents a novel model for further investigation into how environmental factors, together with genetic factors, modulate breast cancer risk in the context of breast developmental stage.

Author Summary

A woman's lifetime risk of developing breast cancer is affected by both genetic and environmental risk factors that can be influenced by breast developmental stage. Indeed, there are certain periods of a woman's life when she is more susceptible to environmental insults that may affect her future breast cancer risk. Such time periods are referred to as windows of susceptibility (WOS), and although the effects of WOS-specific exposure to environmental hazards are well documented, the molecular mechanisms governing these windows are unknown. Our work presented here on the characterization of the rat Mammary Carcinoma Susceptibility 5c, *Mcs5c*, locus has identified a region within *Mcs5c* that interacts with the neighboring gene, *Pappa*, in an age-dependent manner to influence gene expression via genotype-dependent DNA methylation. Importantly, *Mcs5c*-mediated gene regulation occurs specifically within a WOS, and these finding represent the first identified molecular mechanisms by which a WOS influences the ability of a locus to affect mammary/breast cancer risk. This work highlights the importance a developmental stage can have on genetic risk factor function and we anticipate that the *Mcs5c* locus will serve as a model for future studies on windows of susceptibility in combination with genetic and environmental risk factors.

Introduction

In the United States, breast cancer is the most frequently diagnosed cancer and second leading cause of cancer death among women [1]. Its etiology is complex, consisting of the interaction of both genetic and environmental risk factors whose contribution to overall risk can vary depending on the developmental context of the individual. In general, time periods in which women are more susceptible to initiating events affecting their long term breast cancer risk are broadly referred to as windows of susceptibility (WOS) [2]. In humans, the best documentation of a WOS can be found in studies of radiation exposure in women. Women exposed to radiation between 0 and 30 years of age during either

the atomic bombings of Japan or for the treatment of Hodgkin's lymphoma had an increased risk of developing breast cancer later in life compared to women >30 years of age at time of exposure [3,4]. This time period, therefore, represents one of the WOS, and encompasses ages spanning childhood, adolescence, and young adulthood in women. Animals studies performed in rats to model the human WOS phenomenon [5] further suggest the existence of at least two mechanistically distinct susceptibility windows within the larger human WOS, namely, the sexually immature WOS (iWOS) and the adolescent WOS (aWOS). This division of the WOS is most evident in work by Ariazi *et al* [6] on a carcinogen-inducible model of breast cancer, where administration to developmentally immature (3 week) and adolescent-aged (7 week) rats resulted in differential carcinoma development depending on age of administration and the carcinogen used. Additionally, although over 80 genetic loci effecting breast cancer susceptibility have been identified in human genome-wide association studies (GWAS) [summarized in 7], their function in relation to developmental stages has not been characterized. In general, while the effects of window specific exposures are well documented, the cellular mechanisms responsible for their function and governing their interactions with environmental and genetic risk factors are poorly understood.

To begin to understand the complex interactions between WOS, genetics, and the environment, we turned to a comparative genomics approach, utilizing a rat model of breast cancer. The rat is an excellent model for this type of study, as not only does its mammary gland and mammary tumor development mimic that of the human condition [8], but, as previously mentioned, it too displays the WOS phenomenon [5,6]. Additionally, inbred rat strains vary in their susceptibility to carcinogen-induced mammary cancer, allowing for the identification of genetic susceptibility loci through quantitative trait loci (QTL) analysis. This approach was applied in our lab, utilizing the mammary cancer resistant Wistar-Kyoto (WKy) and susceptible Wistar-Furth (WF) inbred rat strains resulting in the

identification and subsequent fine-mapping of the Mammary Carcinoma Susceptibility 5c, *Mcs5c*, locus [9-11]. *Mcs5c* maps to a 170kb region located in a large gene desert on rat chromosome *5* that shares homology with mice and humans (Fig 1). In both chemical carcinogen and oncogene-induced models of mammary cancer, congenic lines homozygous for the resistant WKy *Mcs5c* allele showed an approximately 50% reduction in carcinoma number compared to susceptible WF-homozygous controls [11].

Fig 1: Comparative genomic map of the *Mcs5c* locus. The rat *Mcs5c* locus, along with neighboring genes, is shown together with orthologous regions in the mouse and human genomes. Map coordinates for the rat, mouse, and human were obtained via the UCSC Genome Browser using assemblies from March 2012 (rn5), December 2011 (mm10), and February 2009 (hg19), respectively. The WKy-homozygous region of resistant congenic line 5C-27 is shown in light gray relative to the rat *Mcs5c* locus. The remainder of the 5C-27 line is WF-homozygous.

Using the *Mcs5c* locus as a model, we sought to examine the interaction between a genetic risk factor and WOS. We have characterized a temporal control element (TCE) within *Mcs5c* affecting the expression of neighboring gene Pregnancy-associated plasma protein A, *Pappa*, in a genotype-dependent manner in mammary epithelial cells (MECs). The function of the Pappa/PAPP-A protein makes it an attractive candidate for involvement in both the WOS phenomenon and breast cancer development. PAPP-A is a protease that acts to positively regulate bioavailability and signaling of the Insulin-like growth factors, IGFs, through the cleavage of IGF binding proteins 2, 4 and 5, IGFBP2/4/5 [12-15]. The specific role of PAPP-A in normal breast development has not been studied, but the IGF-I pathway, in general, is an essential component of breast/mammary gland development, as evident by the severe mammary gland defects of *Igf-I* and *Igf-I* receptor (*Igf1r*) knockout mice [16-18]. The role of

IGF-I in breast cancer development is supported by numerous studies which associate the IGF-I signaling pathway with breast cancer initiation and progression [19]. Indeed, in transgenic mice, overexpression of IGF-I in the mammary gland resulted in increased susceptibility and decreased latency to spontaneous and carcinogen-induced mammary adenocarcinomas [20]. Limited studies of PAPP-A function in cancer have demonstrated that increased PAPP-A activity enhanced tumor growth in ovarian and lung cancer cell lines [21,22], and inhibition of its proteolytic function reduced tumor growth in a murine mammary cancer cell line [23].

In this study, we have identified age-specific differences in *Mcs5c* activity which support the existence of mechanistically distinct susceptibility windows. We have functionally characterized the non-coding *Mcs5c* locus, finding that it acts within the aWOS to regulate *Pappa* expression through age-dependent chromatin looping and genotype-dependent DNA methylation. To our knowledge, this study represents the first identification of a molecular genetic mechanism underlying the aWOS phenomenon and emphasizes the ability of developmental age to influence the activity of a susceptibility locus.

Results

Mcs5c acts in a mammary gland autonomous manner

To determine if *Mcs5c* exerts its effect on carcinoma multiplicity via the mammary gland, transplant experiments were performed. Donor mammary gland tissue from either the *Mcs5c* resistant 5C-27 line or a *Mcs5c* susceptible control line was transplanted onto the interscapular fat pad of recipient rats from both genotypes, creating four donor-recipient groups. This direct transplant design allowed for the detection of mammary gland-host interactions and did not result in differential tissue rejection rates, as the lines are isogenenic except at the *Mcs5c* locus. Transplant tissue rejection rates were not statistically significant between transplant groups consisting of donors and recipients with the

same genotype versus groups with different genotypes (Chi-squared test, X = .10, df = 1, p-value = 0.75). Results from the mammary gland transplant experiment are shown in Fig 2. Resistant and susceptible rats receiving resistant donor tissue had a transplant site carcinoma incidence of 21% and 27%, respectively (n=76, 49), while resistant and susceptible rats receiving susceptible tissue had incidences of 42% and 38%, respectively (n=69, 39). Recipient rats of either genotype that received susceptible donor tissue had higher transplant site carcinoma incidences than those that received tissue from resistant rats. In this way, the carcinoma phenotype was dependent on the donor tissue genotype and was not influenced by the recipient genotype, suggesting that *Mcs5c* acts within the mammary gland. Indeed, logistic regression analysis found a statistically significant donor effect (p-value = 0.0043; recipient effect p-value = 0.825). Thus, it was concluded that *Mcs5c* acts in a mammary gland autonomous manner to influence carcinoma multiplicity.

Fig 2: *Mcs5c* acts in a mammary gland autonomous manner to influence carcinoma multiplicity. The four mammary gland transplant groups are listed on the x-axis (R = *Mcs5c* resistant 5C-27 line, S = *Mcs5c* susceptible control line), with the genotype of the donor listed first and the genotype of the recipient listed second. The number of animals per transplant group were: R->R, n = 76; R->S, n = 49; S->R, n = 69; S->S, n = 39. The y-axis indicates the percentage of animals in each group that had one or more carcinomas at the transplant site 15 weeks after DMBA administration. Logistic regression analysis found a statistically significant donor effect (p-value = 0.0043; recipient effect p-value = 0.825).

Pappa is differentially expressed in MECs in an age-dependent manner

Quantitative real-time PCR (qPCR) was used to investigate expression levels of nearby genes in mammary epithelial cells (MECs) of *Mcs5c* resistant and susceptible rats at 4 – 12 weeks of age. This age range was chosen as it captures multiple mammary gland developmental windows, including the iWOS

(4 weeks), aWOS (6 -9 weeks), and adult (12 weeks) time periods. *Pappa*, located over 517kb downstream of *Mcs5c*, was found to be differentially expressed in MECs (Fig 3). Compared to *Mcs5c* resistant rats, *Pappa* expression was increased in susceptible rats by 43% at 6 weeks (Mann-Whitney U test, p-value = 0.015, n = 13 and 15, respectively), 14% at 7 weeks (p-value = 0.05, n = 23 and 19), and 31% at 9 weeks (p-value = 0.0003, n = 23 and 18). Differential expression disappeared by 12 weeks of age (n = 9 and 18), at which point the mammary gland is fully developed and rats are past the aWOS stage [5]. Expression trends were reversed in 4 week old rats, with susceptible animals showing a sharp decrease in expression relative to resistant rats (p-value = 8e-5, n = 9 and 8, respectively). It appears that *Mcs5c* influences *Pappa* expression in MECs in an age-specific manner and functions during both the iWOS and aWOS. Differential expression in MECs was not observed for neighboring genes Tenascin C, *Tnc*, and Tumor Necrosis Factor (Ligand) Superfamily, Member 15, *Tnfsf15*, during the aWOS (S1 Fig). However, differential expression was observed in 4 week-old, immature MECs for both genes, highlighting the complexity and age-specific nature of *Mcs5c* locus activity.

Fig 3: *Pappa* expression is altered in *Mcs5c* susceptible rats in an age-dependent manner. *Pappa* expression was examined in MECs of *Mcs5c* susceptible and *Mcs5c* resistant rats at various ages. Relative gene expression was determined via quantitative real-time PCR and standardized to *Tbp* expression. P-values were obtained using the non-parametric Mann-Whitney U test (*, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$).

Mcs5c and *Pappa* physically interact in an age-dependent and tissue-specific manner

Genotype dependent differential expression seen in MECs led to the hypothesis that *Mcs5c* contained a long-distance acting regulatory element influencing *Pappa* expression. Such a relationship could be mediated by a physical association between the two regions, resulting in the looping out of

intervening DNA sequence. Chromosome conformation capture (3C) was used to identify such an interaction. To create 3C templates, MECs were isolated from the mammary glands of *Mcs5c* resistant and susceptible animals at 4, 6, 7, and 12 weeks of age. The first fixed bait region chosen for investigation, P3-1, begins 2kb upstream of *Pappa*'s first exon, and encompasses an area of high conservation that likely includes the proximal promoter and also contains a conserved CpG island (Fig 4A). This fixed region was negative for any interaction with *Mcs5c* at 4, 7, and 12 weeks of age (S2 Fig). The next bait region investigated, P4-1, is located within the first intron of the *Pappa* gene, 2.5kb downstream of exon 1 (Fig 4A). As seen in Fig 4B, an 8.5kb region within *Mcs5c* displayed a high relative interaction frequency (IF) with P4-1 in 6 and 7 week templates, indicative of a physical interaction between the two regions occurring over a distance of 590kb. 4 and 12 week templates had a much lower IF at this -590 region, leading to the formation of two distinct, age-dependent interaction groups displaying either a strong (6 and 7 week) or weak (4 and 12 week) IF. The difference in IF for these two groups was statistically significant (Mann-Whitney U test, p-value = 1.02e-10, n = 27 and 38 biological replicates, respectively). For all ages, there was no difference in IF between genotypes, indicating that the interaction is age-dependent but not genotype-dependent. We will therefore refer to the -590 looping region of *Mcs5c* as the temporal control element (TCE; chr5:84,428,694 – 84,437,192; RGSC 5.0/rn5). To determine if this interaction is also tissue-specific, 3C profiles were analyzed from 4 and 7 week colon epithelial cells and 7 week liver hepatocytes from *Mcs5c* resistant rats. The TCE did not interact with P4-1 in these tissues (Fig 4C), implying that the interaction between *Pappa* and the *Mcs5c* TCE is tissue-specific in addition to age-dependent. Sequencing of the resistant WKy and susceptible WF TCE alleles revealed 10 variants between the two (S7 Table). Although our 3C results show that age-specific looping occurs independent of genotype, we speculate that one or more variants may be involved in genotype-dependent expression differences observed during this time period.

Fig 4: *Mcs5c* displays an age-dependent and tissue-specific interaction with the *Pappa* gene in MECs. (A) The *Mcs5c* locus is shown in relation to the full length *Pappa* gene on rat chromosome *5*. Bait regions P3-1 (blue line) and P4-1 (red line) used for 3C analysis are highlighted in the zoomed in image (UCSC Genome Browser, March 2012, rn5). Region P3-1 encompasses *Pappa* exon 1, while P4-1 begins 2.5kb downstream. Tick marks indicate *BglII* cut sites, and the green box highlights a conserved CpG island. (B) MEC 3C profiles showing the relative interaction frequency (IF, y-axis) between the bait region P4-1 and regions within the entire *Mcs5c* locus. The x-axis indicates the distance between the tested region in *Mcs5c* and P4-1 (UCSC Genome Browser, March 2012, rn5). The shaded box highlights the temporal control element (TCE) within *Mcs5c* that displayed an age-specific interaction with P4-1, with 6 and 7 week samples showing a strong IF (dotted lines) and 4 and 12 week samples a weak interaction (solid lines). (C) Colon epithelial cell and liver hepatocyte 3C profiles showed a negative/weak IF (y-axis) between the TCE (shaded box) and bait region P4-1. The x-axis is the same as in (B), however, not all sites within *Mcs5c* were tested, resulting in a truncated axis. Only *Mcs5c* resistant rats were used for this analysis. For both (B) and (C), each point represents multiple biological and technical replicates, and standard error bars are included. P-values were obtained using the non-parametric Mann-Whitney U test (*, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$).

Removal of the *Mcs5c* TCE *in vitro* decreases *Pappa* expression

In an effort to causally tie the *Mcs5c* TCE to *Pappa* expression, the entire 8.5kb region was targeted for deletion in the rat mammary carcinoma cell line, LA7. Two CRISPR guides were used to target the region, and clones were screened via PCR across the cut site, with validation by sequencing (S1 Table). We were unable to identify a clone with all copies of the TCE removed, despite much effort. This was likely due to the aneuploid nature of LA7 cells, and mutations incurred at CRISPR guide target regions (S2 & S3 Tables). Copy number analysis of 9 positive CRISPR edited clones showed that we were

able to delete a majority of TCE copies, reducing the copy number by 3.5-fold across all clones (Fig 5A). 3C analysis of positive clones confirmed that removal of multiple TCE copies resulted in decreased *Mcs5c-Pappa* looping (Fig 5B). Expression analysis revealed a significant reduction in *Pappa* expression in CRISPR clones compared to wild-type LA7 cells, with *Pappa* decreased 4-fold across all clones (Fig 5C). A Pearson correlation coefficient was computed to determine the relationship between *Pappa* expression and TCE copy number, and a positive correlation between the two was observed (R = 0.6245, n = 13, p-value = 0.0225; Fig 5D). These data highlight the importance of the *Mcs5c-Pappa* chromatin loop on *Pappa* gene expression, and support our initial hypothesis that *Mcs5c* contains a long-distance regulatory element.

Fig 5: Removal of *Mcs5c* TCE copies resulted in decreased *Pappa* expression *in vitro*. (A) CRISPR guides targeting the TCE were transfected into LA7 cells, and clones were screened for removal of the target region. Nine positive clones were assessed for remaining TCE copy number via qPCR, standardized to a non-targeted region within the *Pappa* gene. Copy number in wild-type LA7 cells (n = 4 independent cultures) was also assessed, and results were normalized to diploid MECs. (B) 3C analysis on select positive clones (n = 3) and a WT LA7 sample showed a decrease in interaction frequency (IF, y-axis) between bait region P4-1 and the *Mcs5c* TCE in clones (TCE). The IF for a positive control region, two nearby *BglII* fragments, is also shown. (C) *Pappa* expression in positive clones and WT LA7 cells (n = 6) was analyzed via qPCR and standardized to *Tbp* expression. P-values for (A) and (C) were obtained using the non-parametric Mann-Whitney U test (*, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$). (D) A scatterplot of *Pappa* expression and *Mcs5c* copy number showed a statistically significant positive correlation between the two (Pearson correlation coefficient, R, = 0.6245, n = 13, p-value = 0.0225). A linear trend line is shown with the dotted line (slope = 5.327).

Pappa CpG island shore is differentially methylated *in vivo* in a genotype-dependent manner

The *Pappa* looping fragment, P4-1, resides in a CpG island (CGI) shore region (Fig 6A). CGI shores are regions located around 2kb of CGIs, and are routinely the site of tissue specific differential methylation associated with gene expression changes [24]. As this shore region is the target site of *Mcs5c* TCE looping, we hypothesized that *Mcs5c* may mediate observed differential *Pappa* expression through an epigenetic mechanism specific to this region. Methylation levels for 12 CG dinucleotides within and proximal to P4-1 were examined in MECs of *Mcs5c* resistant and susceptible rats at 4, 6, 7, 9 and 12 weeks of age using custom designed pyrosequencing assays (Fig 6A). Selection of these timepoints allowed for the examination of methylation patterns before, during, and after the aWOS. In general, methylation levels were dynamic across this region, with sites 2 – 4 consistently displaying the lowest methylation levels (average = 13% methylated) and sites 9 – 12 displaying the highest levels (average = 68% methylated). Individual methylation levels for all ages and CpG sites are shown in S4 Table. Additionally, there appeared to be few age-specific differences in methylation levels for animals within the aWOS, therefore data for animals aged 6, 7, and 9 weeks were combined within genotypes. Of the 12 sites examined, 6 showed statistically significant genotype-dependent differences in methylation levels after adjusting for multiple comparisons (Mann-Whitney U-test with Bonferroni correction). The percent change in methylation levels along with p-values are shown in Table 1. All statistically significant, genotype-dependent methylation differences occurred during the aWOS and were directionally identical, with methylation levels decreased in *Mcs5c* susceptible MECs. The percent decrease in methylation levels ranged from 5.0% – 22.7%. Additionally, a number of other sites displayed a similar trend, although these differences were not significant after Bonferroni correction. At ages outside of the aWOS, there were no statistically significant genotype-dependent differences in methylation, although sites 1 and 2 displayed a non-significant trend of increased methylation in *Mcs5c* susceptible MECs at the 4 week time point. We also investigated the methylation state of the *Pappa* CGI

using 2 pre-made pyrosequencing assays (Fig 6A). Methylation levels for both assays were assessed in 4 week old animals, while one assay was examined at the remaining timepoints. For all CGI assays and timepoints, there were no genotype-dependent differences in methylation levels and, in general, the *Pappa* CGI is hypomethylated at all ages, with site specific methylation levels ranging from 0.16% - 8.41% (S5 Table). The observation of decreased shore methylation and increased *Pappa* expression in *Mcs5c* susceptible MECs strongly supports the canonical role of DNA methylation in gene regulation, that is, that the two are negatively correlated. Indeed, for 6 week MECs, for which we had both DNA and RNA samples, *Pappa* expression was negatively correlated with the average methylation percentage of the 6 significant shore sites (Fig 6B; Pearson correlation coefficient, R, = -0.67, n = 18, p-value = 0.0023). By contrast, no correlation was observed between *Pappa* expression and the average methylation percentage of the CGI-2 assay sites (Fig 6C; Pearson correlation coefficient, R, = 0.16, n = 18, p-value = 0.52). The identification of genotype-dependent methylation differences during the aWOS suggests that *Mcs5c* facilitates genotype-dependent *Pappa* expression differences observed during this time period through epigenetic regulation of the *Pappa* CGI shore.

Fig 6: *In vivo* methylation analysis of the *Pappa* CGI and CGI shore. (A) The first and second exons of the *Pappa* gene (purple boxes) are shown in relation to a conserved CGI (green box) and the P4-1 looping fragment (gray box). Although both exons are shown, exon 2 begins 27.5kb downstream of P4-1. The location of the 12 shore CG dinucleotides investigated in this report are indicated and numbered, as are the regions covered by the two pre-made CGI pyrosequencing assays. The CGI assays each examined 5 CG dinucleotides within the island. (B) A scatterplot showing a statistically significant negative correlation between 6 week MEC *Pappa* expression (x-axis) and shore methylation (y-axis; Pearson correlation coefficient, R, = -0.67, n = 18, p-value = 0.0023) is shown. Shore methylation values were obtained by averaging the absolute methylation percentages of the 6 significant shore sites (Sites 1, 3, 6

– 9) for each individual sample. A linear trend line is shown with the dotted line (slope = -7.88). (C) No correlation was observed between 6 week MEC *Pappa* expression (x-axis) and CGI methylation (y-axis; Pearson correlation coefficient, R, = 0.16, n = 18, p-value = 0.52). CGI methylation values were obtained by averaging the absolute methylation percentages of the 5 sites examined by the CGI-2 assay for each individual sample. A linear trend line is shown with the dotted line (slope = 0.544).

Discussion

Previous work on *Mcs5c* had fine-mapped the locus to a 170kb non-coding region on rat chromosome *5*. This locus resulted in an approximately 50% decrease in both chemical carcinogen and oncogene-induced mammary carcinoma development when homozygous for the resistant WKy allele [11]. In this study, since we showed that the *Mcs5c* locus affects carcinoma multiplicity in a mammary gland autonomous manner through the use of transplant experiments (Fig 2), functional studies were therefore focused on the mammary gland. Expression analysis of neighboring genes revealed genotype-dependent and age-specific differential expression of *Pappa* in MECs (Fig 3). Specifically, *Mcs5c* susceptible MECs from 6 to 9 week old rats, but not older, showed increased expression of *Pappa* compared to *Mcs5c* resistant rats. Importantly, the 6 to 9 week age range encompasses a time period of rapid mammary gland development and maturity, falling within the aWOS [5]. We hypothesized that *Pappa* expression changes were mediated by a control element within *Mcs5c*. Although *Mcs5c* lies over 500kb upstream of the *Pappa* gene, there is precedent for enhancers/suppressors working over such distances [25,26], including the previously characterized *Mcs1a* locus [27]. Through 3C experiments we identified a region within *Mcs5c* which we have termed the temporal control element (TCE) that interacts with an intronic region of *Pappa*, P4-1, in an age-dependent and mammary-specific manner over a distance of 590kb (Fig 4). This interaction occurs specifically in MECs at ages within the aWOS, and takes place regardless of *Mcs5c* genotype, indicating that looping alone cannot explain genotype-

dependent expression differences. The importance of this regional, long-range looping interaction to *Pappa* expression was demonstrated *in vitro* in CRISPR edited rat mammary carcinoma cells (LA7 cell line), where removal of TCE copies and consequent reduction in *Mcs5c-Pappa* looping correlated with decreased *Pappa* expression (Fig 5).

As looping occurs in a genotype-independent manner, additional mechanisms must be responsible for the differential expression observed between *Mcs5c* resistant and susceptible rats. With the intronic *Pappa* looping region, P4-1, falling in a CGI shore, DNA methylation of this region became an important mechanistic candidate to explain genotype-dependent expression differences. The importance of differentially methylated CGI shores to gene expression was first highlighted by Irizarry and colleagues in 2009 [24]. Since then, many studies have shown an association between differentially methylated shore regions and gene expression changes [28-35]. Our study identified 6 CG dinucleotides within and proximal to the P4-1 looping region that were differentially methylated between *Mcs5c* resistant and susceptible MECs (Table 1). Significant methylation differences were only observed during the aWOS, with lower methylation levels found in susceptible MECs. A negative correlation was observed between shore methylation levels and *Pappa* expression within the aWOS, strongly suggesting that DNA methylation may be a mechanism involved in *Mcs5c* regulation of *Pappa* expression (Fig 6B).

Overall, our results point to a complex set of mechanisms underlying *Mcs5c* regulation of *Pappa*. We have shown that during the aWOS time period, rats carrying the susceptible *Mcs5c* allele have increased expression of *Pappa* in MECs compared to rats carrying the resistant allele, and we have identified two mechanisms associated with this activity, chromatin looping and DNA methylation. 3C experiments have revealed the aWOS-specific nature of *Mcs5c-Pappa* looping, and *in vitro* evidence support its role in regulating *Pappa* expression. Similarly, methylation analysis has identified a strong

negative correlation between shore methylation and *Pappa* expression, supporting a role for epigenetic regulation of the *Pappa* gene. An unresolved issue is precisely how *Mcs5c* is mediating these activities. ChIP experiments performed by Schmidt *et al* [36] on rat liver tissue have identified 2 CTCF binding sites within the TCE of *Mcs5c* (identified via CTCFBSDB v2.0) [37,38], suggesting the possibility of a CTCF-mediated looping mechanism, consistent with its role in chromatin organization [39]. ENCODE data support the conservation of these binding sites in both humans and mice [40,41]; however, their relevance to *Mcs5c-Pappa* looping is, at this point, purely speculative. Sequencing of the resistant WKy and susceptible WF *Mcs5c* TCE alleles (chr5:84,428,694 – 84,437,192; RGSC 5.0/rn5) revealed 10 variants (S7 Table). Although looping itself occurs in a genotype-independent manner, we hypothesize that one or more variants are responsible for the differential binding/recruitment of transcription factors, cofactors, and/or methyltransferases that ultimately affect *Pappa* expression. This could be achieved through either their association with transcription machinery at the promotor or by their involvement with differential methylation. Additionally, methylation itself may also be responsible for differential cofactor recruitment at the *Pappa* shore region. Such a mechanism was demonstrated by Bockmühl *et al* [29], where shore methylation disrupted binding of the repressor protein YY1, leading to differential expression of the nearby gene.

While the exact variants and proteins responsible for mediating this interaction remain unknown, our data support the existence of a long-distance acting regulatory element within the *Mcs5c* locus affecting *Pappa* expression in MECs. Mechanistically, *Mcs5c* achieves this through a combination of chromatin looping and genotype-dependent differential DNA methylation that occurs specifically during the aWOS. During this window, *Mcs5c* resistant rats have decreased expression of *Pappa* in MECs and a lower susceptibility to carcinogen-induced mammary carcinogenesis, supporting a protective benefit of reduced *Pappa* levels during this developmental time period. Decreased levels of Pappa in the

developing mammary gland would result in reduced Igf-I bioavailability through a reduction in Igfbp cleavage [12]. During mammary gland development, the Igf-I signaling pathway acts to promote proliferation and inhibit apoptosis, and is essential for proper growth [42]. It is therefore likely that a reduction in free Igf-I would reduce the proliferative index of MECs. As the effects of many environmental mutagens, such as radiation and chemical carcinogens, are dependent on interactions with the DNA of proliferating cells [43], this would result in fewer targets for mutagenesis. This represents one possible method by which reduced *Pappa* expression during the aWOS may result in a mammary carcinoma-resistant phenotype; however, these and alternative hypotheses remain to be tested.

Interestingly, a reversal of genotype-dependent expression differences was observed in 4 week old MECs compared to differences during the aWOS (Fig 2). Specifically, *Pappa* expression was decreased in 4 week old *Mcs5c* susceptible rats, while it was increased in aWOS-aged susceptible rats. As the 4 week time point falls within the iWOS, this suggests that *Mcs5c* activity is window-dependent. Our focus remained on characterizing *Mcs5c* in the context of the aWOS; however, we included the 4 week time point in all mechanistic studies for comparison. MECs from 4 week old rats lacked strong looping between the TCE and the intronic P4-1 region of *Pappa* (Fig 4B), and had no statistically significant genotype-dependent CGI shore methylation differences (Table 1), both of which were observed in aWOS-aged rats. This suggests that these interactions are specific to aWOS-aged regulation of *Pappa* by *Mcs5c*. However, our measure of looping from the *Mcs5c* locus to the *Pappa* gene was not exhaustive, and we speculate that a different looping interaction between *Mcs5c* and *Pappa* occurs during the iWOS to regulate *Pappa* expression at the 4 week time point. Additionally, although statistically significant shore methylation differences were not observed, a trend towards increased methylation in *Mcs5c* susceptible MECs is mechanistically consistent with the reduction of *Pappa*

expression observed during this time period. It is possible that these sites are indicative of significant methylation differences occurring at sites not examined in this study, and shore methylation may still, therefore, be relevant to *Mcs5c* activity during the iWOS. Furthermore, differential expression of *Tnc* and *Tnfsf15* exclusively at the 4 week time point (S1 Fig) indicates that *Mcs5c* regulates multiple genes during the iWOS, and may therefore exhibit a more complex chromatin interaction than the looping model identified in this study. These age-specific differences in function could be explained by interactions with proteins specific to these developmental time points. For example, Tian *et al* [44] showed that ERα expression, which is dependent on developmentally regulated estrogen activity, affected the downstream Igf-I signaling pathways of transgenic mice. Identifying proteomic differences between the immature and adolescent mammary gland will be crucial in understanding the players driving the window-specific mechanistic differences in *Mcs5c* activity.

Our study highlights the importance of characterizing genetic risk factors in the context of developmental windows of susceptibility (G x WOS). Understanding these types of G x WOS interactions will play a crucial role in assessing breast cancer risk and may help identify targets for prevention therapeutics in young women at high risk for developing breast cancer due to genetic mutations or family history. We believe that the *Mcs5c* locus will serve as a model to further study the mechanisms behind G x WOS interactions and to understand how environmental risk factors influence these interactions. There is growing concern over the impact adolescent exposure to a broad range of environmental factors, including early-life stress (ELS), diet, phytoestrogens and other chemicals, may have on long-term breast cancer risk [45]. Indeed, CGI shore methylation has already been found to be affected by environmental factors such as ELS and diet [29,30,46,47]. The *Mcs5c* locus provides a robust model to functionally characterize how environmental factors may affect breast cancer risk by

influencing G x WOS interactions and may provide insight for the characterization of other such cancer susceptibility loci.

Materials and Methods

Animals

Congenic rat lines were maintained in an AAALAC-accredited facility as previously described [11]. All protocols were approved by the University of Wisconsin – Madison School of Medicine and Public Health Animal Care and Use Committee. Congenic rat lines are defined as having the resistant Wistar-Kyoto (WKy) *Mcs5c* allele introgressed on a susceptible Wistar-Furth (WF) background. The resistant congenic line used in this study, 5C-27, is WKy-homozygous for a genomic region that includes the entirety of the *Mcs5c* locus (275kb) [11]. Susceptible control animals are WF-homozygous at the *Mcs5c* locus.

Mammary Gland Transplantation

Mcs5c WKy-homozygous congenic rats from line 5C-27 were used as resistant donors and recipients (*Mcs5c* resistant), and *Mcs5c* WF-homozygous rats were used as susceptible controls (*Mcs5c* susceptible). Abdominal and inguinal mammary glands were collected from female donor rats aged 30-35 days old, scissor minced, and split into four equal volumes. One volume was then grafted onto the interscapular white fat pad of four different 30-35 day old female recipient rats. Three weeks after transplantation, recipients were administered the chemical carcinogen 7,12-dimethylbenz(*a*)anthracene (DMBA), as a single oral dose dissolved in sesame oil at 65 mg/kg of body weight to induce mammary carcinoma formation. At 15 weeks post-DMBA, animals were removed from the study and carcinomas present at the transplant site greater than 3x3mm were counted. Generally, recipient rats developed ≤1 carcinoma at the transplant site, so incidence values were used as opposed to multiplicity. Fat pads

were whole mounted and stained with aluminum carmine to verify transplant mammary gland growth. Four transplant groups were studied, with resistant 5C-27 or susceptible donor glands transplanted into both resistant and susceptible recipients (R:R, R:S, S:R, S:S). The tissue rejection rate for the transplant groups consisting of donors and recipients with the same genotype (R:R, S:S) was compared to the rejection rate for groups with differing donor and recipient genotypes (R:S, S:R) via a Chi-squared test. The effect of donor and recipient genotype on carcinoma incidence, converted to a binary response value, was analyzed using logistic regression with two independent variables (donor genotype and recipient genotype) and no interaction term.

Tissue collection and DNA/RNA processing

For all experiments, mammary epithelial cell (MEC) isolation began with fresh mammary glands (abdominal and inguinal, with lymph nodes removed) that were finely minced and digested for 2 hours at 37°C in 10 mL of GIBCO Dulbecco's modified Eagle's medium/F12 (DMEM/F12; Life Technologies) containing 0.01 g/mL of type III collagenase (Worthington). Cell pellets were collected by centrifugation and resuspended in 5mL DMEM/F12. The suspension was loaded onto a 40µm nylon filter to eliminate stromal cells and collect mammary ductal fragments, consisting of an enriched MEC population. DNA was isolated from cells via the DNeasy Blood and Tissue kit (Qiagen). To isolate RNA, cells were homogenized in TRI Reagent (Ambion), followed by RNA extraction using the MagMAX-96 for Microarrays Total RNA kit (Ambion). LA7 cells used for downstream analysis were collected via treatment with 0.25% trypsin/EDTA (Life Technologies). RNA was extracted using the RNeasy Mini Kit (Qiagen) and DNA was extracted using the DNeasy Blood and Tissue kit (Qiagen).

Quantitative real-time PCR

MECs were collected at 4-12 weeks of age from female *Mcs5c* resistant and susceptible control rats. RNA was isolated as described above. For *in vivo* and *in vitro* expression analysis, cDNA was prepared from 1-2µg total RNA using Superscript II reverse transcriptase (Invitrogen). Gene expression was quantified using pre-designed or custom made TaqMan qPCR assays (*Pappa*, Rn01458295_m1, FAM; *Tbp*, Rn01455646_m1, VIC; *ActB*, cat #4352340E; *Tnc*, probe-FAM 5' CGAGAGCTGTGATTAGA 3', primers 5' GGCTGTCAGAAGGCCAGATG 3' and 5' TGCCATGAAGGGATTTGAAGA 3'; *Tnfsf15*, Rn00595596_m1, FAM) and run on an ABI Prism 7900HT (Applied Biosystems). cDNA was diluted 1:4 or 1:8 and run using reaction conditions described previously [11]. Transcript quantities were calculated as described in Smits *et al* [27], using a standard curve method to calculate C_t values and extrapolate quantity values. Sample measurements are an average of 3-4 replicates and data were analyzed using SDS software version 2.2.2 (Applied Biosystems).

Chromosome conformation capture (3C) assay

Sample templates were prepared from MECs, colon epithelial cells, liver hepatocytes, and LA7 cells. MECs were isolated from 4, 6, 7, and 12 week old *Mcs5c* resistant and susceptible rats and the resulting cell suspension was diluted in PBS and fixed via the addition of 1.5% formaldehyde. Colon epithelial cells were isolated from 4 and 7 week old resistant rats, processed as described in Whitehead *et al* [48], and fixed in formaldehyde. To isolate hepatocytes, the livers of 7 week old resistant rats were digested via cannulation of the portal vein and blanching of the liver using a pre-warmed solution of HBSS (Gibco) + 0.5mM EGTA followed by digestion via pre-warmed DMEM-low glucose (Gibco) + 1000CDU/mL Collagenase type IV (Worthington). Digested livers were collected in DMEM/F12 + 10% FBS on ice and cells dispersed manually. The suspension was filtered through a 100µm nylon filter and the filtrate spun for 2 minutes at 50x*g*. Supernatant was removed and cell pellets were washed until media became clear, followed by fixation in formaldehyde. Bacterial artificial chromosomes (BACs)

encompassing the rat *Mcs5c* and *Pappa* promoter regions (CH230-433D12, CH230-498D4, CH230-256M9, and CH230-244C7) were ordered from Children's Hospital Oakland Research Institute (CHORI) and used as positive control templates. Subsequent template preparation for all cell types and for BAC controls continued as described in detail in Smits *et al.* [49]. The restriction enzyme used was *BglII*. Chromatin interactions were detected via PCR, with bait primers chosen near the predicted promoter regions of *Pappa* tested against *Mcs5c* primers spanning the entire locus. Primer sequences are listed in S6 Table. Reaction components were 1X Herculase reaction buffer, 0.2mM dNTPs, 0.4µM primers, 0.3µL Herculase Enhanced polymerase (5U/µL, Agilent) in a total volume of 25µL. The amount of DNA template to add and optimal annealing temperatures were determined empirically. PCR reactions were performed using the following cycling conditions: 95°C for 1 min, 36 cycles of 95°C for 30 s, T_a for 30 s, 72°C for 20 s, followed by a final extension of 72°C for 8 min. Reactions were analyzed by agarose gel electrophoresis and visualized by ethidium bromide staining. Band intensities were quantified using ImageQuant software (GE Healthcare). A relative interaction frequency (IF) was calculated by dividing the band intensity of the sample templates by that of the BAC control.

Sequencing

Sequencing of the 8.5kb *Mcs5c* looping region (TCE; chr5:84,428,694 – 84,437,192; RGSC 5.0/rn5) identified in 3C experiments was performed on MEC DNA from *Mcs5c* resistant and susceptible rats to assess polymorphisms between the WKy and WF alleles. Sequencing primers are listed in S7 Table. Traditional Sanger sequencing was performed at the University of Wisconsin – Madison Biotechnology Center DNA sequencing facility as described in Smits *et al* [27].

Cell culture and *in vitro* CRISPR editing

The rat mammary carcinoma cell line, LA7, was obtained from the American Type Culture Collection and maintained in DMEM/F12 supplemented with 100 IU/mL penicillin, 100 µg/mL streptomycin (Life Technologies), 5% FBS (HyClone), and 0.005mg/mL insulin. Gene expression analysis proceeded as described above, and copy number analysis was performed via SYBR Green qPCR (Life Technologies). For quantification of *Mcs5c* copies, a primer set located in the CRISPR target region was used (5' CAATCACGTTCACTGTGGGT 3' and 5' TCACCTCACACTACCCCATG 3') and as a control region, a primer set located in *Pappa* was used (5' TCCTCCTGACCACTCTGAGA 3' and 5' CCCTACAAACAGCAGAGGGA 3'). The CRISPR-*Cas9* plasmid pSpCas9(BB)-2A-Puro (PX459) was provided by Dr. Feng Zhang (Addgene plasmid #48139) [50]. Guide sequences were designed using the CRISPR Design Tool (http://crispr.mit.edu) and phosphorylated and annealed guide oligos were inserted into the PX459 plasmid via a combination digestion/ligation reaction. 100ng PX459 plasmid was mixed with 2µL of oligos (diluted 1:250), 1µL Fast Digest *BbsI* (Thermo Scientific), 1X Fast Digest Buffer, 1mM DTT, 1mM ATP, and 1500 units T7 ligase (New England BioLabs) and incubated in a thermocycler for 5 minutes at 37°C followed by 5 minutes at 23°C for a total of 6 cycles. The resulting reaction was then treated with Exonuclease V (NEB) according to the manufacturer's protocol. The product was transformed into competent cells, and colonies expanded and verified via sequencing of the guide insertion site. LA7 cells were transfected with two CRISPR guides flanking the 8.5kb target region. Transfection was performed via electroporation using a Nucleofector® II Device and Amaxa® Cell Line Nucleofector® Kit V (Lonza), according to the manufacturer's instructions. Stable clones were isolated following puromycin selection, and clonal colonies were expanded. Removal of the targeted region was determined via PCR screening and sequencing. Primers used to create guides and screen clones are listed in S8 Table.

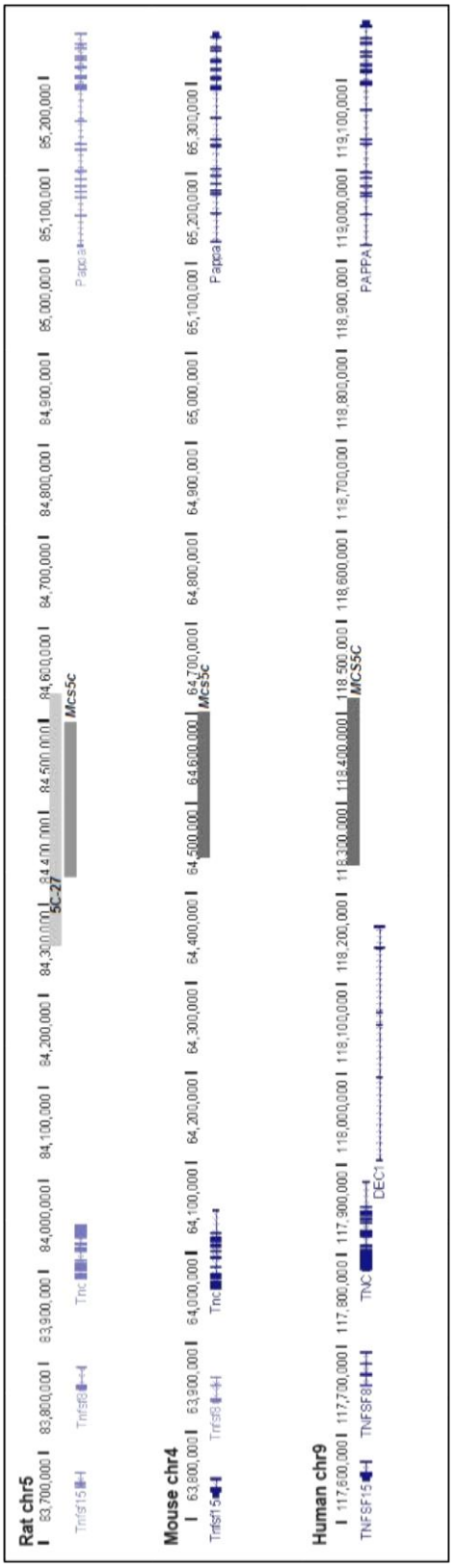
Methylation analysis of the *Pappa* intron

DNA was isolated from *Mcs5c* resistant and susceptible MECs at 4, 6, 7, 9, and 12 weeks of age. Bisulfite conversion was carried out on 500ng of DNA using the EZ DNA Methylation-Lightening kit (Zymo Research), according to the manufacturer's instructions. Four primer sets were designed to amplify the 12 CpG sites of interest within the *Pappa* CpG island (CGI) shore. Their sequences, along with sequencing primers used for pyrosequencing, are listed in S9 Table. Optimal amounts of template DNA, $MgCl_2$, primers, and annealing temperature were experimentally determined (S9 Table). In general, PCR reactions were performed using the following cycling conditions: 95°C for 5 min, 50 cycles of 95°C for 15 s, T_a for 30 s, 72°C for 30 s, followed by a final extension of 72°C for 5 min. 15µL of PCR product was used for pyrosequencing on a PyroMark MD instrument (Qiagen), with 2-3 technical replicates per sample. Data were analyzed using PyroMark CpG software (v 1.0; Qiagen). For analysis of the *Pappa* CGI, 2 pre-made assays were obtained from Qiagen (CGI-1, Rn_D3ZNQ7_01_PM; CGI-2, Rn_D3ZNQ7_02_PM), with PCR conditions following manufacturer's recommendations. Both pre-made assay CGI-1 and CGI-2 amplified 5 CpG sites within the *Pappa* CGI, for a total of 10 sites in the island examined. For statistical analysis of methylation differences, the non-parametric Mann-Whitney U test was used, with a Bonferroni correction applied for multiple comparisons.

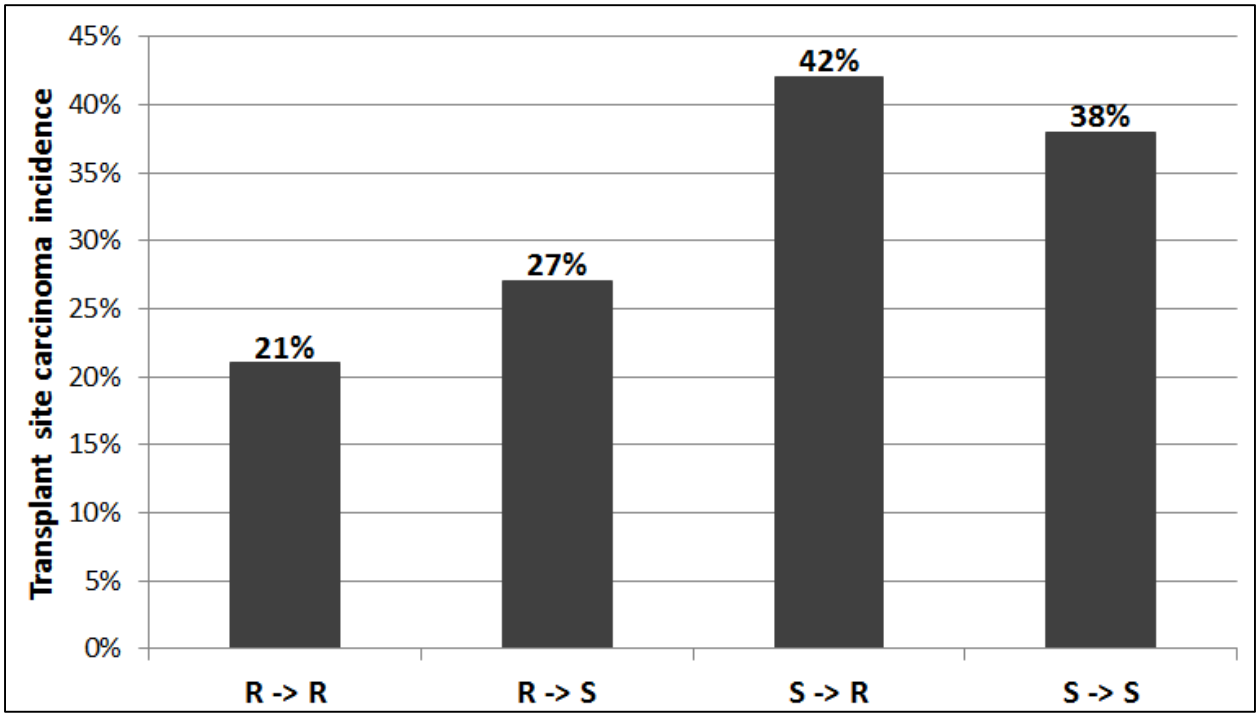
Acknowledgements

The authors would like to thank Dr. James Shull for his critical review of this manuscript, and Amanda Corners and Rachael Baird for their assistance with transplant experiments and tissue collection. Thank you to the UWCCC Pharmacokinetics, Pharmacodynamics, and Pharmacogenetics Laboratory for equipment usage for bisulfite pyrosequencing experiments, and to the UW-Madison Biotechnology Center for their sequencing services.

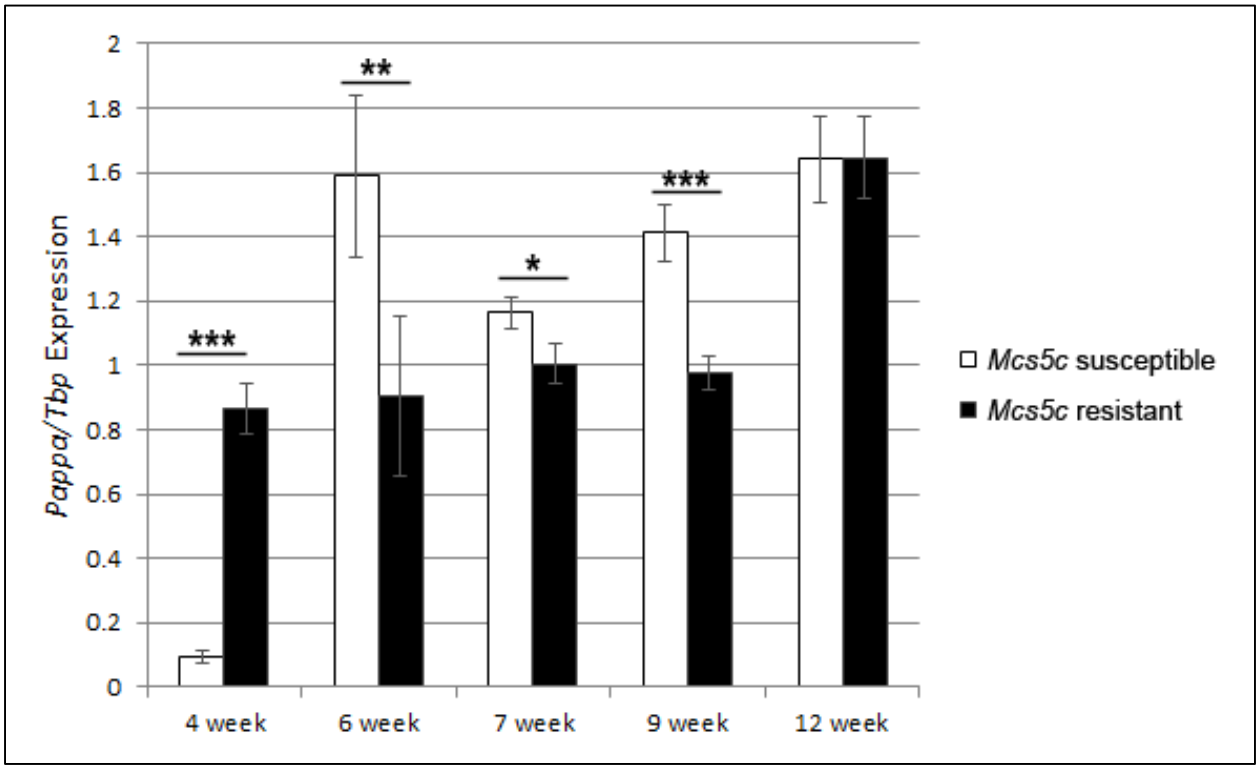
Chapter 2, Fig 1: Comparative genomic map of the *Mcs5c* locus



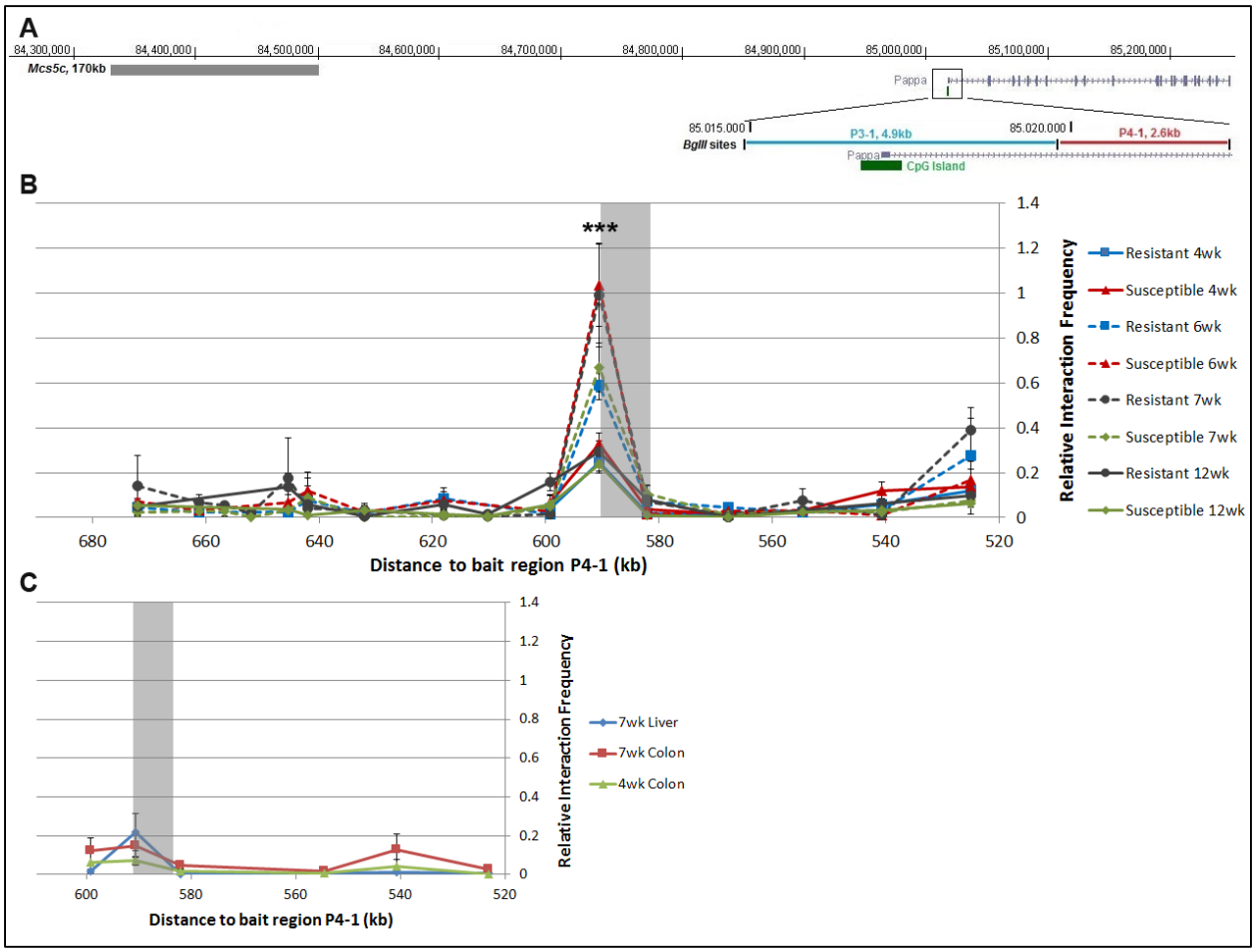
Chapter 2, Fig 2: *Mcs5c* acts in a mammary gland autonomous manner to influence carcinoma multiplicity



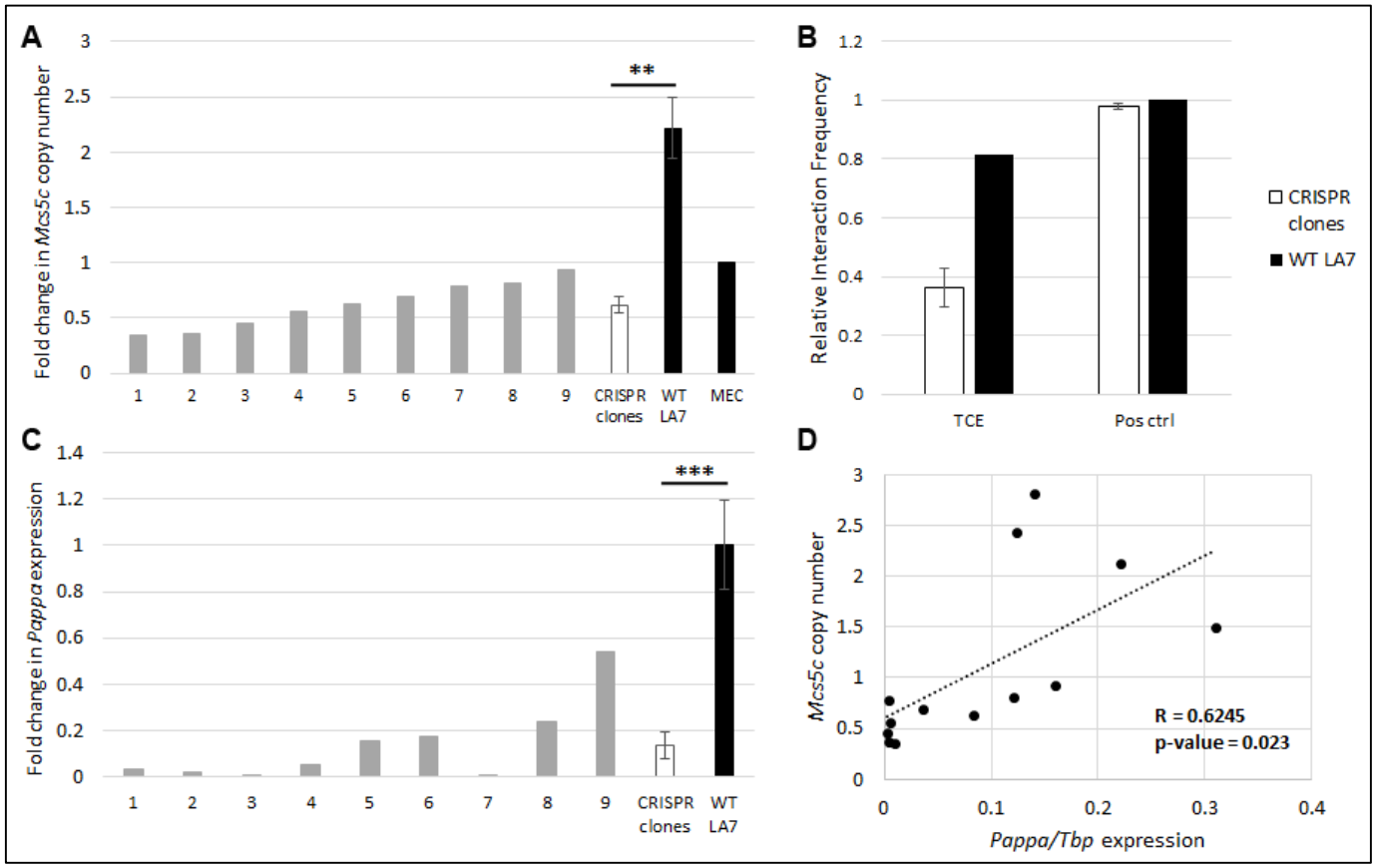
Chapter 2, Fig 3: *Pappa* expression is altered in *Mcs5c* susceptible rats in an age-dependent manner



Chapter 2, Fig 4: *Mcs5c* displays an age-dependent and tissue-specific interaction with the *Pappa* gene in MECs



Chapter 2, Fig 5: Removal of *Mcs5c* TCE copies resulted in decreased *Pappa* expression *in vitro*



Chapter 2, Fig 6: *In vivo* methylation analysis of the *Pappa* CGI and CGI shore

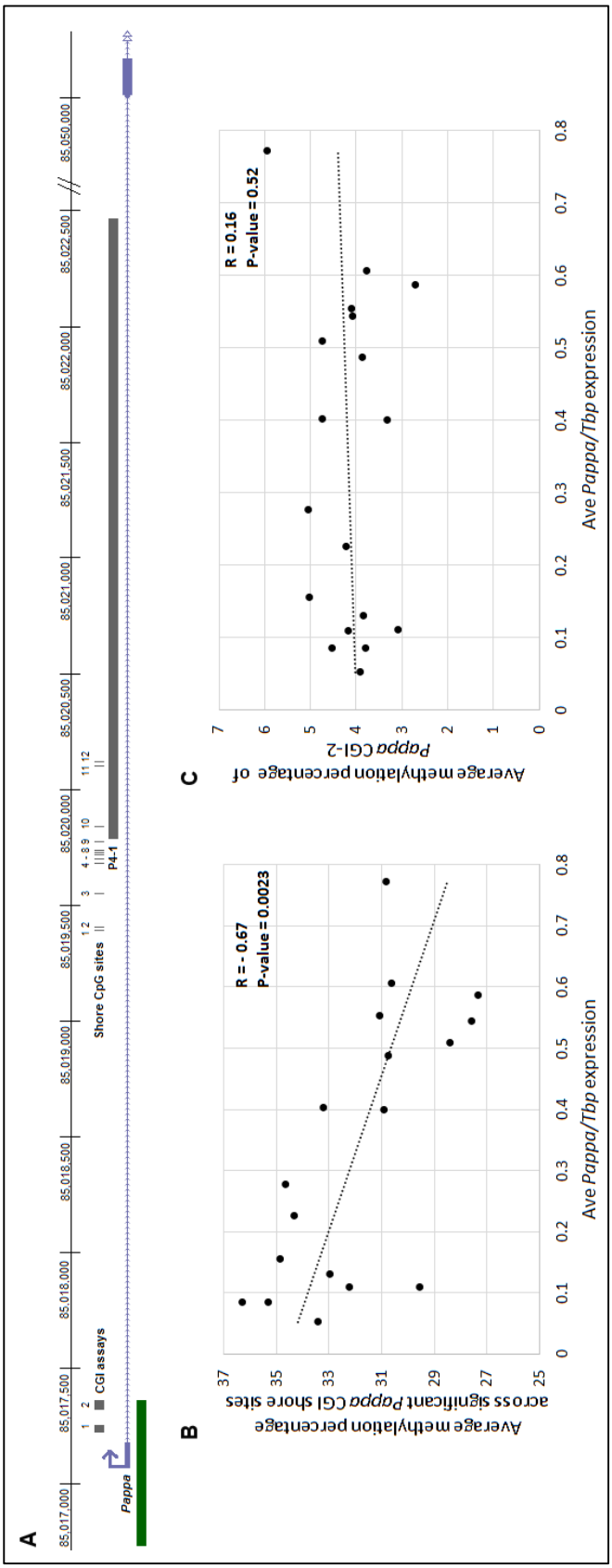


Table 1: Percent change in *Mcs5c* susceptible MEC methylation at sites within the *Pappa* CpG island shore relative to *Mcs5c* resistant MECs

	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Site 9	Site 10	Site 11	Site 12
iWOS - 4 wks	+22.6 (NS)	+26.1 (NS)	(NS)	(NS)	(NS)	(NS)	(NS)	(NS)	(NS)	(NS)	(NS)	(NS)
aWOS - 6/7/9 wks	-20.4 (0.0002)	-23.1 (NS)	-22.7 (0.0004)	-12.6 (NS)	-5.1 (NS)	-12.2 (0.004)	-13.4 (0.003)	-8.9 (0.0114)	-5.0 (0.023)	-5.6 (NS)	-2.2 (NS)	(NS)
Outside WOS - 12 wks	(NS)	(NS)	(NS)	(NS)	(NS)	(NS)	(NS)	(NS)	(NS)	(NS)	(NS)	(NS)

Direction of percent change of methylation levels in *Mcs5c* susceptible rats relative to *Mcs5c* resistant rats is indicated by a + (increase) or - (decrease). Bonferroni corrected p-values are shown in parentheses (Mann-Whitney U Test; NS, not significant). Non-significant percent change trends are also shown.

Supplemental Figure and Table Legends

S1 Fig: *Tnc* and *Tnfsf15* expression in mammary epithelial cells. *Tnc* (A) and *Tnfsf15* (B) expression was examined in MECs of *Mcs5c* susceptible and *Mcs5c* resistant rats at various ages. Gene expression relative to *Mcs5c* susceptible levels was determined via qPCR with *Tnc* standardized to *ActB*, and *Tnfsf15* standardized to *Tbp* expression. There were an average of n = 12 animals per group, and p-values were obtained using the non-parametric Mann-Whitney U test (*, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$).

S2 Fig: Chromosome conformation capture profile for *Mcs5c* and bait region P3-1. MEC 3C profile showing the relative interaction frequency (IF, y-axis) between the bait region P3-1 and regions spanning the entire *Mcs5c* locus. The x-axis indicates the distance between the tested region and P3-1 (UCSC Genome Browser, March 2012, rn5). At the tested timepoints of 4, 7, and 12 weeks of age, there was no significant interaction between *Mcs5c* and P3-1. Every point represents multiple biological and technical replicates, and standard error bars are included.

S1 Table. Sequencing across target cut site of LA7 CRISPR clones

S2 Table. Sequencing of proximal target region of LA7 CRISPR clones

S3 Table. Sequencing of distal target region of LA7 CRISPR clones

S4 Table. MEC methylation levels of *Pappa* CpG island shore

S5 Table. MEC methylation levels of Pappa CpG island

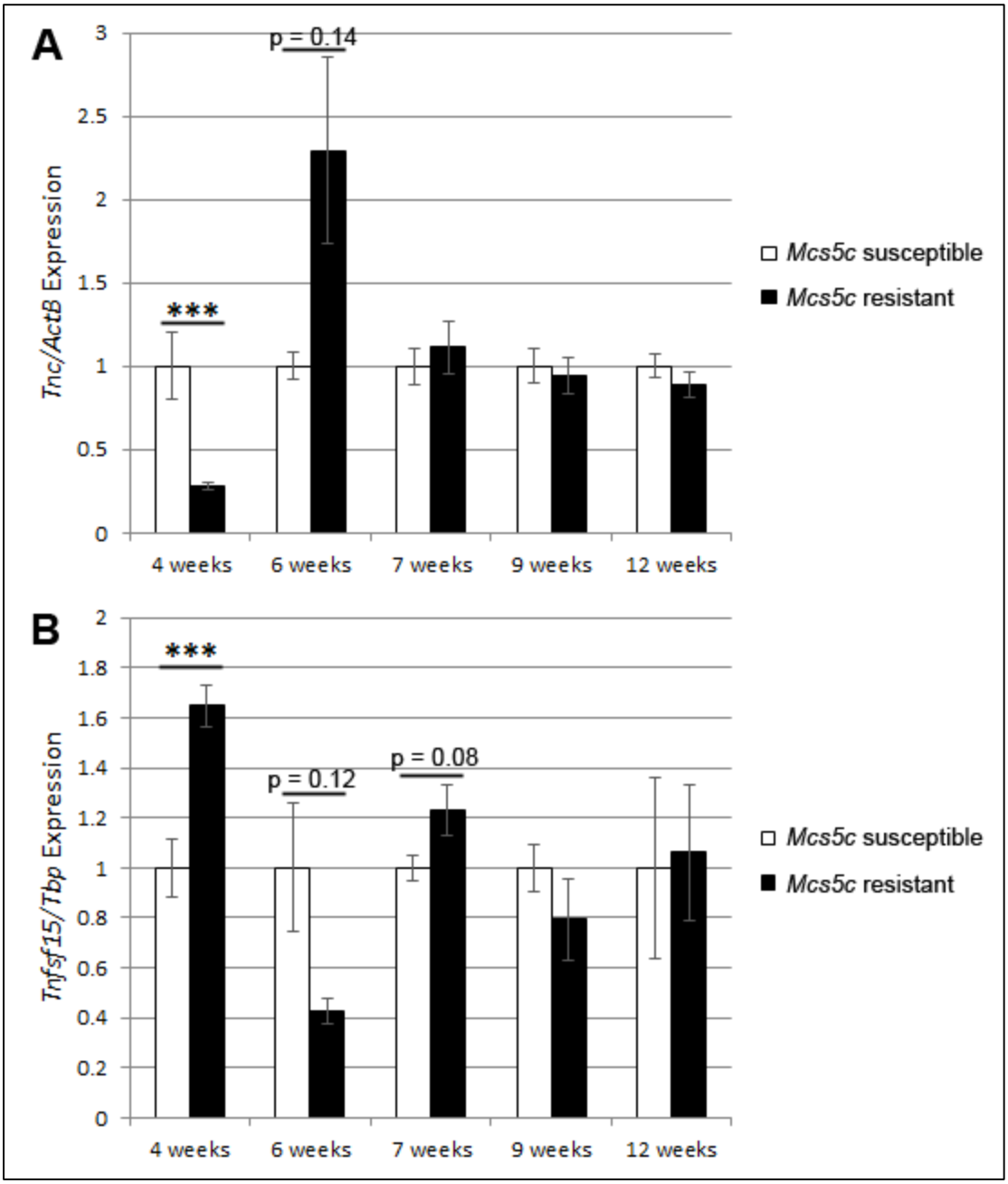
S6 Table. Chromosome conformation capture (3C) primers

S7 Table. Sequencing primers and *Mcs5c* looping variants between the WF and WKy inbred rat strains

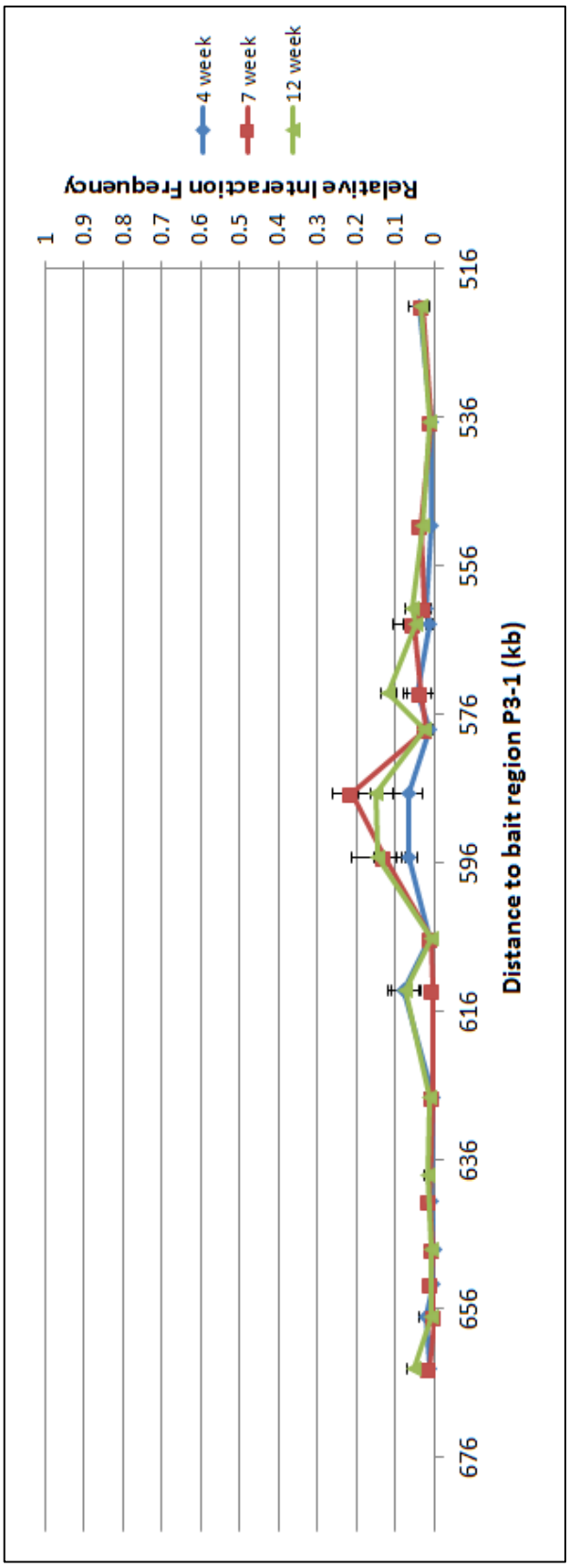
S8 Table. CRISPR gene editing primers

S9 Table. Custom made primers for bisulfite pyrosequencing methylation analysis

Chapter 2, S1 Fig: *Tnc* and *Tnfsf15* expression in mammary epithelial cells



Chapter 2, S2 Fig: Chromosome conformation capture profile for *Mcs5c* and bait region P3-1



S1 Table: Sequencing across target cut site of LA7 CRISPR clones	
Original *Mcs5c* sequence*	TAGAGCCAAA GAAAGTACAC AGAAGGGATG TAGAACTTGG ATCTCTAGGG AGTGTGGAAA AGTCCAGAGA CTGA.....8.1kb.....AACACAACTG TGAAGACTAG AGTTCAGATT CTCAGGATCT ATCTGGGGGT AGCAAAACAA CTCATTTGTA ATTCCAGCCT AGGATGTCAG GGACAGGGGA AATCCTCCTA TAAGCTA**GTT AGATAC****AGTA**
Clone 1	TAGAGCCAAA GAAAGTACAC AGAAGGGATG TAGAACTTGG ATCTCTAGGG A**TAT**TGGTGA GATCTGA
Clone 2	TAGAGCCAAA GAAAGTACAC AGAAGGGATG TAGAACTATA AGCTAGTTAG ATACAGTACC TATATTGGTG AGATCTGA
Clone 3	TAGAGCCAAA GAAAGTACAC AGAAGGGATG TAGAACTATA AGCTAGTTAG ATACAGTACC TATATTGGTG AGATCTGA
Clone 4	TAGAGCCAAA GAAAGTACAC AGAAGGGATG TAGAACTATA AGCTAGTTAG ATACAGTACC TATATTGGTG AGATCTGA
Clone 5	TAGAGCCAAA GAAAGTACAC AGAAGGGATG TAGAACTATA AGCTAGTTAG ATACAGTACC TATATTGGTG AGATCTGA
Clone 6	TAGAGCCAAA GAAAGTACAC AGAAGGGATG TAGAACTATA AGCTAGTTAG ATACAGTACC TATATTGGTG AGATCTGA
Clone 7	TAGAGCCAAA GAAAGTACAC AGAAGGGATG TAGAACTATA AGCTAGTTAG ATACAGTACC TATATTGGTG AGATCTGA
Clone 8	TAGAGCCAAA GAAAGTACAC AGAAGGGATG TAGAACTATA AGCTAGTTAG ATACAGTACC TATATTGGTG AGATCTGA
Clone 9^	GAAAGTACAC AGAAGGGATG TAGAACTTGG ATCTCTAGGG ATATTGGTGA GATCTGA

* Targeting guides are indicated in the sequence using the following key: Guide 2, underlined; Guide 3, red lettering; Guide 4, blue lettering; Guide 5, bold; Guide 6, double underlined

^ Clone 9 was targeted with Guides 2 and 6. Sequencing revealed that double stranded break repair completely eliminated the Guide 6 region.

S2 Table: Sequencing of proximal target region of LA7 CRISPR clones	
Original *Mcs5c* sequence*	TAGAGCCAAA GAAAGTACAC AGAAGGGATG TAGAACTTGG ATCTCTAGGG AGTGTGGAAA AGTCCAGAGA CTGA
Clone 1	TAGAGCCAAA GAAAGTACAC AGAAGGGATG TAGAACTTGG ATCTCTAGGG ATGTGGAAAA GTCCAGAGAC TGA
Clone 2	TAGAGCCAAA GAAAGTACAC AGAAGGGATG TAGAACTTGG ATCTCTAGGG AAGTGTGGAA AAGTCCAGAG ACTGA
Clone 3	TAGAGCCAAA GAAAGTACAC AGAAGGGATG TAGAACTTGG ATCTCTAGGG AAGTGTGGAA AAGTCCAGAG ACTGA
Clone 4	TAGAGCCAAA GAAAGTACAC AGAAGGGATG TAGAACTTGG ATCTCTAGGG AAGTGTGGAA AAGTCCAGAG ACTGA
Clone 5	TAGAGCCAAA GAAAGTACAC AGAAGGGATG TAGAACTTGG ATCTCTAGGG AAGTGTGGAA AAGTCCAGAG ACTGA
Clone 6	TAGAGCCAAA GAAAGTACAC AGAAGGGATG TAGAACTTGG ATCTCTAGGG AAGTGTGGAA AAGTCCAGAG ACTGA
Clone 7	TAGAGCCAAA GAAAGTACAC AGAAGGGATG TAGAACTTGG ATCTCTAGGG AAGTGTGGAA AAGTCCAGAG ACTGA
Clone 8	TAGAGCCAAA GAAAGTACAC AGAAGGGATG TAGAACTTGG ATCTCTAGGG AAGTGTGGAA AAGTCCAGAG ACTGA
Clone 9	TAGAGCCAAA GAAAGTACAC AGAAGGGATG TAGAACTTGG ATCTAGGGAG TGTGGAAAAG TCCAGAGACT GA

* Targeting guides are indicated in the sequence using the following key: Guide 2, underlined; Guide 3, red lettering

S3 Table: Sequencing of distal target region of LA7 CRISPR clones	
Original *Mcs5c* sequence*	AACACAACTG TGAAGACTAG AGTTCAGATT CTCAGGATCT ATCTGGGGGT AGCAAAACAA CTCATTTGTA ATTCCAGCCT AGGATGTCAG GGACAGGGGA AATCCTCCTA TAAGCTA**GTT AGATAC**A**GTA CCTATAT**TGG TGAGATCTGA
Clone 1	AGCAAAACAA CTCATTTGTA ATTCCAGCCT AGGATGTCAG GGACAGGGGA AATCCTCCTA TAAGCTA**GTT AGATACAGTA CCTA**C**TAT**TG GTGAGATCTG A
Clone 2	AGCAAAACAA CTCATTTGTA ATTCCAGCCT AGGATGTCAG GGACAGGGGA AATCCTCCTA ATAAGCTAGT TAGATACAGT ACCTATATTG GTGAGATCTG A
Clone 3	AGCAAAACAA CTCATTTGTA ATTCCAGCCT AGGATGTCAG GGACAGGGGA AATCCTCCTA ATAAGCTAGT TAGATACAGT ACCTATATTG GTGAGATCTG A
Clone 4	AGCAAAACAA CTCATTTGTA ATTCCAGCCT AGGATGTCAG GGACAGGGGA AATCCTCCTA ATAAGCTAGT TAGATACAGT ACCTATATTG GTGAGATCTG A
Clone 5	AGCAAAACAA CTCATTTGTA ATTCCAGCCT AGGATGTCAG GGACAGGGGA AATCCTCCTA ATAAGCTAGT TAGATACAGT ACCTATATTG GTGAGATCTG A
Clone 6	AGCAAAACAA CTCATTTGTA ATTCCAGCCT AGGATGTCAG GGACAGGGGA AATCCTCCTA ATAAGCTAGT TAGATACAGT ACCTATATTG GTGAGATCTG A
Clone 7	AGCAAAACAA CTCATTTGTA ATTCCAGCCT AGGATGTCAG GGACAGGGGA AATCCTCCTA ATAAGCTAGT TAGATACAGT ACCTATATTG GTGAGATCTG A
Clone 8	AGCAAAACAA CTCATTTGTA ATTCCAGCCT AGGATGTCAG GGACAGGGGA AATCCTCCTA ATAAGCTAGT TAGATACAGT ACCTATATTG GTGAGATCTG A
Clone 9^	GTGGCTGGAG GAGGGATGTA TTCATTGGAC TGTGCCTCAA TGGCCAAGGA GAAAGATCAT TAGAAGGTGA TTGAGAATAT AAATTTGCAC C

* Targeting guides are indicated in the sequence using the following key: Guide 4, blue lettering; Guide 5, bold; Guide 6, double underlined

^ Sequencing of Clone 9 revealed that 402bp of the targeted site, including Guide 6 and surrounding regions, was deleted following repair. The junction point of this repair is indicated by the red nuc

S4 Table: MEC methylation levels of *Pappa* CpG island shore													
		CpG Site 1	**CpG Site 2**	**CpG Site 3**	**CpG Site 4**	**CpG Site 5**	**CpG Site 6**	**CpG Site 7**	**CpG Site 8**	**CpG Site 9**	**CpG Site 10**	**CpG Site 11**	**CpG Site 12**
4wk	*Mcs5c* Resistant Ave (n=6)	23.81%	11.31%	17.32%	10.18%	45.45%	27.33%	24.52%	34.97%	59.84%	55.65%	84.45%	55.39%
	std err	0.57	0.61	0.98	0.64	1.28	0.76	0.81	0.52	0.46	1.02	1.08	0.60
	Mcs5c Susceptible Ave (n=6)	29.18%	14.27%	16.02%	11.31%	48.00%	27.83%	24.89%	36.01%	59.46%	58.40%	84.74%	55.91%
	std err	1.87	1.07	0.80	0.70	1.00	0.99	1.07	1.10	0.81	1.24	1.01	0.88
	p-value	0.0411	0.0260	0.6991	0.4848	0.1797	0.4848	0.9372	0.6991	0.9372	0.1320	0.7553	0.6297
6wk	*Mcs5c* Resistant Ave (n=10)	26.95%	13.73%	17.67%	14.16%	50.37%	30.59%	26.86%	39.31%	60.91%	56.12%	88.65%	60.87%
	std err	1.14	1.06	0.94	0.40	1.11	0.64	0.51	0.78	0.57	1.24	0.66	0.83
	Mcs5c Susceptible Ave (n=8)	19.92%	9.36%	12.43%	13.09%	48.01%	26.40%	23.58%	36.06%	59.95%	55.83%	87.51%	62.34%
	std err	0.74	0.40	0.50	0.71	1.17	1.07	0.84	0.82	1.01	1.14	0.94	0.80
	p-value	3.20E-04	0.0021	1.83E-04	0.2842	0.3154	0.0085	0.0021	0.0085	0.6334	0.8884	0.2897	0.2398
7wk	*Mcs5c* Resistant Ave (n=6)	23.22%	8.70%	13.70%	13.75%	51.16%	28.65%	26.39%	37.68%	62.85%	62.72%	89.99%	65.50%
	std err	0.86	0.67	0.69	1.02	1.07	1.02	1.26	1.20	0.48	1.17	0.72	0.75
	Mcs5c Susceptible Ave (n=8)	20.81%	8.87%	12.05%	12.01%	47.40%	26.01%	22.34%	35.04%	58.26%	56.47%	86.47%	61.26%
	std err	1.31	0.56	0.35	0.60	0.78	0.85	0.43	0.58	0.75	0.95	0.81	0.74
	p-value	0.1419	0.8518	0.0639	0.2824	0.0127	0.0813	0.0047	0.1079	6.66E-04	0.0013	0.0063	6.70E-04
9wk	*Mcs5c* Resistant Ave (n=5)	27.56%	11.57%	14.35%	13.50%	53.30%	31.96%	29.57%	37.30%	64.35%	63.36%	87.57%	65.12%
	std err	1.00	0.32	0.82	1.33	1.72	0.43	0.87	1.37	1.17	1.02	0.95	1.19
	Mcs5c Susceptible Ave (n=4)	22.21%	8.81%	11.89%	10.50%	52.43%	28.43%	26.71%	32.54%	59.35%	63.25%	86.57%	63.43%
	std err	0.49	0.69	0.23	0.23	1.01	1.10	1.34	1.51	1.10	1.95	2.24	1.79
	p-value	1.83E-04	0.0062	0.0159	0.1905	1.0000	0.0635	0.1905	0.0635	0.0317	0.9048	0.7302	0.5556
6/7/9wk	*Mcs5c* Resistant Ave (n=21)	26.03%	11.78%	15.74%	13.89%	51.29%	30.36%	27.37%	38.36%	62.28%	60.16%	88.83%	63.27%
Combined	std err	0.77	0.71	0.65	0.44	0.74	0.49	0.53	0.60	0.50	1.09	0.45	0.78
	Mcs5c Susceptible Ave (n=20)	20.73%	9.05%	12.17%	12.14%	48.65%	26.65%	23.71%	34.95%	59.16%	56.79%	86.91%	62.12%
	std err	0.62	0.32	0.24	0.42	0.72	0.60	0.57	0.56	0.55	0.67	0.60	0.57
	p-value	3.0303E-06	0.0027	5.96E-07	0.0079	0.0325	5.5507E-05	4.213E-05	0.0002	0.0002	0.0132	0.0128	0.1748
12wk	*Mcs5c* Resistant Ave (n=5)	24.89%	10.32%	12.02%	13.32%	53.42%	30.12%	29.59%	36.03%	62.73%	61.62%	90.32%	64.85%
	std err	3.13	1.65	0.81	0.68	1.45	0.82	1.19	1.30	1.14	2.87	0.84	1.33
	Mcs5c Susceptible Ave (n=5)	23.59%	11.86%	13.00%	13.09%	57.37%	31.57%	30.64%	37.20%	61.76%	63.60%	88.23%	63.57%
	std err	1.80	1.54	0.60	0.59	1.74	0.80	1.85	1.03	1.67	2.34	1.16	1.25
	p-value	1.0000	0.7857	0.4206	0.6905	0.1508	0.3095	0.6905	0.5476	1.0000	0.5556	0.1903	0.5288

Non-corrected p-values calculated via the non-parametric Mann-Whitney U Test are shown. New familywise significance threshold after Bonferroni correction = 0.00069. P-values significant at this threshold are highlighted in gray.

S5 Table: MEC methylation levels of *Pappa* CpG island

CGI-1: Chr5:85,017,189 - 85,017,285 (Rn5)

		CpG Site 1	CpG Site 2	CpG Site 3	CpG Site 4	CpG Site 5
4wk	*Mcs5c* Resistant Ave (n=6)	1.20%	0.43%	2.26%	0.45%	0.65%
	std err	0.15	0.17	0.14	0.13	0.11
	Mcs5c Susceptible Ave (n=6)	1.18%	0.82%	2.20%	0.16%	0.43%
	std err	0.05	0.12	0.12	0.11	0.12
	p-value	>0.05	>0.05	>0.05	>0.05	>0.05

CGI-2: Chr5:85,017,266 - 85,017,405 (Rn5)

		CpG Site 1	CpG Site 2	CpG Site 3	CpG Site 4	CpG Site 5
4wk	*Mcs5c* Resistant Ave (n=6)	1.80%	2.81%	4.92%	7.20%	4.56%
	std err	0.25	0.50	0.40	0.53	0.61
	Mcs5c Susceptible Ave (n=6)	1.87%	4.26%	5.99%	7.42%	3.05%
	std err	0.28	0.55	0.48	0.39	0.41
	p-value	>0.05	>0.05	>0.05	>0.05	0.0411
6wk	*Mcs5c* Resistant Ave (n=10)	1.43%	3.00%	5.68%	7.42%	3.71%
	std err	0.25	0.32	0.39	0.51	0.30
	Mcs5c Susceptible Ave (n=8)	1.13%	2.78%	5.36%	7.80%	2.99%
	std err	0.30	0.68	0.64	0.77	0.63
	p-value	>0.05	>0.05	>0.05	>0.05	>0.05
7wk	*Mcs5c* Resistant Ave (n=6)	1.63%	2.24%	4.61%	6.52%	3.06%
	std err	0.45	0.24	0.61	0.69	0.12
	Mcs5c Susceptible Ave (n=8)	1.55%	2.52%	5.12%	6.23%	3.86%
	std err	0.28	0.40	0.50	0.34	0.57
	p-value	>0.05	>0.05	>0.05	>0.05	>0.05
9wk	*Mcs5c* Resistant Ave (n=5)	2.58%	3.90%	4.90%	8.41%	4.91%
	std err	0.59	0.23	0.47	0.89	0.53
	Mcs5c Susceptible Ave (n=4)	2.99%	3.05%	4.63%	7.19%	4.07%
	std err	0.35	0.23	0.33	0.43	0.51
	p-value	>0.05	>0.05	>0.05	>0.05	>0.05
6/7/9wk Combined	*Mcs5c* Resistant Ave (n=21)	1.76%	3.00%	5.19%	7.40%	3.81%
	std err	0.23	0.22	0.29	0.39	0.24
	Mcs5c Susceptible Ave (n=20)	1.70%	2.73%	5.11%	7.05%	3.55%
	std err	0.23	0.30	0.32	0.37	0.35
	p-value	>0.05	>0.05	>0.05	>0.05	>0.05
12wk	*Mcs5c* Resistant Ave (n=5)	1.63%	3.34%	5.02%	7.52%	4.91%
	std err	0.13	0.59	0.65	0.57	0.51
	Mcs5c Susceptible Ave (n=5)	2.06%	3.29%	5.27%	7.63%	4.75%
	std err	0.22	0.51	0.29	0.41	0.54
	p-value	>0.05	>0.05	>0.05	>0.05	>0.05

Non-corrected p-values calculated via the non-parametric Mann-Whitney U Test are shown. New familywise significance threshold after Bonferroni correction = 0.00143.

S6 Table: Chromosome conformation capture (3C) primers		
Name	**Position on rat ch5 (rn5)**	**Sequence (5' – 3')**
Mcs5c		
L619-E5-1	84347357	TGTTCACATCCCCTAGACAACTTTA
L619-E5-2	84351297	GAAAAGCTGAACTGTCCCAAAGTAT
L619-E2-2	84358162	CGTAAAAGAAGCACCATAGTTTTCC
L619-E2-2B	84362744	TTGGCAGCTGAAGCAATAGTAAA
L619-E2-3A	84367399	CCAGCATGAACTTTGATGGTAATAG
L619-E2-1	84373935	CTCCCAATTTGTCTGTAGAGATGAC
L619-E2-1C	84377423	ACATGCAGGGAAAAATAGACAGATT
L619-E3-1B	84387454	TGCTCATGTTAGGTTCCAGTTTAAT
L619-E3-1A	84387722	GAATGAAAGGGTAGATTGTTGGAAT
L619-E4-1A	84401386	AGTTTTGGGTTTTGGTGTACATTTT
L619-E4-1B	84402212	ATACCTCCCTCCACCAATTATGAA
L619-E4-2	84409109	AATGAGTCAGAACAACACCAAAAAT
L619-E4-3	84420136	ATGTGGAGCAAAAGGAGCAC
L619-E1-2Aa*	84428804	GATCCAAGTTCTACATCCCTTCTGT
L619-E1-1	84437228	ACCTTCTAATGATCTTTCTCCTTGG
L619-E1-1D	84442267	CTATGGTTTATCTCTGGCCTCTTGT
L619-E1-3B	84451535	TACATAATACAGACAGCACCAGCAC
L619-E1-3A	84453612	AGCTTAGGGACTGAGATTGACAGTT
L619-E1-4	84464788	CAATAAACAGCCAGACAGAAGAGAG
L619-E1-5	84478703	TGGAAAATTCCTGACACTGTATGTT
L619-E1-6	84494352	CCACTTATGGGATACAACTTCCTCT
L619-E1-6B	84496188	CCAGGATTCATATCAGTGTGTTTGT
Pappa		
L619-P3-4	85005792	ACAGATGATGGTGTAAGTCCAGAAA
L619-P3-3	85009054	TGGCCATAAACAATGTTCGTAGTAT
L619-P3-2	85011824	CCATTAAAAATTCATGTCAGACCAA
L619-P3-1	85015002	CCTTCCTTAGAGTTAGGATCTTTGC
L619-P4-1	85019840	TTCAGTGACTGTCTGAACCAGTGT
L619-P4-2	85028953	GGGAATTGGAGTCTCAGAAAGATAC
L619-P4-3	85038363	CAATGAGAACACATGAGAATGTTTG

*Primer L619-E1-2Aa marks the TCE of *Mcs5c*

Positions are in basepairs (UCSC version 5.0/rn5)

S7 Table: Sequencing primers and *Mcs5c* looping variants between the WF and WKy inbred rat strains

Sequencing primers for *Mcs5c* looping region (temporal control element, TCE)

Fragment	Forward Primer (5' – 3')	Reverse Primer (5' – 3')	PrimerF position (rn5)	PrimerR position (rn5)	# Variants	Variant class
Fragment 1	CCTTACCATGGTCACAACATCT	GTTTCCATAAGCTTGTTTCTGTATT	84428565	84429012	0	
Fragment 2	AATGGATGGGAATGGATCAG	CCATCCTCCCATAGTTTGGA	84428903	84430815	1	SNP
Fragment 3	TGGAGGAAGGAACAGGAATGT	TCCCTCGCTCTCAGTTTTGT	84430737	84431212	0	
Fragment 4	CTTAAGGGAGATGGCAGGGT	GAAGAATGGGGAGGGGAACA	84431046	84431447	0	
Fragment 5	CAATCACGTTCACTGTGGGT	TCACCTCACACTACCCCATG	84431490	84432328	4	SNPs
Fragment 6	CCATGGGGTAGTGTGAGGTG	CTGGAAGCAGGCAGTTGAGA	84432327	84434139	2	1 SNP, 1 indel
Fragment 7	CCACAGCTATGCACCAATCC	CCCTTTTACCCTGAGACCCA	84433910	84434359	0	
Fragment 8	CCAGGGTTCAGGAGACAAGT	TTGCCTCCATTCTCTCAGCA	84434214	84434687	0	
Fragment 9	TGTGTGTGGTTCCTTTAATGCA	CATCACCTTCCCACCTCAGA	84434578	84435156	1	SNP
Fragment 10	GTGGGAAGGTGATGCTTTGG	ACCTCCCTTCTGTTTCCATAGT	84435162	84435795	1	SNP
Fragment 11	TGGCAAAGGAATACACTATGGA	AGCCCATTCCTCACTGAAGA	84435781	84436737	1	SNP
Fragment 12	TCTTCAGTGAGGAATGGGCT	TTGGCAGCTCATGATCAACA	84436737	84437428	0	

Variants identified in TCE

Variant	Variant sequence (WF allele / WKy allele)	Fragment	Mutation position (rn5)	Variant class
1	GAAACTATGAATAAACTGGGTAG T/A TCAAAACAATAGAAAGACCAGTTGACTTCTAG	Fragment 2	84430146	SNP
2	TCTTGGCAGGACCAAGAAAAGCCT G/C CATTCTTATTCCTCTGCTCAGCTTG	Fragment 5	84431538	SNP
3	GTGTTACAGAAGCTCACCAGTCTAC C/T TGGCTTTTGCTCCCTGTTCTTTCAAGGAA	Fragment 5	84431907	SNP
4	GACATTGGATTCAGAATAG A/G TCCCACGGCTGCTGGCAGACTATAGGATACTCCAGTTAA	Fragment 5	84431956	SNP
5	CGATGATTTCAACAGTAACATCCATATAG T/A CCTTTTTCCAAAGTCCAGATCCAGTCTCCT	Fragment 5	84432102	SNP
6	TATAACTTAGCTTCATTTCCAGAAATC ---/TTT TTTTTTTTTGTTTTGAGACAGAGGCTCCTGAA	Fragment 6	84432568	indel
7	TTTGCAGAATGGGATATTAAAATATTCAC T/C TGGGGGATGGGTCATGGGGTTCCGTCAATA	Fragment 6	84433667	SNP
8	TTGTGAACATGTGCTGAGAGAATGGAGGCAAAT A/G ACCTTGTGTAATGTAGGATAGTGTGA	Fragment 9	84434709	SNP
9	GTGCCCCTCTTCTATGTCAGGGAGC G/A ATATTTACTACTGCTGTTACTGTT	Fragment 10	84435437	SNP
10	AAATCAATCCACTTACTAATCTAAG A/G GTTACTGACTATGTAGTTGGGACATTATTCAGGAT	Fragment 11	84436672	SNP

Positions are in basepairs (UCSC version 5.0/rn5)

For variant 6 the WKy allele matched the reference genome sequence (Brown Norway inbred strain)

For all other variants, the WF allele matched the reference sequence

S8 Table: CRISPR gene editing primers		
Name	**Position on rat ch5 (rn5)**	**Sequence (5' – 3')**
Guide inserts*		
L633-exp1-G2a	84428812 (+ strand)	CACCGATGTAGAACTTGGATCTCT
L633-exp1-G2b	84428812 (- strand)	AAACAGAGATCCAAGTTCTACATC
L633-exp1-G3a	84428820 (+ strand)	CACCGACTTGGATCTCTAGGGAGTG
L633-exp1-G3b	84428820 (- strand)	AAACCACTCCCTAGAGATCCAAGTC
L633-exp1-G4a	84437156 (- strand)	CACCGTATCTAACTAGCTTATAGG
L633-exp1-G4b	84437156 (+ strand)	AAACCCTATAAGCTAGTTAGATAC
L633-exp1-G5a	84437167 (+ strand)	CACCGTTAGATACAGTACCTATAT
L633-exp1-G5b	84437167 (- strand)	AAACATATAGGTACTGTATCTAAC
L633-exp1-G6a	84437073 (+ strand)	CACCGTCAGATTCTCAGGATCTATC
L633-exp1-G6b	84437073 (- strand)	AAACGATAGATCCTGAGAATCTGAC
Primers to check guide insertion in PX459		
L633-Gchk-2F	NA	TTTCTTGGGTAGTTTGCAGTTTT
L633-Gchk-2R	NA	CGACTCGGTGCCACTTTT
Primers to screen deletion of target region		
L619-5Cseq1F	84428565	CCTTACCATGGTCACAACATCT
L619-5Cseq6R	84437428	TTGGCAGCTCATGATCAACA
L619-5Cseq1R	84429012	GTTTCCATAAGCTTGTTTCTGTATT
L619-5Cseq6F	84436737	TCTTCAGTGAGGAATGGGCT
L619-5Cseq3dF	84431490	CAATCACGTTCACTGTGGGT
L619-5Cseq3dR	84432328	TCACCTCACACTACCCCATG

*For insertion into the PX459 plasmid, *BbsI* compatible overhangs were added onto the guide sequences. Those additional bases are indicated in red.

S9 Table: Custom made primers for bisulfite pyrosequencing methylation analysis

Assay Name	Shore sites examined	Primer name	5' modification	Sequence (5' – 3')	Rxn volumn	[$MgCl_2$]	[F primer]	[Rb primer]	Template dilution	T_{anneal}	Product size
Shore_1	Sites 1 & 2	L634-BSP-2F2	none	AGAAGTATTTTTTAATTTTTTGATTTGTG	50ul	3mM	200nM	100nM	1ul, 1:4 dilution	60°	162bp
		L634-BSP-2R2b	biotin	AAATAATAATACCCCCCCCAAAAAAAACTAAC							
		L634-BSP-seq2c*	none	ATTTGTGGTTAAATTG							
Shore_2	Site 3	L634-BSP-3F2	none	GTGGGTGATAATGAGTATTTTTAG	50ul	3mM	200nM	100nM	1ul, undiluted	61°	186bp
		L634-BSP-3R2b	biotin	AAAAAAATCCCTAAACCTAAAATAACAATTACAAACC							
		L634-BSP-seq3b*	none	GGGGTATTATTATTTT							
Shore_3a	Sites 4 - 9	L634-BSP-4F6	none	GTTTGGTAGTAGTAGGTTTTGTAGTTTTTGGGAG	50ul	3mM	200nM	200nM	1ul, 1:4 dilution	57°	178bp
		L634-BSP-4R3b	biotin	ATAAAAAACATAATTTCTATACTTCAAAAACCTACTC							
		L634-BSP-seq4*	none	TTTGTTTTATATTATTAGATTGT							
Shore_3b	Site 10	L634-BSP-4F8	none	TTTTAGTTTGTTTTATATTATTAGATTG	50ul	3mM	200nM	200nM	2ul, undiluted	57°	265bp
		L634-BSP-4R3b	biotin	ATAAAAAACATAATTTCTATACTTCAAAAACCTACTC							
		L634-BSP-seq4c*	none	TTTTGTTTGGGTTTAAATTTT							
Shore_4	Sites 11 & 12	L634-BSP-5F	none	AGAGAATAGTAGTTTGTGTAGGGAAATTTG	50ul	3mM	200nM	200nM	1ul, 1:4 dilution	59°	211bp
		L634-BSP-5Rb	biotin	AACACACCCTCTAATCCTCTATAAATCTC							
		L634-BSP-seq5*	none	GAAAGTATTGAGGGTGTAGGA							

*Indicates sequencing primers for each assay used in the pyrosequencing reaction

References

1. American Cancer Society. Cancer Facts & Figures 2015. Atlanta: American Cancer Society; 2015. Available: http://www.cancer.org/acs/groups/content/@research/documents/document/acspc-046381.pdf

2. Biro FM and Deardorff J. Identifying opportunities for cancer prevention during preadolescence and adolescence: Puberty as a window of susceptibility. J Adolesc Health. 2013; 52: S15-S20.

3. Land CE, Tokunaga M, Koyama K, Soda M, Preston DL, Nishimori I, et al. Incidence of female breast cancer among atomic bomb survivors, Hiroshima and Nagasaki, 1950-1990. Radiat Res. 2003; 160: 707-717.

4. Hancock SL, Tucker MA, Hoppe RT. Breast cancer after treatment of Hodgkin's disease. J Natl Cancer Inst. 1993; 85: 25-31.

5. Russo J, Wilgus G, Russo IH. Susceptibility of the mammary gland to carcinogenesis: I. Differentiation of the mammary gland as determinant of tumor incidence and type of lesion. Am J Pathol. 1979; 96: 721-736.

6. Ariazi JL, Haag JD, Lindstrom MJ, Gould MN. Mammary glands of sexually immature rats are more susceptible than those of mature rats to the carcinogenic, lethal, and mutagenic effects of N-nitroso-N-methylurea. Mol Carcinogen. 2005; 46: 155-164.

7. Fachal L and Dunning AM. From candidate gene studies to GWAS and post-GWAS analyses in breast cancer. Curr Op Genet Dev, 2015, 30:32-41.

8. Russo J. Significance of rat mammary tumors for human risk assessment. Toxicol Pathol. 2015; 43: 145-170.

9. Samuelson DJ, Haag JD, Lan H, Monson DM, Shultz MA, Kolman BD, et al. Physical evidence of Mcs5, a QTL controlling mammary carcinoma susceptibility, in congenic rats. Carcinogenesis. 2003; 24: 1455-1460.

10. Samuelson DJ, Aperavich BA, Haag JD, Gould MN. Fine mapping reveals multiple loci and a possible epistatic interaction within the mammary carcinoma susceptibility quantitative trait locus, Mcs5. Cancer Res. 2005; 65: 9637-9642.

11. Veillet AL, Haag JD, Remfert JL, Meilahn AL, Samuelson DJ, Gould MN. Mcs5c: A mammary carcinoma susceptibility locus located in a gene desert that associates with Tenascin C expression. Can Prev Res (Phila). 2011; 4: 97-106.

12. Conover CA, Bale LK, Overgaard MT, Johnstone EW, Laursen UH, Fuchtbauer EM, et al. Metalloproteinase pregnancy-associated plasma protein A is a critical growth regulatory factor during fetal development. Development. 2004; 131: 1187-1194.

13. Monget P, Mazerbourg S, Delpuech T, Maurel MC, Manière S, Zapf J, et al. Pregnancy-associated plasma protein-A is involved in insulin-like growth factor binding protein-2 (IGFBP-2) proteolytic degradation in bovine and porcine preovulatory follicles: Identification of cleavage site and characterization of IGFBP-2 degradation. Biol Reprod. 2003; 68: 77-86.

14. Lawrence JB, Oxvig C, Overgaard MT, Sottrup-Jensen L, Gleich GJ, Hays LG, et al. The insulin-like growth factor (IGF)-dependent IGF binding protein-4 protease secreated by human fibroblasts is pregnancy-associated plasma protein-A. Proc Natl Acad Sci USA. 1999; 96: 3149-3153.

15. Laursen LS, Overgaard MT, Søe R, Boldt HB, Sottrup-Jensen L, Giudice LC, et al. Pregnancy-associated plasma protein-A (PAPP-A) cleaves insulin-like growth factor binding protein (IGFBP)-5 independent of IGF: implications for the mechanism of IGFBP-4 proteolysis by PAPP-A. FEBS Lett. 2001; 504: 36-40.

16. Kleinberg DL, Wood TL, Furth PA, Lee AV. Growth hormone and Insulin-like growth factor-I in the transition from normal mammary development to preneoplastic mammary lesions. Endocr Rev. 2009; 30: 51-74.

17. Ruan W and Kleinberg DL. Insulin-like growth factor I is essential for terminal end bud formation and ductal morphogenesis during mammary development. Endocrinology. 1999; 140: 5075-5081.

18. Bonnette SG and Hadsell DL. Targeted disruption of the IGF-I receptor gene decreases cellular proliferation in mammary terminal end buds. Endocrinology. 2001; 142: 4937-4945.

19. Christopoulos PF, Msaouel P, Koutsilieris M. The role of the insulin-like growth factor-1 system in breast cancer. Mol Cancer. 2015; 14:43. doi 10.1186/s12943-015-0291-7

20. de Ostrovich KK, Lambertz I, Colby JK, Tian J, Rundhaug JE, Johnston D, et al. Paracrine overexpression of insulin-like growth factor-1 enhances mammary tumorigenesis in vivo. Am J Pathol. 2008; 173: 824-834.

21. Boldt HB and Conover CA. Overexpression of Pregnancy-associated plasma protein-A in ovarian cancer cells promotes tumor growth in vivo. Endocrinology. 2011; 152: 1470-1478.

22. Pan H, Hanada S, Zhao J, Mao L, Ma MZ. Protein secretion is required for pregnancy-associated plasma protein-A to promote lung cancer growth in vivo. PLoS One. 2012; 7: e48799. doi:10.1371/journal.pone.0048799

23. Ryan AJ, Napoletano S, Fitzpatrick PA, Currid CA, O'Sullivan NC, Harmey JH. Expression of a protease-resistant insulin-like growth factor-binding protein-4 inhibits tumour growth in a murine model of breast cancer. Br J Cancer. 2009; 101: 278-286.

24. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, et al. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat Genet. 2009; 41: 178-186.

25. Lettice LA, Heaney SJ, Purdie LA, Li L, de Beer P, Oostra BA, et al. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum Mol Genet. 2003; 12: 1725-1735.

26. Benko S, Gordon CT, Mallet D, Sreenivasan R, Thauvin-Robinet C, Brendehaug A, et al. Disruption of a long distance regulatory region upstream of SOX9 in isolated disorders of sex development. J Med Genet. 2011; 48: 825-830.

27. Smits BM, Haag JD, Rissman AI, Sharma D, Tran A, Schoenborn AA, et al. The gene desert mammary carcinoma susceptibility locus Mcs1a regulates Nr2f1 modifying mammary epithelial cell differentiation and proliferation. PLoS Genet. 2013; 9: e1003549. doi:10.1371/journal.pgen.1003549

28. Acevedo N, Reinius LE, Greco D, Gref A, Orsmark-Pietras C, Persson H, et al. Risk of childhood asthma is associated with CpG-site polymorphisms, regional DNA methylation and mRNA levels at the GSDMB/ORMDL3 locus. Hum Mol Genet. 2015; 24: 875-890.

29. Bockmühl Y, Patchev AV, Madejska A, Hoffmann A, Sousa JC, Sousa N, et al. Methylation at the CpG island shore region upregulates Nr3c1 promoter activity after early-life stress. Epigenetics. 2015; 10: 247-257.

30. Farkas SA, Böttiger AK, Isaksson HS, Finnell RH, Ren A, Nilsson TK. Epigenetic alterations in folate transport genes in placental tissue from fetuses with neural tube defects and in leukocytes from subjects with hyperhomocysteinemia. Epigenetics. 2013; 8: 303-316.

31. Park JL, Kim HJ, Seo EH, Kwon OH, Lim B, Kim M, et al. Decrease of 5hmC in gastric cancers is associated with TET1 silencing due to with DNA methylation and bivalent histone marks at TET1 CpG island 3'-shore. Oncotarget. 2015; 6: 37647-37662.

32. Perisic T, Holsboer F, Rein T, Zschocke J. The CpG island shore of the GLT-1 gene acts as a methylation-sensitive enhancer. Glia. 2012; 60: 1345-1355.

33. Rao X, Evans J, Chae H, Pilrose J, Kim S, Yan P, et al. CpG island shore methylation regulates caveolin-1 expression in breast cancer. Oncogene. 2013; 32: 4519-4528.

34. Sugawara H, Iwamoto K, Bundo M, Ueda J, Miyauchi T, Komori A, et al. Hypermethylation of serotonin transporter gene in bipolar disorder detected by epigenome analysis of discordant monozygotic twins. Transl Psychiatry. 2011; 1: e24. doi:10.1038/tp.2011.26

35. Yang R, Pfütze K, Zucknick M, Sutter C, Wappenschmidt B, Marme F, et al. DNA methylation array analyses identified breast cancer-associated HYAL2 methylation in peripheral blood. Int J Cancer. 2015; 136: 1845-1855.

36. Schmidt D, Schwalie PC, Wilson MD, Ballester B, Gonçalves A, Kutter C, et al. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. Cell. 2012; 148: 335-348.

37. Bao L, Zhou M and Cui Y. CTCFBSDB: a CTCF binding site database for characterization of vertebrate genomic insulators. Nucleic Acids Research. 2008; 36: D83-D87.

38. Ziebarth JD, Bhattacharya A, Cui Y. CTCFBSDB 2.0: a database for CTCF-binding sites and genome organization. Nucleic Acids Research. 2013; 41: D188-D194.

39. Phillips JE and Corces VG. CTCF: master weaver of the genome. Cell. 2009; 137: 1194-1211.

40. The ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature. 2007; 447: 799-816.

41. Stamatoyannopoulos JA, Snyder M, Hardison R, Ren B, Gingeras T, Gilbert DM, et al. An encyclopedia of mouse DNA elements (Mouse ENCODE). Genome Biol. 2012; 13:418. doi: 10.1186/gb-2012-13-8-418

42. Kleinberg DL and Barcellos-Hoff MH. The pivotal roal of Insulin-like growth factor I in normal mammary development. Endocrinol Metab Clin N Am. 2011; 40: 461-471.

43. Russo J, Tay LK, Ciocca DR, Russo IH. Molecular and cellular basis of the mammary gland susceptibility to carcinogenesis. Environmental Health Perspectives. 1983; 49: 185-199.

44. Tian J, Berton TR, Shirley SH, Lambertz I, Gimenez-Conti IB, DiGiovanni J, et al. Developmental stage determines estrogen receptor alpha expression and non-genomic mechanisms that

control IGF-1 signaling and mammary proliferation in mice. J Clin Invest. 2012; 122: 192-204. doi: 10.1172/JCI42204

45. Interagency Breast Cancer and Environmental Research Coordinating Committee (IBCERCC). Breast Cancer and the Environment: Prioritizing Prevention. 2013. Available: https://www.niehs.nih.gov/about/assets/docs/breast_cancer_and_the_environment_prioritizing_prevention_508.pdf

46. Vijayendran M, Beach SRH, Plume JM, Brody GH, Philibert RA. Effects of genotype and child abuse on DNA methylation and gene expression at the serotonin transporter. Front Psychiatry. 2012. 3:55. doi: 10.3389/fpsyt.2012.00055

47. Voisin S, Almén MS, Moschonis G, Chrousos GP, Manios Y, Schiöth HB. Dietary fat quality impacts genome-wide DNA methylation patterns in a cross-sectional study of Greek preadolescents. Eur J Hum Genet. 2015; 23: 654-662.

48. Whitehead RH, Demmler K, Rockman SP, Watson NK. Clonogenic growth of epithelial cells from normal colonic mucosa from both mice and humans. Gastroenterology. 1999; 117: 858-865.

49. Smits BM, Traun BD, Devries TL, Tran A, Samuelson D, Haag JD, et al. An insulator loop resides between the synthetically interacting elements of the human/rat conserved breast cancer susceptibility locus MCS5A/Mcs5a. Nucleic Acids Res. 2012; 40: 132-147.

50. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. Nat Protoc. 2013; 8: 2281-2308.

Chapter 3

Expression of the orphan nuclear receptor, *NR2F1*, in TNBC cells reduces carcinoma growth *in vivo* through the promotion of a luminal molecular profile

Amanda N. Henning[1], Sid H. Kiblawi[2], Irene Ong[3], Jill D. Haag[1], Mark Craven[2], Michael N. Gould[1]

[1]Department of Oncology, McArdle Laboratory for Cancer Research, University of Wisconsin-Madison School of Medicine and Public Health, Madison, WI, USA

[2]Department of Computer Sciences, Department of Biostatistics and Medical Informatics, USA Genome Center of Wisconsin, University of Wisconsin-Madison, Madison, WI, USA

[3]Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison. University of Wisconsin Carbone Comprehensive Cancer Center, University of Wisconsin-Madison, WI, USA

Abstract

Breast cancer mortality rates have declined in recent years thanks, in part, to the advent of targeted therapies that have shown substantial benefit for hormone receptor-positive and HER2+ clinical subtypes. Unfortunately, the triple-negative class of breast cancers, lacking expression of commonly targeted receptors, have been left out of such progress, and their overall prognosis remains worse by comparison. Triple negative breast cancer (TNBC) tends to be more aggressive and affect younger women than its receptor-positive counterparts, and the identification of targeted therapies for this clinical subtype remains a high priority. Previous work on a carcinogen-induced rat model of breast cancer has identified *Nr2f1/Coup-TFI* expression as an important factor mediating the activity of the Mammary carcinoma susceptibility locus, *Mcs1a*. In this model, *Nr2f1* expression promoted the luminal differentiation of mammary epithelial cells that displayed a reduced proliferative capacity. Comparative analysis in humans suggested NR2F1's potential as a therapeutic agent in TNBC, and we sought to test this hypothesis *in vivo* using a TNBC cell line with inducible *NR2F1*-overexpression. *NR2F1*-overexpression in the triple-negative MDA-MB-468 cell line resulted in a reduction in carcinoma volume and growth rate in a xenograft model. Expression analysis identified upregulation of the classical luminal marker *FOXA1*, and downregulation of a number of triple-negative/basal-associated genes, including *NGFT* and *TNFAIP2*, as well as a novel association with the vacuolar-ATPase complex genes. These results suggest that NR2F1 may reduce carcinoma growth through the promotion of a less aggressive, luminal molecular profile, or the "luminalization" of TNBC cells. Network mapping of regulatory interactions identified a number of central regulatory pathways by which NR2F1 may mediate differential expression of target genes, with FOXA1 representing a critical node. This work supports the therapeutic potential of NR2F1 in TNBC, and suggests that the luminalization of triple-negative tumors may represent a relevant therapeutic strategy.

Introduction

Breast cancer is the second leading cause of cancer death among women in the United States; however, cancer mortality rates have been on a steady decline in the past decades, with overall rates decreasing by 36% between 1989 and 2012 [1]. This reduction is attributable, in part, to advances in both cancer detection and treatment [2]. The advent of targeted therapies has been particularly beneficial for the treatment of hormone receptor-positive and HER2+ breast cancers. 5-year adjuvant treatment of ER+ breast cancers with the selective estrogen receptor modulator tamoxifen has been found to reduce the recurrence and mortality rate by 50% and 30%, respectively [3], while treatment of HER2+ breast cancers with the HER2-antagonist trastuzumab has resulted in a 50% improvement in disease-free survival interval [4]. Treatment options for the triple-negative class of breast cancers, which lack expression of ER, PR, and HER2, however, are limited. Current treatment strategies for triple-negative breast cancers (TNBCs) consist of systemic cytotoxic chemotherapy, but even with treatment, 3-year overall survival rates are lower and recurrence rates higher for TNBC compared to non-TNBC patients [5]. Clearly, to achieve the success that endocrine therapies and HER2-antagonists have had in the treatment of hormone receptor-positive and HER2+ breast cancer, similarly targeted therapies for TNBC need to be developed.

In an effort to identify new therapeutic agents for breast cancer treatment and prevention, we have taken a comparative genomics approach, using a carcinogen-induced rat model of breast cancer. The inherent susceptibility differences among inbred rat strains to carcinogen-induced mammary cancer were exploited in quantitative trait loci (QTL) analyses between the susceptible Wistar-Furth (WF) and resistant Copenhagen (COP) and Wistar-Kyoto (WKy) rat strains. This led to the identification of 8 Mammary carcinoma susceptibility (*Mcs1-8*) loci capable of affecting carcinoma number following carcinogen treatment, with several loci containing subloci characterized through further analysis [6–11].

Functional characterization of the *Mcs5a* locus [12,13] has resulted in the identification of novel therapeutic prevention and treatment strategies that are currently being explored in women [14] and validate the translational potential of comparative studies. Recently published work on the non-coding *Mcs1a* locus has suggested the therapeutic relevance of its target gene *Nr2f1/Coup-TFI* [15]. Characterization of the *Mcs1a* locus by Smits *et al* [15] found that congenic rat lines homozygous for the resistant COP allele introgressed on a WF background showed an approximately 50% reduction in carcinoma number in both a chemical carcinogen and oncogene-induced mammary cancer model. The non-coding *Mcs1a* locus affects expression of *Nr2f1* in the mammary gland through a long-range chromatin interaction between the gene and a conserved Nr2f1 binding site within the locus. This site is involved in self-repression of *Nr2f1* transcription and is disrupted in the resistant allele, leading to increased expression of *Nr2f1* in resistant rats. The mammary gland of resistant rats displayed an increase in the luminal cell population, whose cells exhibited a reduced proliferative capacity compared to the luminal cells of susceptible rats. These data suggested that *Nr2f1* may play a role in mammary gland proliferation and differentiation, pushing the gland towards a more luminal phenotype. RNA-sequencing (RNA-seq) analysis on mammary glands from a mouse homologous *Mcs1a* region megadeletion model revealed that *Nr2f1* was anti-correlated with genes involved in the cell cycle, proliferation, and DNA-damage response. These results were recapitulated in publically available human breast cancer microarray data when orthologous genes were examined, and further support a role for *Nr2f1* in cell proliferation. Finally, *NR2F1* expression was examined in relation to a number a clinical features, and was found to decrease with increasing histological grade and was lower in TNBC compared to receptor-positive cancers. Both high histological grade and triple-negative status are indicative of aggressive, poorly differentiated and highly proliferative breast cancers, and this association further supports a role for *Nr2f1/NR2F1* in mammary gland/breast proliferation and differentiation and highlights NR2F1 as a target for therapeutic intervention.

TNBCs account for approximately 12% of all breast cancer cases [1] and occur more frequently in younger women and African American patients [16,17]. They generally have a higher histological grade and worse prognosis when compared to other clinical subtypes [16,18]. The correlation of low *NR2F1* transcript levels with high-grade/TNBC-status, as well as its inverse association with mammary gland proliferation, suggested that modification of the NR2F1 pathway in TNBC may result in the promotion of a less aggressive, more treatable phenotype. NR2F1 is an orphan nuclear receptor within the steroid/thyroid hormone receptor superfamily of proteins. It can act as both an activator and repressor of transcription, and frequently dimerizes with a number of other nuclear receptors, particularly the retinoid X receptor (RXR). It is NR2F1's interactions with other transcription factors or coactivators present in the cell that appear to control its regulatory action [19–23]. During mouse embryogenesis, *Nr2f1* is highly expressed in a distinct pattern within the central nervous system, and knock-out mice die perinatally with severe neuronal developmental defects [24,25]. *Nr2f1,* therefore, appears to be necessary for proper neuronal development and differentiation, and our work on the *Mcs1a* locus suggests that *Nr2f1* may have a similar function in the mammary gland. Studies on the role of NR2F1 in breast cancer are limited. In one study, NR2F1 was found to mediate retinoic acid-induced expression of Retinoic acid receptor, beta (*RARB*) in a Retinoic acid receptor, alpha (RARA)-dependent manner in a TNBC cell line. This interaction resulted in growth inhibition and apoptosis *in vitro* [26]. In a separate study, NR2F1 was identified as part of a 49-gene signature for tumor cell dormancy, with high expression associated with a higher dormancy score. Dormancy scores tended to be higher in ER+ cancers and showed an inverse correlation with proliferation in ER+ cell lines. Knockdown of *NR2F1* mRNA in ER+ MCF-7 cells resulted in an increase in cell growth in a xenograft model [27].

Given the association of NR2F1 with decreased proliferation and less aggressive clinical subtypes, we hypothesized that increasing *NR2F1* expression in TNBC may have therapeutic potential as a targeted treatment for this class of tumors. We have overexpressed *NR2F1* in the triple-negative MDA-MB-468 breast cancer cell line and observed a reduction in carcinoma volume and growth rate *in vivo* for cells overexpressing *NR2F1*. Our data support NR2F1 as a possible therapeutic agent in the treatment of TNBC, and network modeling of *NR2F1*-dependent differentially expressed genes has identified relevant signaling pathways and proteins that may hold additional therapeutic benefits.

Results

MDA-MB-468 cells engineered to overexpress *NR2F1* in an inducible manner

NR2F1 expression levels were investigated in three triple-negative breast cancer (TNBC) cell lines. MDA-MB-231 and BT-549 both showed considerable *NR2F1* expression, while expression in MDA-MB-468 was significantly lower, although not negligible (Fig 1A). MDA-MB-468 was therefore the cell line chosen for *NR2F1* engineering. Cells were stably transfected with two plasmids encoding either a constitutively active tetracycline repressor (TetR) protein or the *NR2F1* gene proceeded by two copies of the tetracycline operator sequence. Without tetracycline/doxycycline, TetR binds the operator sequences, preventing *NR2F1* expression. With addition of tetracycline (tet) or doxycycline (dox), TetR is removed and *NR2F1* expression is induced [28,29]. Three tet-inducible clones were isolated, and analysis of NR2F1 protein and RNA levels indicated two were positive for tet-inducible *NR2F1* expression (Fig 1B and C). Positive clones C6.3 and C6.8 both showed significant induction of NR2F1 protein levels and RNA, with C6.8 exhibiting a higher *in vitro* response than C6.3.

Fig 1: Inducible overexpression of *NR2F1* in TNBC cell line. (A) Endogenous *NR2F1* expression was assessed in 3 triple negative breast cancer (TNBC) cell lines using quantitative real-time PCR (qPCR).

Average expression levels are shown relative to expression of the reference gene *PPIA*. (B) Protein levels of NR2F1 are shown in three MDA-MB-468 clones stably transfected with the *tetR* gene and a tetracycline (tet)-inducible *NR2F1* expression vector. Clones C6.3 and C6.8 were positive for NR2F1 induction, while C6.5 was negative. A total of 10µg of protein was used for western blot analysis. (C) The fold change of *NR2F1* mRNA levels in positive clones following tet administration was examined by qPCR. *PPIA* was used as the reference gene. For both (B) and (C), cells were exposed to 1µg/mL tet for 48 hours prior to cell collection.

NR2F1 overexpression in TNBC cell line slows carcinoma growth in a xenograft model

Preliminary xenograft experiments were performed with both C6.3 and C6.8 cells. Although C6.8 cells showed strong *NR2F1* induction *in vitro*, they failed to grow sizable carcinomas *in vivo*, regardless of dox treatment (data not shown). C6.3 cells showed appreciable carcinoma growth, and were therefore used in further xenograft experiments. A total of 37 nude mice were injected with 2 million *NR2F1*-inducible C6.3 cells. Half of the mice were sorted into the treatment group (n = 18), receiving a dox-containing diet, while the other half remained on a standard rodent diet (n = 19). Carcinoma cells in mice fed a dox-containing diet to induce overexpression of *NR2F1* (*NR2F1*-high) achieved a lower final carcinoma volume than cells with non-induced *NR2F1* expression (*NR2F1*-low). At necropsy, 45 days post-injection, *NR2F1*-high carcinomas were 37% smaller than *NR2F1*-low carcinomas (Fig 2A; p-value = 0.026). When the growth rates (mm^3/day) of *NR2F1*-high and *NR2F1*-low carcinomas were examined in 3 approximately 15-day growth windows throughout the course of the experiment, a statistically significant increase in growth rate was observed between the second and third growth periods of the study for *NR2F1*-low cells (Fig 2B; p-value = 5.76e-6). Conversely, growth rates remain relatively steady for *NR2F1*-high cells. During the final growth period, a statistically significant *NR2F1*-dependent difference in growth rates was observed, with *NR2F1*-high carcinomas growing at a rate of 3.9mm^3/day,

while the growth rate for *NR2F1*-low carcinomas was 1.6mm^3/day (2.4-fold difference, p-value = 0.0022). As a control experiment, mice were injected with cells containing only the *tet*R plasmid, constitutively expressing the TetR protein, but no *NR2F1*-containing inducible construct (TR1 cells). No difference in carcinoma volume was detected between dox and no dox diets, suggesting that the presence of doxycycline does not affect carcinoma growth (S1 Fig).

Fig 2: NR2F1 reduces TNBC growth in a xenograft model. (A) 2 million C6.3 cells containing a constitutively active *tet*R plasmid and doxycycline (dox)-inducible *NR2F1* expression plasmid were injected into the hind flank of 6 week old female nude mice. 24 hours post-injection, animals were placed on a dox-containing (*NR2F1*-high/Dox; n=18) or standard (*NR2F1*-low/no Dox; n=19) rodent diet. Carcinoma volume (mm^3, y-axis) was measured with a caliper using the formula ½(L*W*H) over the course of 45 days (x-axis). The difference in carcinoma volume at the final time point was statistically significant between the two groups. (B) The 45 day measurement period was divided into 3 growth windows consisting of the first (0-14 days post-injection, dpi), second (15-27 dpi), and the third (28-45 dpi) growth periods of the study (x-axis). The growth rate (mm^3/day, y-axis) was calculated by taking the average growth rate between measurement points within a growth period for all samples included in (A). For both (A) and (B) standard error bars are shown, and p-values were obtained using the non-parametric Mann Whitney U test (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$).

RNA sequencing and GO term analysis of *NR2F1*-high and -low xenograft carcinomas

RNA sequencing (RNA-seq) was performed on a total of 8 xenograft carcinomas collected at necropsy, 4 from the *NR2F1*-low group and 4 from the *NR2F1*-high dox treatment group. An average of 20.2 million reads/sample were obtained, and mapping to the human reference genome (GRCh37/hg19) was done using RSEM. Differentially expressed (DE) genes were identified using the EdgeR software

package in R. A total of 356 genes were found to be DE between the *NR2F1*-high and -low carcinoma samples (adjusted p-value <0.05), with 218 genes found to be upregulated in *NR2F1*-high carcinomas, and 138 genes downregulated. The top 40 up and downregulated genes by adjusted p-value are shown in Table 1, and a list of all DE genes is available in S3 Table. Select genes were validated by qPCR in both the samples used for RNA-seq analysis as well as additional *NR2F1*-high and -low carcinoma samples (n = 11 and 12, respectively; Fig 3).

Fig 3: qPCR validation of differentially expressed genes. A selection of differentially expressed genes identified in our RNA-seq analysis were verified through quantitative real-time PCR (qPCR) analysis. Gene expression was examined in the 8 carcinoma samples used for RNA-seq (4 *NR2F1*-high/Dox and 4 *NR2F1*-low/no Dox), as well as additional xenograft samples for a total of 11 *NR2F1*-high/Dox and 12 *NR2F1*-low/no Dox samples. Relative gene expression levels were standardized to the reference gene *PPIA*, and standard error bars are shown. P-values were obtained using the non-parametric Mann Whitney U test (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$).

Functional annotation clustering of the up and downregulated gene sets was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) [30,31]. The clustering of annotation terms by their functional relatedness and inclusion of similar genes helps to maximize the biological meaning obtained from gene sets. A group enrichment score of >1.3 was indicative of a statistically significant enrichment cluster, representing the geometric mean (-log scale) of cluster-associated annotation term p-values that is <0.05. Both up and downregulated gene sets saw clusters enriched for secreted proteins, signal peptides, extracellular proteins, plasma membrane proteins, and proteins involved in cell adhesion (Tables 2 and 3). Unique clusters in the upregulated gene set saw enrichment for terms related to tissue development (Cluster 4) and proliferation and apoptosis (Cluster

15) (Table 2). Predictably, Forkhead box A1 (*FOXA1*), one of our top upregulated genes and a known driver of mammary gland development [32], was included in Cluster 4. In Cluster 15, terms for both positive and negative regulation of apoptosis are seen, with the same genes appearing in multiple GO categories, suggesting that their involvement in apoptosis is likely context dependent. Cluster enrichment scores for the downregulated gene set were generally lower than those for the upregulated genes, and overlapped in functionality (Table 3). Cluster 17, a trending but not significantly enriched cluster, stood out as it was a large cluster of 56 annotation terms that were grouped together almost exclusively based on the inclusion of four genes that are components of the vacuolar-ATPase complex. These genes, *ATP6V0D2*, *ATP6V1C2*, *ATP6V1B1*, and *ATP6V0A4*, are downregulated between 2 and 4 fold in *NR2F1*-high carcinomas (S3 Table), and represent a novel association with *NR2F1* expression. Lists of the top 25 functional annotation clusters for the up and downregulated gene sets can be found in S4 and S5 Tables (electronic version only).

Network mapping of *NR2F1*-dependent differentially expressed genes

NR2F1 acts as a transcription factor, although its function is highly dependent on interactions with other transcription factors and co-factors present in the cell. In an effort to identify through which pathways NR2F1 may be regulating the observed DE genes, we created a regulatory network map (Fig 4A). To do this, we integrated our RNA-seq data with transcription factor binding site data, protein-protein interactions, and literature-based data mining. These data were combined and the network visualized using the Cytoscape software platform [33]. To begin, ChIP-chip data from Kittler *et al* [34] on NR2F1 and other nuclear receptor transcription factors were combined with ENCODE ChIP-seq data to optimize a set of gene regulatory proteins (GRPs) (Fig 4, pink boxes) capable of interacting with the maximum number of DE genes. These direct regulatory interactions are indicated by the red arrows in Fig 4. Next, to connect the GRPs to NR2F1, intermediary factors (IFs) (Fig 4, orange boxes) were added

based on protein-protein interactions, represented by the green lines in Fig 4. Relationships between genes and proteins identified via literature-based data mining are represented by purple lines. These interactions may be direct or indirect, and add an additional layer of information to our regulatory network. These proteins and their associated interactions form the basic backbone of our network, and is shown in Fig 4B. The only DE gene that appears in the backbone network, aside from NR2F1, is FOXA1 (red box). We have included it here with a unique color as it is both a major GRP and an upregulated DE gene. Our full network was capable of reaching 330 of the 356 DE genes (92.7%; Fig 4A), with up and downregulated genes indicated in blue and light blue, respectively. There is a moderate amount of inter-connectively within the network, with each DE gene the target of an average of 3.3 GRPs. Likely, all possible pathways to a DE gene shown in the network are not functionally relevant, but this provides for a starting point to detect relevant transcriptional nodes and pathways. For instance, an examination of the number of direct targets for each GRP (Table 4) shows that FOS, FOXA1, and RARA represents critical nodes in our network, hitting 43%, 47%, and 59% of all DE network-mapped genes. To examine the possible pathways from NR2F1 to specific DE genes of interest, target genes can be placed into the backbone network with all relevant interactions present, but excluding all other DE genes, allowing for a more readable network. This was done for the top upregulated gene Chitinase 3-Like 1 (*CHI3L1*; Fig 4C) and for the four vacuolar-ATPase complex genes, *ATP6V0D2*, *ATP6V1C2*, *ATP6V1B1*, and *ATP6V0A4* (Fig 4D). From these simplified figures we can see that *CHI3L1* is a direct binding target of a number of GRPs, including FOXA1 and NR2F1. This latter interaction pathway represents the most direct way for NR2F1-mediated regulation of *CHI3L1*, and perhaps the most functionally relevant interaction. Two of the ATPase genes are direct targets of FOXA1 binding (*ATP6V0D2* and *ATP6V1C2*), and their expression may therefore be affected by NR2F1 regulation of *FOXA1* expression. Alternatively, both *ATP6V0D2* and *ATP6V0A4* are targets of RARA, PPARG, and GATA3 binding, and may be co-regulated through a NR2F1-containing protein complex. This kind of visualization of possible NR2F1 regulatory interactions allows

for the identification of potential critical nodes and regulatory pathways mediating regulation of DE genes and provides hypotheses that can be tested *in vitro* and used to refine the network.

Fig 4: Multi-platform network mapping of regulatory interactions between NR2F1 and DE genes. The Cytoscape software platform was used to visually depict the integrated RNA-seq, ChIP-chip/seq, and literature-mined data used to create a regulatory network of possible NR2F1 interactions with DE genes. For all network maps, proteins are depicted in colored boxes with categories consisting of gene regulatory proteins (GRPs, pink), intermediary factors (IFs, orange), upregulated DE genes (blue), and downregulated DE genes (light blue). Exceptions to this are NR2F1 (gray box) and the gene regulatory protein/upregulated DE gene FOXA1 (red box). Protein-protein interactions are indicated by green lines. Protein-DNA and literature-mined interactions are indicated by the red and purple arrows, respectively, with the arrow head pointing to the regulatory target. Protein-DNA interactions were pulled from publically available ChIP-chip and ChIP-seq datasets and represent direct regulatory interactions, while literature interactions were mined from research abstracts and may be direct or indirect in nature. (A) The full NR2F1 network showing all GRPs, IFs, and DE genes and their interactions. (B) The backbone NR2F1 network showing only GRPs and IFs and their interactions. (C) DE gene-specific network showing the possible regulatory pathways from NR2F1 to the DE gene *CHI3L1*. (D) DE gene-specific network showing the possible regulatory pathways from NR2F1 to the DE vacuolar-ATPase complex genes *ATP6V0D2*, *ATP6V1C2*, *ATP6V1B1*, and *ATP6V0A4.*

Discussion

In the present study, we have demonstrated that overexpression of *NR2F1* in the TNBC cell line MDA-MB-468 (468) resulted in a 37% decrease in carcinoma volume and 2.4-fold reduction in growth rate *in vivo* (Fig 2). RNA-seq analysis of *NR2F1*-high and -low carcinomas identified 356 differentially

expressed genes resulting from *NR2F1* overexpression, with significant enrichment for secreted and cell membrane proteins. Predictably, upregulated genes were enriched for proteins involved in tissue development, with one of the top DE genes, *FOXA1*, a known driver of mammary gland development whose expression is associated with luminal, ER+ breast cancers and low tumor grade [32]. Downregulated genes showed enrichment for proteins that form the vacuolar-ATPase (V-ATPase) complex, representing a novel association between these genes and *NR2F1* expression, and identifying a potential druggable target for TNBC. A network building strategy was developed to identify possible pathways and binding partners responsible for the *NR2F1*-dependent expression differences observed. This network incorporated our expression data with publically available ChIP-chip and ChIP-seq data, known protein-protein interactions, and literature-based text mining. Through investigation of the network via the Cytoscape interface, central hubs and transcription factors were identified. FOS, FOXA1, and RARA represent critical gene regulatory proteins (GRPs), as they are proposed to interact with a large percentage of DE genes (Table 4), and the NR2F1 binding partners RXRA and nuclear receptor corepressor 2 (NCOR2) represent central hubs, as they interact with a number of other GRPs and IFs in the network (Fig 4B).

Our third most significantly upregulated gene, *FOXA1*, stood out due to its well characterized relationship with breast development and breast cancer. FOXA1 is a transcription factor involved in the development of a number of epithelial tissues [32], and its expression in the mammary gland is required for proper ductal epithelial development through the regulation of ERα expression [35]. FOXA1 in breast cancer is predictably associated with ERα expression and also with low histological grade, increased overall survival, and, more broadly, with the luminal A molecular subtype [36]. However, high FOXA1 levels are still an independent predictor of improved overall prognosis when patients with luminal [36] or ERα-negative tumors [37] were examined independently. The luminal A subtype, along with the

luminal B, HER2+, basal-like, and normal breast-like subtypes, were identified through gene expression profiling of breast cancer samples [38,39]. Generally speaking, luminal tumors have expression profiles that mimic luminal breast epithelial cells, while basal-like tumors are triple-negative histologically and have expression patterns characteristic of basal epithelial and myoepithelial breast cells [18,40]. In our study, overexpression of *NR2F1* resulted in an 11-fold increase in *FOXA1* expression (Table 1), likely through the direct targeting of the *FOXA1* gene by NR2F1 (Fig 4B). Additionally, FOXA1 represents a major regulatory node in our network, targeting 47% of all DE genes (Table 4). These data together with FOXA1's known association with a luminal phenotype, suggest that FOXA1 may be an important mediator of the increased *Nr2f1*-associated luminal cell differentiation previously described in the rat [15]. Furthermore, it suggests that the reduction in growth observed in *NR2F1*-high TNBC cells may be due to the promotion of a more luminal, less aggressive phenotype through increased *FOXA1* expression.

Our focus on FOXA1 was further bolstered by the recent work of Marcotte and colleagues [41], who identified potential drivers of breast cancer cell survival through shRNA dropout screens performed on 77 breast cancer cell lines. The breadth of cell lines used allowed for the clustering of cell lines into major clinical molecular subtypes (luminal A, luminal B, HER2+, and basal) and allowed for the identification of subtype-specific driver genes, whose expression is essential to cell line survival. In this essentiality analysis, three of our top 40 significantly upregulated genes, *FOXA1*, *CHI3L1*, and Nephronectin (*NPNT*), were found to be essential in the luminal/HER2+ molecular subtype. *FOXA1* was the top essential gene in this author-defined subtype, and also in the luminal/HER2- molecular subtype based on the Neve classification [42] and in the luminal androgen receptor (LAR) molecular subtype based on the Lehmann classification [40]. Breast tumors in the LAR subtype are histologically triple-negative, but are enriched for the expression of genes involved in hormonally-regulated pathways that

typically define the luminal subtype, including *FOXA1* [40]. Importantly, both *CHI3L1* and *NPNT* are direct targets of FOXA1, and *CHI3L1* is also a direct target of NR2F1. CHI3L1 is a secreted protein that has been shown to be expressed by luminal mammary epithelial cells [43]. Its expression is increased during involution, where overexpression inhibits differentiation and polarization in a lactogenic hormone-dependent manner *in vitro*; however, these results were not observed in a normal mammary epithelial cell line *in vivo.* CHI3L1 is therefore not considered to be tumorigenic [43]. Examination of CHI3L1 in breast cancer samples has had conflicting results, as it has been both negatively [44,45] and positively [46] correlated with luminal cancer cell markers ER and PR. NPNT is a secreted, extracellular matrix protein whose role, if any, in normal breast development is not known, but is involved in kidney development [47] and the promotion osteoblast differentiation [48]. These three genes, therefore, function broadly in tissue development and/or differentiation and the possibility of their involvement in luminal epithelial cell-specific development/differentiation, whether confirmed (FOXA1) or unconfirmed (CHI3L1 and NPNT), is bolstered by their classification as luminal-essential genes by Marcotte *et al* [41]. Their upregulation following *NR2F1* overexpression and their direct regulation by either FOXA1 or NR2F1, further supports our hypothesis that increased *NR2F1* expression in TNBC 468 cells is facilitating reduced tumor growth through the promotion of a less-aggressive, more luminal molecular profile resulting in the "luminalization" of TNBC cells.

Examination of significantly downregulated DE genes (Table 1) identified a number of genes implicated in cancer cell survival. The top most significant gene Interferon, alpha-inducible protein 27 (*IFI27*), as well additional genes Monoamine oxidase A (*MAOA*), Nerve growth factor receptor (*NGFR*), and Tumor necrosis factor, alpha-induced protein 2 (*TNFAIP2*) are all membrane proteins whose expression is associated with cell proliferation, migration, and invasion in breast cancer or other cancer types [49–52]. Additionally, expression of *NGFR* and *TNFAIP2* are associated with the basal/TNBC

phenotype [52,53]. The association of these genes with tumor cell proliferation and invasion and their localization to the plasma membrane suggests that the therapeutic effect of *NR2F1* overexpression may also be due, in part, to the modulation of the extracellular environment through downregulation of these, and other, membrane/extracellular-associated proteins. Furthermore, their positive association with TNBC tumors further supports our assertion that *NR2F1* overexpression results in a shift away from basal-related gene expression to a more luminal molecular profile, or "luminalization", that gives rise to a less-aggressive phenotype.

Functional annotation clustering (Table 3) identified a unique cluster consisting predominately of four V-ATPase complex genes, *ATP6V0D2*, *ATP6V1C2*, *ATP6V1B1*, and *ATP6V0A4,* that were downregulated between 2.2 and 4.3 fold with *NR2F1* overexpression (Table 1, S3 Table). The V-ATPase is a large, multi-subunit complex responsible for the acidification of either intracellular compartments or the extracellular environment through the hydrolysis of ATP. It consists of the transmembrane V_0 and cytoplasmic V_1 domains, which each contain multiple subunits and isoforms [54]. The V_0 *a* subunit is responsible for localization of the complex [54], and knockdown of DE gene *ATP6V0A4* has been shown to reduce plasma membrane V-ATPase complex localization in the TNBC MDA-MB-231 cell line [55]. This suggests that plasma membrane V-ATPase localization may be compromised in 468 cells following *NR2F1*-overexpression. In breast cancer, knock down of individual subunits has resulted in decreased tumor growth *in vivo* [56] and cell invasiveness *in vitro* [55,57], and targeting of the V-ATPase complex through inhibitory drugs has resulted in reduced invasion of TNBC cells *in vitro* [58] and pancreatic cancer cell growth *in vivo* [59]. Interestingly, *in vivo* growth reduction was only observed once carcinomas became large, which mirrors our *in vivo* growth results showing *NR2F1*-high tumors had a reduced growth rate only in the latter growth period of the study (Fig 2B). The importance of V-ATPase genes was further supported by the essentiality analysis of Marcotte and colleagues [41], as they

identified *ATP6V1B2* as the third most essential gene for the basal-like molecular subtype. Although this gene was not DE in our study, we did observe differential expression of its isoform, *ATP6V1B1*, as well as three other complex genes, with two genes directly targeted by the DE transcription factor FOXA1 (Fig 4D). Together, these results suggest that V-ATPase activity is important for TNBC/basal cell survival, and that *NR2F1*-mediated downregulation of these genes may contribute to the reduction in carcinoma growth observed *in vivo*, making them a potential target for TNBC therapies.

Altogether, the examination of DE genes resulting from overexpression of *NR2F1* in the triple-negative 468 cell line has identified a number of promising therapeutic targets and strategies for TNBC treatment. We are particularly interested in the possibility that *NR2F1* expression results in a shift from a basal to more luminal-like molecular profile. This may partially explain the observed reduction in carcinoma growth, as luminal tumors have a less aggressive clinical prognosis. It is also possible that this *NR2F1*-induced change in molecular profile may make cells more amenable to therapies not previously associated with basal breast cancer, and raises the possibility of utilizing combination therapies. To this end, network modeling of DE genes has helped in identifying regulatory pathways that may be exploited for therapeutic purposes. A caveat of our modeling strategy is that many of our protein-DNA interactions came from analysis in the ER+ MCF-7 breast cancer cell line [34], and interactions would therefore need to be verified in triple-negative cells to ensure ER is not a necessary cofactor. Even so, our network provides a useful model for investigation of regulatory interactions and patterns, and was used to identify RARA as a critical gene regulatory protein. Lin *et al* [26] have previously demonstrated that the anti-cancer effects of retinoic acid are mediated by NR2F1- and RARA-dependent induction of *RARB* expression. Indeed, in our analysis *RARB* was upregulated 6.2-fold in *NR2F1*-high cells (adj p-value = 3.76E-5; S3 Table), suggesting that the overexpression of *NR2F1* may result in cells that are more sensitive to the effects of retinoic acid treatment. Importantly, while NR2F1 is an orphan nuclear

receptor with an as of yet unidentified ligand, analysis of its ligand binding domain shows near identical homology with its subfamily member, NR2F2 [20]. The crystal structure of NR2F2 has been identified, and its function appears to be ligand-regulated [60]. Due to their high sequence homology, it is therefore likely that NR2F1 functions in a similarly ligand-dependent manner. This suggests the possibility that a ligand or other molecule could be identified that could enhance NR2F1 activity in TNBC to achieve the therapeutic "luminalization" effects achieved in our overexpression model. These hypotheses, and others, could be tested *in vitro* and *in vivo* using our *NR2F1*-inducible system, and emphasize the potential of this approach to identify and validate combination therapies for TNBC treatment.

Materials and Methods

Cell culture conditions and creation of inducible *NR2F1* overexpression cell lines

Triple negative human breast adenocarcinoma cells, MDA-MB-468, were obtained from the American Type Culture Collection and maintained in DMEM/F12 supplemented with 100 IU/mL penicillin, 100 µg/mL streptomycin (Life Technologies), and 10% FBS (HyClone). Cells were cultured in a humidified tissue culture incubator at 37° C with 5% CO_2. The T-Rex™ System (Life Technologies) was used to create tetracycline/doxycycline-inducible cell lines. The pcDNA6/TR© vector was stably transfected into MDA-MB-468 cells, and a clone expressing the *tetR* gene was chosen for subsequent experiments. Human *NR2F1* cDNA (OriGene; SKU: SC116616) was removed from the pCMV6-XL4 expression vector via *NotI* digestion and subcloned into the pcDNA™4/TO inducible expression vector. The resulting *NR2F1* expression vector was transfected into MDA-MB-468 cells expressing the *tetR* gene, and stable clones obtained following zeocin selection (500µg/mL). All transfections were performed via electroporation using a Nucleofector® II Device and Amaxa® Cell Line Nucleofector® Kit V (Lonza),

according to manufacturer's instructions. For *in vitro* experiments where induction of *NR2F1* was desired, media was changed to one containing 10% Tet System Approved FBS (Clontech).

Western blot analysis

Cells were lysed in RIPA buffer (50mM Tris-HCL, pH 8, 150mM NaCl, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS, 1mM EDTA) containing protease inhibitor cocktail tablets (Roche), agitated for 20 minutes at 4°C, and centrifuged at 14,000 rpm for 10 minutes at 4°C. The supernatant was used for analysis. Protein was fractionated by 10% SDS-PAGE electrophoresis and transferred to a nitrocellulose membrane. Membranes were blocking in a 5% (w/v) non-fat powdered milk solution for 1 hour and incubated overnight at 4°C with primary antibody. Membranes were then washed three times in PBST (phosphate buffered saline with 0.05% Tween-20) and incubated with a horseradish peroxidase conjugated secondary antibody for 1 hour at room temperature. Membranes were washed three times in PBST and complexes detected using ECL Western Blotting Substrate (Thermo Scientific). Bands were visualized using a ChemiDoc™ MP system (Bio-Rad). The following antibodies and dilutions were used: TetR at 1:1,000 (Clontech, Cat #: 631131); human NR2F1 at 1:1,000 (R&D Systems; Cat #: PP-H8132-00); human GAPDH at 1:10,000 (R&D Systems; Cat #: MAB5718); goat anti-mouse IgG HRP-conjugated secondary antibody at 1:2,000 (Life Technologies; Cat #: A16072).

Mouse xenograft experiments

All animals were maintained in an AAALAC-accredited facility and protocols were approved by the University of Wisconsin – Madison School of Medicine and Public Health Animal Care and Use Committee. Female athymic nude (*Foxn1*^{nu}) mice aged 3-4 weeks were purchased from Harlan Laboratories. To establish xenograft carcinomas, 2 million tetracycline-inducible MDA-MB-468 cells were suspended in 100µL PBS and subcutaneously injected into the right flank of 6 week old nude mice.

Twenty-four hours following injection, half of the mice were put on a rodent diet containing 625 mg/kg doxycycline to induce *NR2F1* expression (Harlan; TD.01306). Carcinoma growth was measured with a caliper over the course of 45 days, and carcinoma volume was determined using the formula ½(L*W*H). To calculate the carcinoma growth rate (mm^3/day) over the course of the experiment, the 45 day measurement period was divided into 3 approximately 15-day growth windows consisting of the first (0-14 days post-injection, dpi), second (15-27 dpi), and the third (28-45 dpi) growth periods of the study. Growth rates between measurement points were calculated for each individual mouse and measurement points within a growth period were averaged for a single, window-specific growth rate/mouse. Individual animal growth rates were then averaged within treatment groups for each growth period of the study.

Quantitative Real-Time PCR

Cells used for gene expression were collected via 0.25% trypsin/EDTA (Life Technologies) and RNA was extracted using the RNeasy Mini Kit (Qiagen). RNA from MDA-MB-231 and BT-549 cell lines used to assess endogenous *NR2F1* expression were kindly provided by Dr. Wei Xu. For xenograft gene expression, carcinomas were snap-frozen in liquid nitrogen at necropsy and homogenized in TRI-Reagent (Life Technologies) using a Polytron® PT 10-35 homogenizer (Kinematica AG). RNA was extracted using MagMAX-96 for Microarrays Total RNA kit (Life Technologies). cDNA was prepared from 2µg total RNA using Superscript II reverse transcriptase (Life Technologies). Gene expression was quantified using pre-designed TaqMan qPCR assays and run on an ABI Prism 7900HT (Life Technologies). cDNA was diluted 1:8 and run using reaction conditions described previously [61]. Transcript quantities were calculated as described in Smits *et al* [15], using a standard curve method to calculate C_t values and extrapolate quantity values. Human *PPIA* (Hs04194521_s1) served as the reference gene, and TaqMan assays used for target genes are listed in S1 Table.

RNA-sequencing and functional enrichment analysis

A total of 8 xenograft carcinomas were chosen for RNA-sequencing (RNA-seq) analysis, 4 from mice receiving doxycycline-containing food (Dox) to induce *NR2F1* expression and 4 on a normal diet (no Dox). RNA was sent to the University of Wisconsin Biotechnology Center Gene Expression Center for sample preparation and sequencing. Sample preparation was done using the TruSeq RNA Library Preparation Kit (Illumina), according to the manufacturer's specifications and sequencing was performed on the Illumina HiSeq 2000. Reads were aligned to the human reference genome, GRCh37/hg19, using the RSEM algorithm [62]. Genes with at least 1 read per million in at least 3 samples in each condition were used in downstream analysis. Differential expression between Dox and no Dox carcinomas was determined using EdgeR [63]. Functional annotation enrichment analysis was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) web-based analysis tool [30,31]. A background list of expressed genes obtained from our RNA-seq results was used for used for the analysis and is available in S2 Table (electronic version only). The full list of up and down regulated genes used for enrichment analyses is available in S3 Table. Sample values in this dataset are normalized counts per million (CPM) values obtained from EdgeR. For functional annotation clustering performed in DAVID, a cluster was considered significant if its group enrichment score was >1.3. This represents the geometric mean (-log scale) of cluster-associated annotation term p-values that is < 0.05.

Creation of an NR2F1-based regulatory network

Background network

To create a regulatory network which would model the possible interactions responsible for NR2F1-mediated expression changes, an integer linear program (IP)-based approach based on the work by Chasman *et al* [64] was used. In this way, data from our RNA-seq experiment as well as publically

available protein-protein, protein-DNA, and literature-mined interactions were incorporated and used to perform subnetwork inference. Protein-protein interactions were gathered from HPRD [65], BioGRID [66], and HIPPIE [67], while the majority of protein-DNA interactions were gathered from MCF-7 ChIP-chip data collected by Kittler *et al* [34]. Additional protein-DNA interactions relevant to our network were identified using the ENCODE ChIP-Seq Significance Tool (encodeqt.simple-encode.org/) [68]. Program parameters consisted of our DE gene list (S3 Table) and an analysis window centered on the TSS/5'end using a 5000bp upstream and downstream pad in all available cell lines. Finally, the Literome Project [69,70] was used to mine biomedical articles for relevant regulatory interactions pertaining to our DE gene list. The network was visualized using the Cytoscape software platform [33]. In this way, DE genes, gene regulatory proteins (GRPs), and intermediary factors (IFs) were represented by nodes, while interactions were represented by edges. Specifically, directed edges represented regulatory protein-DNA or literature-mined interactions, while undirected edges represented physical protein-protein interactions. Post processing of the network was done to verify that undirected edges were true and relevant to the system.

IP method, weighting, and path generation

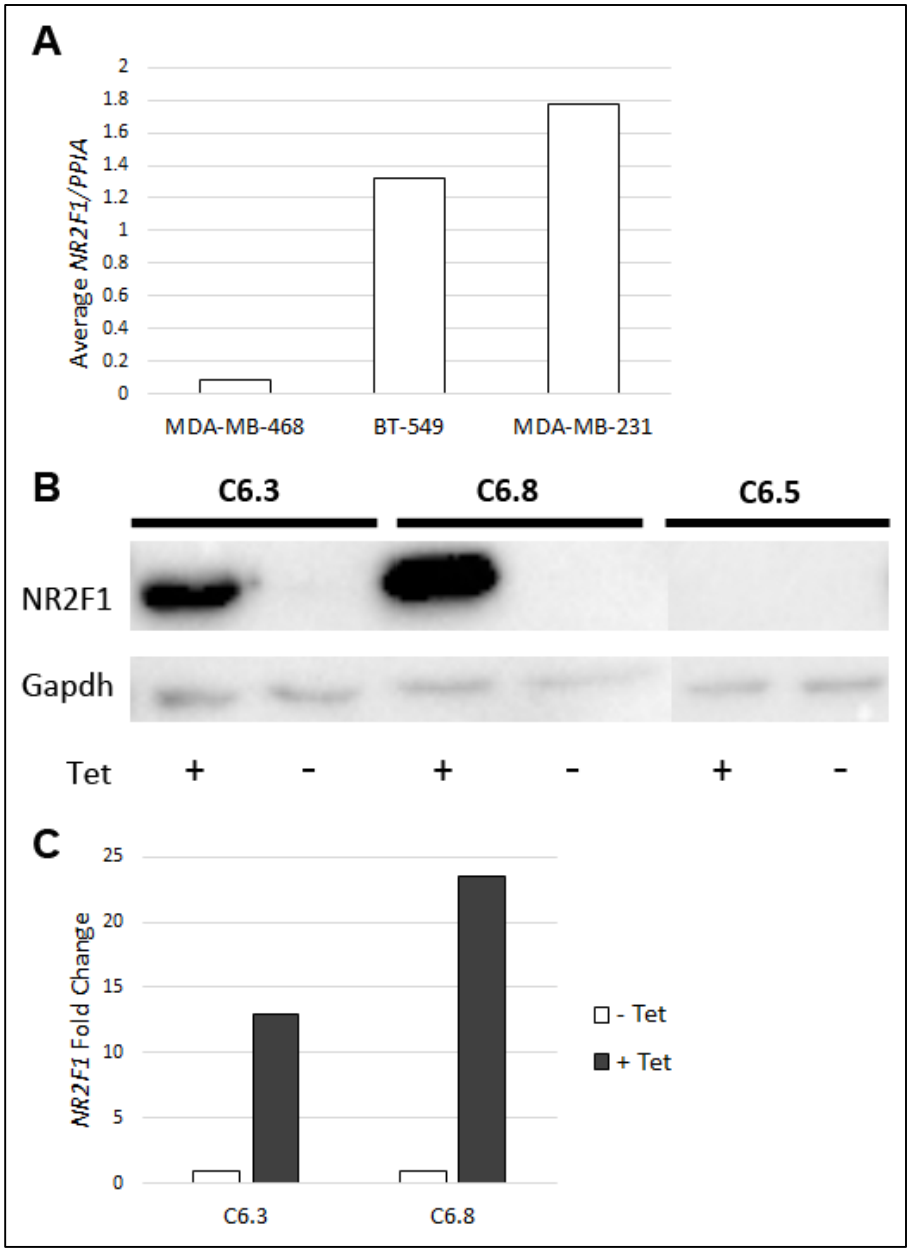
The NR2F1 signaling subnetwork was inferred by solving two integer linear programs (IPs) in a method based on the work by Chasman *et al* [64]. A full description of this method can be found in the Supplemental Methods section. In general, using NR2F1 as the source node, a depth-first traversal search algorithm was used to find all directed acyclic pathways that led from NR2F1 to the DE genes. The stopping criteria for the algorithm were as follows: if the maximum depth was exceeded, stop searching down that branch of the network; if a protein-DE gene edge was reached, do not search further down that path. The maximum depth of the search algorithm was limited to 3, so all paths had at most 4 nodes. All paths had the same directionality, going from a source node to a target node and

not the reverse. The subnetwork inference IP had the potential to generate multiple solutions that satisfied all of the constraints and retrieved the same objective function solution values. To differentiate the solutions and select the one that inferred the most plausible subnetwork, RNA-seq expression data and the scientific literature were used to assign weights to gene nodes. To incorporate RNA-seq data, expression values, represented as normalized counts per million (CPM) values obtained from EdgeR, were used to weight GRPs and IFs for selection, with the requirement that these incorporated nodes be expressed in MDA-MB-468 cells. Literature-based weighting of GRPs and IFs was performed using GADGET (http://gadget.biostat.wisc.edu). GADGET is a tool developed by Dr. Mark Craven and colleagues that allowed for the mining of biomedical literature for genes associated with a user-input query. Using GADGET, the query "Breast Cancer AND Basal NOT Luminal" was used to identify associated genes. These genes were sorted based on their relevance to the query using a GADGET-generated F_1 score. This F_1 score was used to weight GRPs and IFs in the network. The integration of both RNA-seq and literature-based weighting methods assisted in choosing an IP solution that was the most relevant to our TNBC cell model.

Acknowledgements

The authors would like to thank the University of Wisconsin-Madison Biotechnology Center Gene Expression Center for their sequencing services and the Center for Predictive Computational Phenotyping for its support of the GADGET interface. Also, thank you to Dr. Wei Xu for TNBC cell line RNA and discussion of the T-Rex™ System and to Rachael Baird for technical assistance.

Chapter 3, Fig 1: Inducible overexpression of *NR2F1* in TNBC cell line



Chapter 3, Fig 2: NR2F1 reduces TNBC growth in a xenograft model

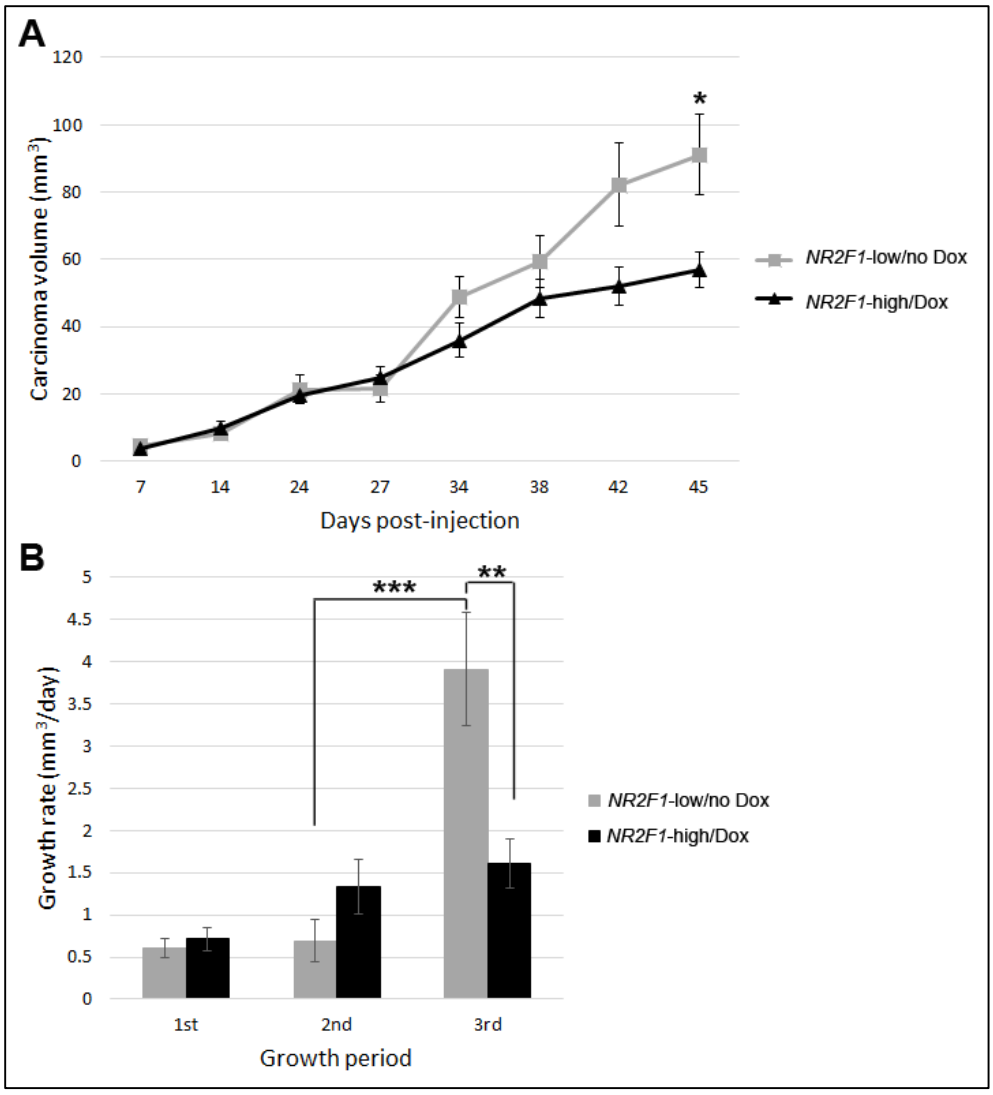


Table 1: Top 40 significantly up and downregulated genes in *NR2F1*-high vs. *NR2F1*-low xenograft carcinomas

Upregulated genes

Gene Symbol	Fold change	adj p-val
CHI3L1	14.18	4.41E-19
FCGBP	13.60	1.21E-16
FOXA1	11.15	3.02E-16
SERPINA5	29.35	8.30E-14
PKP1	12.41	2.07E-13
PTGIS	9.46	7.52E-13
PLAT	19.74	2.91E-12
NPNT	11.07	2.96E-12
TCN1	18.14	7.85E-12
ALOX5	10.10	2.53E-11
PADI3	18.80	2.79E-11
VSTM2L	10.44	7.02E-10
ANO1	4.50	3.63E-09
TGM2	9.41	5.05E-09
TMEM213	5.84	1.47E-08
PECAM1	15.62	1.49E-08
LOC100505989	60.37	1.56E-08
TUB	12.65	2.19E-08
INPP4B	12.46	2.97E-08
MSX1	9.49	7.36E-08
HOXB2	10.08	8.44E-08
RIC3	146.37	1.23E-07
UGT1A6	28.26	2.22E-07
IL1R2	8.52	2.22E-07
UGT1A7	169.33	2.58E-07
CST6	6.89	3.24E-07
GNE	4.48	3.63E-07
CLU	6.52	5.68E-07
SPRR3	45.51	5.94E-07
INHBB	4.02	5.94E-07
IL1B	26.81	7.20E-07
PPBP	11.29	8.68E-07
SUSD2	5.26	9.18E-07
MUCL1	17.42	1.43E-06
SCIN	10.03	1.45E-06
CAPN8	5.31	1.64E-06
NR2F1	4.28	1.72E-06
PPP1R14C	5.28	1.77E-06
THSD7A	9.52	1.85E-06

Downregulated genes

Gene Symbol	Fold change	adj p-val
IFI27	-12.74	1.73E-14
MAOA	-9.98	5.63E-10
NGFR	-7.47	2.60E-08
TNFAIP2	-4.87	1.21E-06
STAC2	-4.50	1.27E-06
TMEM100	-5.13	1.95E-06
MFI2	-4.94	8.76E-06
ITGB2	-4.12	8.76E-06
WIF1	-6.14	9.72E-06
SLC38A3	-5.44	1.29E-05
FAT3	-8.08	1.39E-05
FAM176A	-3.44	1.66E-05
SPON2	-6.81	1.87E-05
LOC645638	-4.59	1.93E-05
SYT8	-5.80	1.99E-05
KIAA1210	-6.33	3.71E-05
SH3TC1	-4.29	4.47E-05
NES	-4.00	4.72E-05
TNS1	-3.45	8.16E-05
EDN2	-3.87	8.41E-05
CA12	-3.49	1.08E-04
DYX1C1-CCPG1	-8.37	1.14E-04
LINC00284	-3.81	1.18E-04
LOC100505746	-3.85	1.24E-04
LRRC55	-3.32	1.30E-04
ITGAL	-4.51	1.33E-04
UNC5B	-3.96	1.33E-04
NT5E	-4.11	1.39E-04
CD22	-9.82	1.48E-04
ARSA	-3.56	1.48E-04
PRSS21	-9.55	1.66E-04
ATP6V0D2	-4.31	2.34E-04
RNF150	-4.71	2.46E-04
ALDH3B1	-3.60	2.57E-04
ADRB2	-4.25	5.23E-04
LOC100499194	-8.93	5.33E-04
LYNX1	-3.57	5.55E-04
LOC100507218	-8.24	5.83E-04
ATP6V1C2	-2.90	6.29E-04

Genes are listed by adjusted p-value

Chapter 3, Fig 3: qPCR validation of differentially expressed genes

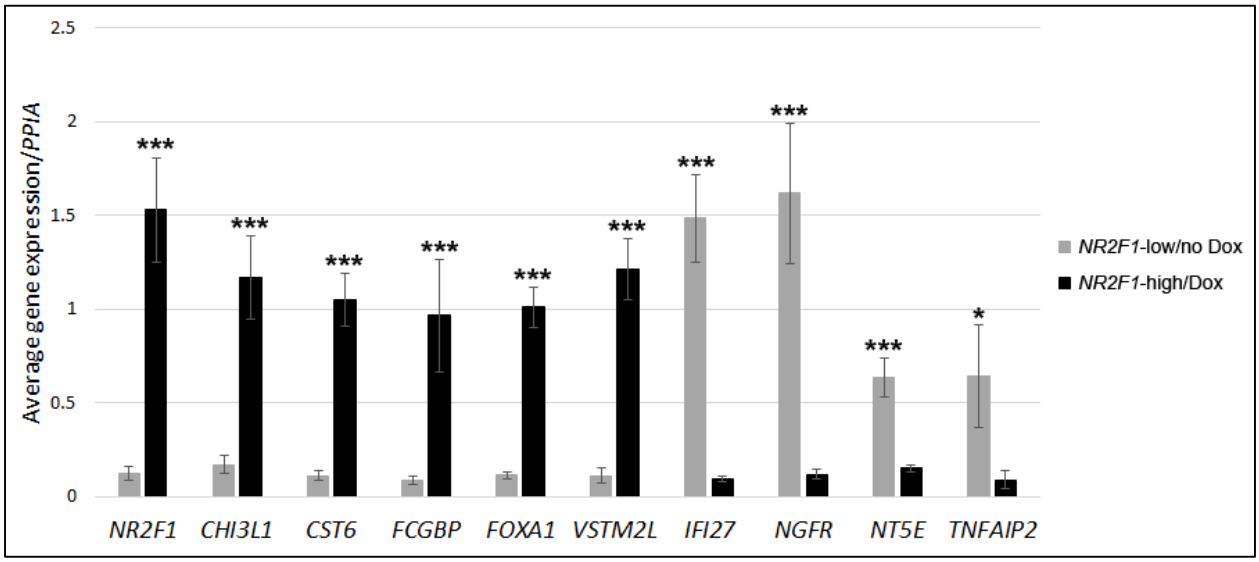


Table 2: Select functional annotation clusters of upregulated gene set			
Annotation Cluster 1	**Enrichment Score: 17.17**		
Category	Term	PValue	Benjamini
SP_PIR_KEYWORDS	glycoprotein	5.48E-24	1.75E-21
SP_PIR_KEYWORDS	disulfide bond	3.28E-18	5.26E-16
UP_SEQ_FEATURE	signal peptide	1.76E-17	7.04E-15
SP_PIR_KEYWORDS	signal	2.20E-17	2.35E-15
UP_SEQ_FEATURE	disulfide bond	6.83E-17	2.95E-14
GOTERM_CC_FAT	GO:0005576~extracellular region	1.82E-16	4.42E-14
SP_PIR_KEYWORDS	Secreted	7.21E-15	5.77E-13
Annotation Cluster 2	**Enrichment Score: 11.15**		
Category	Term	PValue	Benjamini
GOTERM_CC_FAT	GO:0005576~extracellular region	1.82E-16	4.42E-14
GOTERM_CC_FAT	GO:0044421~extracellular region part	6.17E-11	6.14E-09
GOTERM_CC_FAT	GO:0005615~extracellular space	3.19E-08	2.11E-06
Annotation Cluster 3	**Enrichment Score: 8.05**		
Category	Term	PValue	Benjamini
SP_PIR_KEYWORDS	glycoprotein	5.48E-24	1.75E-21
UP_SEQ_FEATURE	glycosylation site:N-linked (GlcNAc...)	2.21E-21	1.76E-18
UP_SEQ_FEATURE	topological domain:Extracellular	3.23E-10	6.44E-08
UP_SEQ_FEATURE	topological domain:Cytoplasmic	6.82E-09	1.09E-06
GOTERM_CC_FAT	GO:0031226~intrinsic to plasma membrane	8.78E-07	4.37E-05
GOTERM_CC_FAT	GO:0005886~plasma membrane	1.63E-06	6.48E-05
GOTERM_CC_FAT	GO:0005887~integral to plasma membrane	4.23E-06	1.40E-04
UP_SEQ_FEATURE	transmembrane region	4.49E-06	5.96E-04
SP_PIR_KEYWORDS	transmembrane	6.67E-06	4.27E-04
SP_PIR_KEYWORDS	membrane	9.31E-06	4.96E-04
GOTERM_CC_FAT	GO:0044459~plasma membrane part	1.31E-05	3.72E-04
GOTERM_CC_FAT	GO:0016021~integral to membrane	0.0017	0.0367
GOTERM_CC_FAT	GO:0031224~intrinsic to membrane	0.0020	0.0391
Annotation Cluster 4	**Enrichment Score: 3.00**		
Category	Term	PValue	Benjamini
GOTERM_BP_FAT	GO:0008544~epidermis development	4.34E-05	0.0183
GOTERM_BP_FAT	GO:0007398~ectoderm development	9.29E-05	0.0311
GOTERM_BP_FAT	GO:0030855~epithelial cell differentiation	4.54E-04	0.0742
GOTERM_BP_FAT	GO:0060429~epithelium development	0.0014	0.1272
GOTERM_BP_FAT	GO:0030216~keratinocyte differentiation	0.0162	0.3897

GOTERM_BP_FAT	GO:0009913~epidermal cell differentiation	0.0252	0.4524
Annotation Cluster 11	**Enrichment Score: 1.82**		
Category	Term	PValue	Benjamini
GOTERM_BP_FAT	GO:0007155~cell adhesion	0.0132	0.3574
GOTERM_BP_FAT	GO:0022610~biological adhesion	0.0134	0.3565
SP_PIR_KEYWORDS	cell adhesion	0.0194	0.2300
Annotation Cluster 15	**Enrichment Score: 1.52**		
Category	Term	PValue	Benjamini
GOTERM_BP_FAT	GO:0042127~regulation of cell proliferation	2.48E-05	0.0208
GOTERM_BP_FAT	GO:0006916~anti-apoptosis	0.0080	0.3092
GOTERM_BP_FAT	GO:0043066~negative regulation of apoptosis	0.0082	0.3005
GOTERM_BP_FAT	GO:0043069~negative regulation of programmed cell death	0.0088	0.3002
GOTERM_BP_FAT	GO:0060548~negative regulation of cell death	0.0090	0.3004
GOTERM_BP_FAT	GO:0042981~regulation of apoptosis	0.0156	0.3852
GOTERM_BP_FAT	GO:0043067~regulation of programmed cell death	0.0170	0.3945
GOTERM_BP_FAT	GO:0010941~regulation of cell death	0.0177	0.3968
GOTERM_BP_FAT	GO:0043068~positive regulation of programmed cell death	0.0246	0.4487
GOTERM_BP_FAT	GO:0010942~positive regulation of cell death	0.0255	0.4467
GOTERM_BP_FAT	GO:0043065~positive regulation of apoptosis	0.0498	0.5794

Table 3: Select functional annotation clusters of downregulated gene set			
Annotation Cluster 1	**Enrichment Score: 5.46**		
Category	Term	PValue	Benjamini
SP_PIR_KEYWORDS	disulfide bond	9.26E-09	2.02E-06
UP_SEQ_FEATURE	disulfide bond	9.89E-09	5.37E-06
SP_PIR_KEYWORDS	signal	1.45E-07	1.59E-05
UP_SEQ_FEATURE	signal peptide	1.52E-07	4.14E-05
SP_PIR_KEYWORDS	glycoprotein	7.80E-07	5.67E-05
UP_SEQ_FEATURE	glycosylation site:N-linked (GlcNAc...)	2.52E-05	0.0045
SP_PIR_KEYWORDS	Secreted	3.24E-05	0.0018
GOTERM_CC_FAT	GO:0005576~extracellular region	7.87E-05	0.0063
GOTERM_CC_FAT	GO:0044421~extracellular region part	9.83E-04	0.0312
GOTERM_CC_FAT	GO:0005615~extracellular space	0.0023	0.0587
Annotation Cluster 2	**Enrichment Score: 3.93**		
Category	Term	PValue	Benjamini
GOTERM_BP_FAT	GO:0007155~cell adhesion	2.70E-05	0.0301
GOTERM_BP_FAT	GO:0022610~biological adhesion	2.76E-05	0.0155
SP_PIR_KEYWORDS	cell adhesion	3.86E-04	0.0105
GOTERM_BP_FAT	GO:0016337~cell-cell adhesion	6.86E-04	0.2282
Annotation Cluster 3	**Enrichment Score: 3.63**		
Category	Term	PValue	Benjamini
GOTERM_CC_FAT	GO:0005886~plasma membrane	7.24E-06	0.0012
SP_PIR_KEYWORDS	cell membrane	8.06E-05	0.0029
SP_PIR_KEYWORDS	membrane	0.0221	0.2770
Annotation Cluster 17*	**Enrichment Score: 1.20**		
Category	Term	PValue	Benjamini
GOTERM_CC_FAT	GO:0033176~proton-transporting V-type ATPase complex	8.08E-04	0.0320
GOTERM_CC_FAT	GO:0016471~vacuolar proton-transporting V-type ATPase complex	0.0033	0.0738
SP_PIR_KEYWORDS	Hydrogen ion transport	0.0052	0.0911
GOTERM_BP_FAT	GO:0015986~ATP synthesis coupled proton transport	0.0055	0.6472
GOTERM_BP_FAT	GO:0015985~energy coupled proton transport, down electrochemical gradient	0.0055	0.6472

*Cluster 17 contains 56 terms and is shown in a condensed form

Chapter 3, Fig 4: Multi-platform network mapping of regulatory interactions between NR2F1 and DE genes

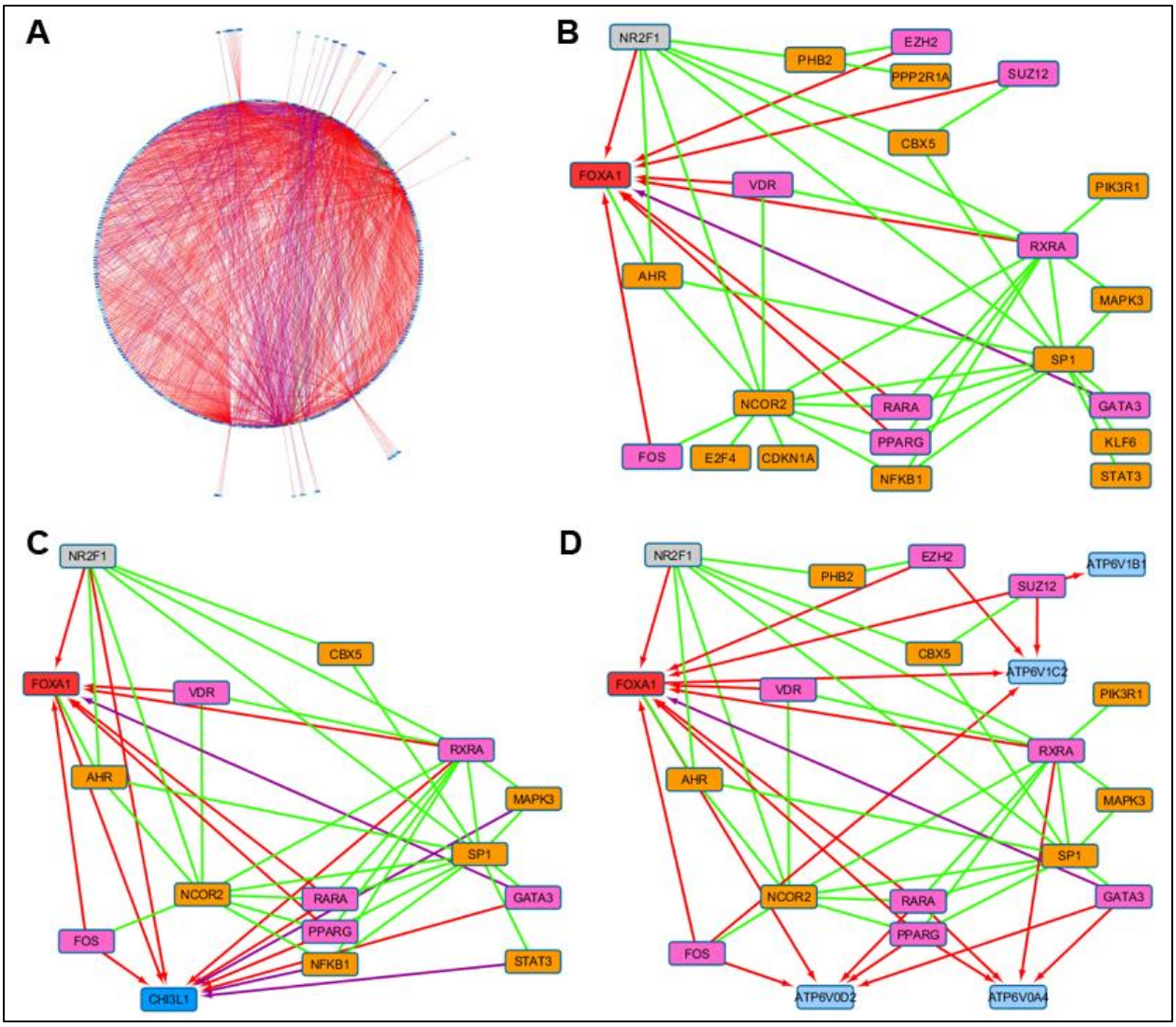


Table 4: Gene regulatory protein interactions within the NR2F1 network

GRP	Direct DE gene targets	Percent of DE network genes
EZH2	131	39.7%
FOS	143	43.3%
FOXA1	155	47.0%
GATA3	83	25.2%
NR2F1	94	28.5%
PPARG	81	24.5%
RARA	195	59.1%
RXRA	77	23.3%
SUZ12	42	12.7%
VDR	93	28.2%

GRP = gene regulatory protein; DE = differentially expressed

Supplemental Figure and Table Legends

S1 Fig: Doxycycline does not affect TNBC growth. (A) 2 million cells containing only the constitutively active *tet*R plasmid (TR1 cells) were injected into the hind flank of 6 week old female nude mice. 24 hours post-injection, animals were placed on a dox-containing (TR1/Dox; n=13) or standard (TR1/no Dox; n=12) rodent diet. Carcinoma volume (mm^3, y-axis) was measured with a caliper using the formula ½(L*W*H) over the course of 45 days (x-axis). The difference in carcinoma volumes was not statistically significant between the two groups. Standard error bars are displayed, and p-values were obtained using the non-parametric Mann Whitney U test (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$).

S1 Table: Expression assays used for qPCR

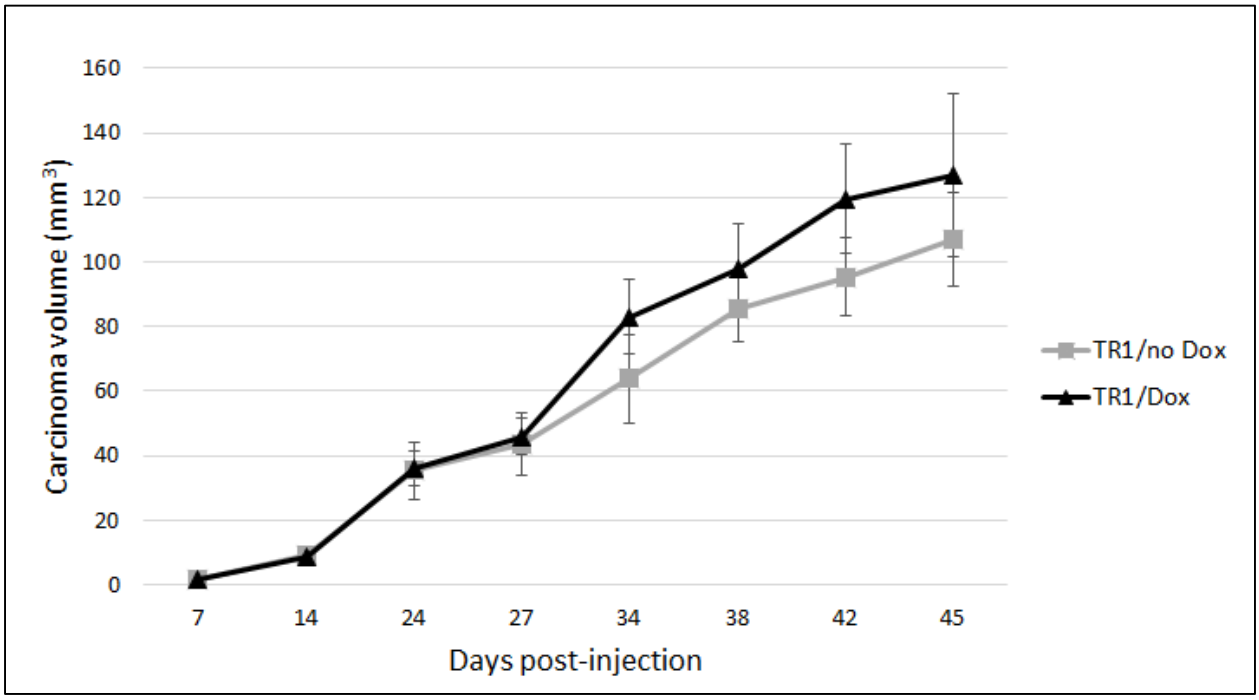
S2 Table: Background gene list used for functional annotation analysis. Gene symbols were converted to Entrez gene IDs in DAVID and used as the background gene list for functional annotation analyses on up and downregulated genes. *Due to its size, this table is only available in the electronic version of this thesis.*

S3 Table: Complete list of statistically significant up and downregulated genes from RNA-seq analysis. An adjusted p-value cut-off of <0.05 was considered statistically significant. Treatment and control sample values are normalized counts per million (CPM) values obtained from EdgeR analysis.

S4 Table: Functional annotation clustering of upregulated gene set using DAVID - top 25 clusters. Significant clusters (enrichment score >1.3) are highlighted in red. The gene list from S2 Table was used as the background list. *Due to its size, this table is only available in the electronic version of this thesis.*

S5 Table: Functional annotation clustering of downregulated gene set using DAVID - top 25 clusters. Significant clusters (enrichment score >1.3) are highlighted in red. The gene list from S2 Table was used as the background list. *Due to its size, this table is only available in the electronic version of this thesis.*

Chapter 3, S1 Fig: Doxycycline does not affect TNBC growth



Supplemental Table 1: Expression assays used for qPCR	
Gene	**Assay ID**
CHI3L1	Hs00609691_m1
CST6	Hs01012810_g1
FCGBP	Hs00175398_m1
FOXA1	Hs04187555_m1
IFI27	Hs01086373_g1
NGFR	Hs00609977_m1
NR2F1	Hs00818842_m1
NT5E	Hs00159686_m1
TNFAIP2	Hs00196800_m1
VSTM2L	Hs00327067_m1

S2 Table: Background gene list used for functional annotation analysis.

This table contains a list of the 15,948 expressed genes (gene symbols and Entrez IDs) used as the background list for DAVID analysis. Due to its size, it is only available in the electronic version of this thesis.

Supplemental Table 3: Complete list of statistically significant up and downregulated genes from RNA-seq analysis

Genes upregulated in *NR2F1*-high samples

				Treatment samples: *NR2F1*-high				Control samples: *NR2F1*-low			
gene_id	fold Change	pvalue	padj	MH01	MH02	MH03	MH04	MH05	MH06	MH07	MH08
CHI3L1	14.18	2.76E-23	4.41E-19	78.42	70.43	60.03	87.98	4.18	6.56	5.40	4.23
FCGBP	13.60	1.52E-20	1.21E-16	292.57	316.47	648.37	488.05	22.11	40.98	31.10	31.88
FOXA1	11.15	5.68E-20	3.02E-16	42.38	45.12	45.82	37.28	3.70	3.86	4.71	2.57
SERPINA5	29.35	2.60E-17	8.30E-14	6.41	5.72	8.33	6.05	0.24	0.29	0.17	0.15
PKP1	12.41	7.78E-17	2.07E-13	53.11	33.68	29.73	51.36	3.14	4.44	3.17	2.42
PTGIS	9.46	3.30E-16	7.52E-13	87.38	45.42	94.34	83.38	8.04	8.00	7.88	8.47
PLAT	19.74	1.46E-15	2.91E-12	38.13	40.40	26.46	47.56	3.30	1.93	0.60	1.21
NPNT	11.07	1.67E-15	2.96E-12	34.04	84.66	70.96	54.93	3.14	2.51	10.11	6.34
TCN1	18.14	4.93E-15	7.85E-12	7.87	11.80	18.54	11.96	0.72	0.58	1.20	0.23
ALOX5	10.10	1.75E-14	2.53E-11	25.70	19.16	18.30	18.31	1.53	2.41	1.46	2.34
PADI3	18.80	2.10E-14	2.79E-11	10.81	8.22	5.64	12.26	0.64	0.58	0.26	0.38
VSTM2L	10.44	6.17E-13	7.02E-10	18.91	13.66	22.30	25.24	1.77	2.41	1.20	2.04
ANO1	4.50	3.41E-12	3.63E-09	343.04	343.57	250.35	293.56	43.42	69.13	90.40	66.17
TGM2	9.41	5.06E-12	5.05E-09	20.84	18.59	21.73	28.23	2.66	2.80	1.11	2.42
TMEM213	5.84	1.57E-11	1.47E-08	28.67	38.29	36.76	26.34	5.16	3.69	8.91	4.15
PECAM1	15.62	1.68E-11	1.49E-08	2.32	3.00	2.53	2.12	0.16	0.00	0.17	0.23
LOC100505989	60.37	1.86E-11	1.56E-08	0.85	2.50	2.70	1.57	0.00	0.00	0.09	0.00
TUB	12.65	2.75E-11	2.19E-08	17.29	13.73	6.45	4.74	0.48	0.48	1.46	0.83
INPP4B	12.46	4.10E-11	2.97E-08	5.02	16.95	16.74	10.51	0.48	0.48	2.23	0.76
MSX1	9.49	1.06E-10	7.36E-08	6.41	6.08	6.86	6.49	0.72	0.58	0.26	0.98
HOXB2	10.08	1.27E-10	8.44E-08	6.33	5.22	8.58	6.13	0.16	0.58	0.86	0.98
RIC3	146.37	1.92E-10	1.23E-07	0.85	3.29	1.14	0.88	0.00	0.00	0.00	0.00
UGT1A6	28.26	3.61E-10	2.22E-07	3.12	4.30	0.82	2.94	0.16	0.09	0.00	0.08
IL1R2	8.52	3.76E-10	2.22E-07	6.10	5.65	5.55	6.78	0.72	0.77	0.34	0.83
UGT1A7	169.33	4.52E-10	2.58E-07	1.23	0.56	2.28	2.39	0.00	0.00	0.00	0.00
CST6	6.89	5.89E-10	3.24E-07	37.75	28.17	40.27	52.89	8.04	5.78	3.34	4.76
GNE	4.48	6.84E-10	3.63E-07	99.50	106.40	87.23	55.56	11.90	13.98	33.67	18.28
CLU	6.52	1.10E-09	5.68E-07	191.90	130.65	351.96	290.98	38.60	46.76	20.39	36.48
SPRR3	45.51	1.23E-09	5.94E-07	0.93	1.00	2.12	1.53	0.00	0.00	0.00	0.08
INHBB	4.02	1.22E-09	5.94E-07	164.88	112.26	116.15	107.82	31.84	30.27	32.73	25.91
IL1B	26.81	1.53E-09	7.20E-07	2.70	6.15	1.06	5.40	0.00	0.19	0.00	0.30
PPBP	11.29	1.91E-09	8.68E-07	2.62	4.36	3.10	2.63	0.00	0.10	0.69	0.30
SUSD2	5.26	2.07E-09	9.18E-07	420.62	262.04	501.59	511.21	63.52	99.69	50.47	99.33

MUCL1	17.42	3.50E-09	1.43E-06	1.00	2.57	4.25	3.36	0.16	0.19	0.17	0.08
SCIN	10.03	3.65E-09	1.45E-06	1.86	3.08	2.21	1.90	0.16	0.10	0.26	0.30
CAPN8	5.31	4.23E-09	1.64E-06	22.54	24.53	10.95	14.74	2.65	3.57	3.17	3.55
NR2F1	4.28	4.54E-09	1.72E-06	343.52	194.41	229.66	278.31	41.81	74.60	54.92	68.43
PPP1R14C	5.28	4.77E-09	1.77E-06	40.99	37.54	18.46	27.43	5.63	4.63	9.51	3.40
THSD7A	9.52	5.12E-09	1.85E-06	2.86	9.94	6.86	4.73	0.82	0.29	0.34	0.83
TMPRSS4	4.53	5.63E-09	1.95E-06	54.50	45.26	42.39	60.62	9.01	14.17	9.25	10.73
BPI	26.10	6.25E-09	2.12E-06	1.93	1.07	3.27	1.39	0.08	0.10	0.09	0.00
KCNK3	12.43	7.33E-09	2.44E-06	2.86	2.86	1.80	4.81	0.32	0.29	0.09	0.23
PLA2G2A	14.44	1.11E-08	3.62E-06	2.70	2.29	8.00	5.03	0.24	0.58	0.34	0.08
GLP1R	7.57	1.23E-08	3.91E-06	10.73	11.80	5.80	7.30	0.40	1.74	1.89	0.60
SCGB3A1	16.90	2.40E-08	7.51E-06	2.32	1.29	5.64	3.50	0.08	0.00	0.26	0.38
MYH16	19.62	2.62E-08	8.04E-06	3.24	4.79	0.57	2.48	0.00	0.10	0.09	0.30
PLAU	4.57	3.11E-08	9.01E-06	97.84	76.89	38.79	85.48	14.11	21.30	12.75	14.02
GPR115	9.76	3.89E-08	1.09E-05	6.25	8.94	2.70	3.28	0.08	0.39	1.29	0.38
F3	4.42	5.62E-08	1.49E-05	130.45	271.43	84.13	134.09	30.72	15.62	62.46	27.42
LBP	8.99	6.57E-08	1.69E-05	12.43	13.23	21.56	26.41	1.45	4.44	1.20	0.76
SPRR1B	88.17	8.66E-08	2.09E-05	1.08	0.79	0.33	1.24	0.00	0.00	0.00	0.00
A2ML1	14.33	1.08E-07	2.58E-05	8.88	6.65	1.72	6.57	0.00	0.96	0.34	0.30
GABRP	3.29	1.16E-07	2.71E-05	105.21	210.93	145.06	79.15	31.52	35.00	48.15	40.64
GRIK2	5.77	1.66E-07	3.72E-05	2.86	8.51	5.06	4.30	0.48	0.58	1.20	1.13
COL17A1	5.29	1.65E-07	3.72E-05	25.47	29.82	9.72	25.75	5.87	3.09	3.34	3.63
RARB	6.16	1.70E-07	3.76E-05	2.86	8.08	7.68	5.54	0.80	0.39	1.63	0.98
KCP	9.08	2.02E-07	4.42E-05	11.50	11.15	1.72	3.79	0.40	0.67	1.03	0.83
PTGER2	9.44	3.13E-07	6.57E-05	2.16	1.93	1.72	1.24	0.24	0.29	0.09	0.08
ID1	4.45	3.28E-07	6.79E-05	9.80	7.58	8.41	8.61	1.45	2.31	1.97	1.81
C15orf48	6.19	3.92E-07	8.02E-05	5.17	6.15	9.56	5.18	0.96	1.74	0.60	0.68
KIAA1755	6.46	4.64E-07	9.14E-05	2.62	2.22	2.53	3.94	0.48	0.48	0.34	0.38
EPHB2	3.68	5.88E-07	1.13E-04	78.66	51.84	59.95	79.83	19.37	19.97	12.51	18.51
IGFBP1	16.62	6.26E-07	1.17E-04	1.47	0.93	1.06	1.46	0.24	0.00	0.00	0.00
MMP28	12.93	7.67E-07	1.34E-04	2.01	1.57	0.74	2.92	0.08	0.29	0.00	0.15
HRH1	4.16	7.92E-07	1.37E-04	16.13	19.02	8.90	11.75	3.14	2.70	5.14	2.04
SORL1	3.34	8.15E-07	1.39E-04	32.19	41.33	44.35	34.51	8.52	7.13	17.22	12.61
CREB3L2	2.81	1.02E-06	1.66E-04	227.95	293.09	243.73	212.00	75.99	53.70	127.01	84.98
RAET1E	5.54	1.21E-06	1.95E-04	4.55	4.93	2.12	4.30	0.88	0.29	0.69	0.83
LHFP	3.88	1.27E-06	2.03E-04	8.49	12.08	7.43	8.97	1.45	1.83	3.20	2.79
C4orf26	15.90	1.32E-06	2.09E-04	0.46	2.22	0.90	1.17	0.00	0.00	0.26	0.00
SLCO2A1	4.00	1.52E-06	2.35E-04	19.92	20.81	17.15	26.77	5.23	5.50	3.17	6.19
DAPK1	3.13	1.59E-06	2.44E-04	63.07	65.35	35.53	39.25	12.38	13.40	22.53	14.96
GJB2	3.21	1.87E-06	2.78E-04	130.92	152.59	169.48	124.53	63.12	27.57	47.73	32.78
TLR7	6.39	2.07E-06	3.05E-04	1.54	4.58	4.33	1.90	0.32	0.19	0.60	0.68

HSD17B11	8.12	2.14E-06	3.13E-04	0.85	1.50	1.23	1.39	0.13	0.10	0.09	0.23
KCNE1	4.87	2.36E-06	3.43E-04	3.32	2.65	3.35	2.99	0.40	0.39	1.11	0.60
CTSL1	3.45	2.82E-06	4.06E-04	19.30	14.89	17.72	14.66	4.66	3.76	4.97	5.36
FIBCD1	5.86	3.21E-06	4.53E-04	1.62	1.83	2.04	2.19	0.08	0.39	0.34	0.45
FOXQ1	2.88	3.19E-06	4.53E-04	105.06	83.73	83.64	84.99	34.26	29.70	29.30	25.76
PODXL	2.61	3.28E-06	4.58E-04	424.63	430.59	299.11	312.57	130.67	113.57	173.12	127.88
MAML3	6.49	4.08E-06	5.55E-04	2.62	2.57	1.96	1.24	0.00	0.19	0.60	0.45
PIGR	6.58	5.11E-06	6.72E-04	8.72	11.51	47.13	10.21	2.73	3.86	1.71	2.95
LOC10013 1551	4.79	6.03E-06	7.66E-04	2.24	3.43	2.29	1.39	0.40	0.29	0.69	0.45
STK39	3.46	6.01E-06	7.66E-04	10.73	20.38	20.91	13.35	3.94	2.89	6.77	4.83
AXIN2	3.33	6.05E-06	7.66E-04	20.14	22.40	12.17	13.13	4.58	3.28	6.60	5.21
EMP1	2.74	6.21E-06	7.74E-04	404.25	329.27	241.12	214.41	92.23	82.34	162.39	91.25
IL1RL1	9.24	6.90E-06	8.53E-04	0.69	2.57	0.74	1.90	0.00	0.10	0.34	0.15
ABLIM3	6.90	7.11E-06	8.72E-04	1.62	2.79	0.82	2.41	0.24	0.39	0.17	0.23
SLC23A1	5.54	8.22E-06	1.00E-03	1.54	2.22	1.63	2.19	0.08	0.19	0.77	0.30
MATN2	3.16	8.54E-06	1.03E-03	15.59	23.10	17.81	19.04	4.50	4.53	9.77	4.91
TRIML2	11.57	8.68E-06	1.04E-03	1.00	2.29	0.49	0.88	0.16	0.10	0.09	0.00
SCARA3	2.81	9.87E-06	1.17E-03	141.26	107.68	142.04	152.40	44.31	56.98	41.21	44.79
OLFM4	4.04	1.09E-05	1.28E-03	182.56	493.94	754.14	314.50	162.76	64.31	121.24	60.28
TNFSF10	3.83	1.12E-05	1.29E-03	9.42	35.32	29.00	17.87	4.34	3.66	10.28	5.06
FZD4	3.32	1.11E-05	1.29E-03	24.24	30.32	19.68	14.74	3.62	4.44	10.88	7.55
KCNK6	3.33	1.14E-05	1.30E-03	13.20	11.15	8.41	9.70	2.65	3.09	3.34	3.25
JAG1	2.60	1.26E-05	1.38E-03	143.58	162.74	79.47	105.27	41.01	36.15	63.75	42.00
ENPP1	3.60	1.37E-05	1.47E-03	12.35	32.67	33.41	17.44	4.42	3.86	10.97	7.03
C11orf16	47.35	1.46E-05	1.55E-03	0.62	0.64	0.33	0.29	0.00	0.00	0.00	0.00
LOC10013 0705	6.84	1.52E-05	1.56E-03	16.67	17.73	1.96	3.79	0.64	1.45	2.48	1.06
WNT7A	24.43	1.95E-05	1.92E-03	0.69	1.00	0.08	1.24	0.00	0.00	0.00	0.08
F2R	3.78	1.96E-05	1.92E-03	5.56	12.30	8.90	8.17	1.77	0.96	4.28	2.12
SYTL5	12.57	2.40E-05	2.32E-03	0.31	1.07	0.57	0.73	0.00	0.10	0.09	0.00
NOSTRIN	5.82	2.64E-05	2.47E-03	1.31	6.72	4.33	2.99	0.24	0.19	1.54	0.60
KANK4	3.25	2.65E-05	2.47E-03	13.51	10.51	8.74	7.66	3.14	2.51	3.51	2.87
TNNT2	5.11	3.01E-05	2.75E-03	15.82	16.73	4.17	5.62	0.80	1.93	4.46	1.06
CABP4	4.07	3.02E-05	2.75E-03	4.48	6.72	2.53	2.41	0.72	0.67	1.63	0.76
RNASE1	4.10	3.49E-05	3.13E-03	7.49	4.93	8.17	7.95	1.61	2.60	0.94	1.51
LGALS9C	6.51	4.20E-05	3.70E-03	4.02	3.34	5.74	5.03	0.16	1.85	0.38	0.32
MGAM	4.31	4.44E-05	3.85E-03	1.70	3.65	2.61	1.46	0.32	0.29	0.94	0.53
ENTPD8	4.04	4.53E-05	3.89E-03	5.91	5.29	2.94	2.48	0.56	1.08	1.71	0.68
LAMB3	3.01	4.54E-05	3.89E-03	756.09	475.14	458.87	726.60	223.38	249.32	117.41	171.31
GPR87	6.43	4.61E-05	3.93E-03	3.86	8.37	1.39	4.89	0.56	0.00	1.54	0.68
CCND1	2.35	4.83E-05	4.08E-03	158.63	151.94	112.31	124.60	57.01	44.43	71.29	54.16

ELOVL4	3.52	5.03E-05	4.22E-03	12.97	50.98	31.61	28.96	7.08	3.57	16.19	7.88
TMEM156	4.01	5.25E-05	4.38E-03	2.01	3.00	2.70	2.19	0.48	0.67	1.03	0.23
S100B	16.59	5.35E-05	4.40E-03	0.31	0.43	0.74	0.58	0.08	0.00	0.00	0.00
SCCPDH	2.44	5.50E-05	4.49E-03	72.17	107.83	91.32	62.01	28.79	23.81	53.21	28.10
HKDC1	3.90	5.60E-05	4.56E-03	2.78	2.93	2.53	2.19	0.64	0.29	0.60	0.98
FLJ43826	9.20	5.87E-05	4.70E-03	1.16	0.43	0.82	1.02	0.08	0.10	0.09	0.08
RNASE7	6.29	5.90E-05	4.70E-03	2.01	2.00	0.49	1.75	0.40	0.10	0.17	0.23
CDC42EP5	4.90	5.92E-05	4.70E-03	3.32	1.57	4.57	4.38	0.88	0.58	0.34	0.91
EFNB1	2.97	5.84E-05	4.70E-03	61.85	38.40	59.63	71.49	24.20	18.32	13.71	18.66
HAS3	3.86	6.02E-05	4.75E-03	25.16	24.24	5.64	24.15	4.91	6.27	4.63	3.63
PRODH	3.46	6.10E-05	4.79E-03	20.69	14.59	8.00	15.03	2.65	5.98	3.86	3.78
FOLR1	3.48	6.27E-05	4.90E-03	134.23	101.03	279.04	204.51	65.41	56.72	24.42	49.85
OLFML2A	3.32	6.39E-05	4.97E-03	23.39	16.02	10.62	21.08	5.87	5.50	3.68	5.36
GPR132	19.14	6.49E-05	5.03E-03	0.85	0.72	0.57	0.29	0.00	0.10	0.00	0.00
CEACAM6	5.45	6.75E-05	5.20E-03	3.24	3.65	13.73	4.01	0.56	1.93	0.94	0.98
GJB4	4.15	6.92E-05	5.28E-03	10.65	6.94	5.55	10.80	1.69	3.47	1.03	1.59
CYP27C1	3.21	6.91E-05	5.28E-03	27.56	44.33	12.58	15.90	5.47	5.78	12.94	5.89
AGR2	2.97	7.19E-05	5.46E-03	23.85	51.98	79.72	42.75	15.04	12.82	24.68	13.45
LINC00341	3.37	8.94E-05	6.70E-03	5.48	8.44	5.06	3.50	1.85	0.87	2.40	1.21
ZIC2	2.64	8.95E-05	6.70E-03	20.69	22.38	20.75	17.22	7.16	7.62	8.74	6.08
OXTR	5.10	9.38E-05	6.99E-03	2.08	12.16	3.35	5.25	0.64	0.29	2.48	0.91
PGC	10.00	9.75E-05	7.23E-03	0.39	0.43	0.90	1.17	0.00	0.10	0.00	0.15
SYNC	2.93	9.93E-05	7.26E-03	15.47	25.29	17.72	12.19	3.72	3.38	10.21	6.52
HLA-DPA1	6.94	1.06E-04	7.72E-03	1.16	0.79	1.25	0.88	0.17	0.29	0.00	0.09
GLRA3	52.30	1.09E-04	7.84E-03	0.00	0.57	1.06	0.44	0.00	0.00	0.00	0.00
CTNNAL1	4.04	1.09E-04	7.84E-03	2.70	17.23	8.09	7.08	1.29	1.06	3.77	2.19
STAT4	4.48	1.20E-04	8.61E-03	1.00	4.15	1.88	2.55	0.64	0.19	0.77	0.38
ZBTB16	11.06	1.24E-04	8.83E-03	0.85	0.50	0.33	0.58	0.00	0.00	0.09	0.08
HES2	3.18	1.30E-04	9.18E-03	21.69	12.94	8.74	12.26	4.34	5.01	4.63	2.95
KLF9	3.15	1.32E-04	9.27E-03	5.17	17.88	14.78	11.09	2.89	2.70	5.40	4.00
FAM105A	3.38	1.33E-04	9.33E-03	2.70	5.08	3.68	2.70	0.56	0.58	1.71	1.21
GPRC5C	3.69	1.43E-04	9.93E-03	11.66	6.57	18.38	17.58	2.98	5.11	2.31	3.96
NMNAT2	2.94	1.45E-04	1.00E-02	11.17	12.42	5.72	7.37	2.25	2.41	4.54	2.95
FHL1	2.74	1.45E-04	1.00E-02	25.86	25.03	26.87	31.00	12.87	7.81	7.80	9.29
RARRES1	2.37	1.45E-04	1.00E-02	347.21	537.63	670.92	420.79	176.99	131.31	311.63	207.20
NEDD9	2.33	1.47E-04	1.01E-02	85.99	76.72	55.38	60.42	30.64	23.43	36.33	25.61
PDLIM3	2.81	1.48E-04	1.01E-02	15.98	11.87	13.80	11.09	4.26	4.92	4.03	4.83
NFASC	2.92	1.59E-04	1.08E-02	18.76	22.81	10.45	14.08	3.62	7.71	4.97	4.99
LYPD6B	2.73	1.74E-04	1.16E-02	9.69	11.73	9.96	8.68	3.22	4.05	4.20	2.64
TFPI2	3.09	1.81E-04	1.20E-02	6.79	6.51	6.04	4.30	1.21	2.22	2.91	1.21
SLC15A2	2.52	1.85E-04	1.22E-02	29.64	58.92	38.39	28.01	10.29	10.12	24.85	14.88

RCAN1	2.24	1.95E-04	1.28E-02	334.10	536.15	499.96	381.61	160.99	133.73	312.57	166.18
MAMLD1	2.24	1.95E-04	1.28E-02	81.05	79.94	59.46	65.36	32.81	26.51	34.79	28.78
PIWIL4	3.96	2.00E-04	1.30E-02	1.47	3.43	1.72	1.17	0.48	0.19	0.69	0.45
SELL	5.78	2.12E-04	1.37E-02	0.85	0.72	0.90	0.66	0.08	0.00	0.17	0.23
LGALS9B	5.52	2.12E-04	1.37E-02	4.86	2.47	4.34	4.65	1.01	0.00	0.38	1.42
S1PR3	2.49	2.50E-04	1.58E-02	25.24	33.89	20.66	24.46	8.12	6.75	17.22	9.37
C1QTNF1	2.83	2.53E-04	1.59E-02	84.37	61.35	93.28	115.34	33.53	40.40	15.68	29.38
RASA3	2.78	2.53E-04	1.59E-02	20.30	13.09	12.09	14.44	5.15	5.88	4.28	5.36
FAM46C	2.80	2.55E-04	1.60E-02	8.26	15.37	15.60	8.68	4.10	2.41	5.31	4.61
LOC100507410	3.39	2.59E-04	1.60E-02	3.63	4.15	1.55	2.77	0.88	0.48	1.11	0.91
NNMT	3.48	2.67E-04	1.65E-02	3.94	2.22	2.61	3.36	0.88	0.96	0.60	0.91
FAS	4.18	2.87E-04	1.77E-02	1.85	9.22	4.98	5.98	0.80	0.29	2.66	1.44
DGKI	35.21	2.90E-04	1.78E-02	0.31	0.72	0.33	0.07	0.00	0.00	0.00	0.00
ABCA4	2.82	3.08E-04	1.86E-02	11.89	15.30	5.80	7.30	2.73	2.89	5.06	3.02
PDZK1	3.81	3.13E-04	1.88E-02	5.31	11.25	2.26	6.20	0.71	0.72	3.07	1.92
GPRASP2	11.88	3.16E-04	1.89E-02	9.06	2.91	2.49	2.48	0.00	0.00	1.62	0.00
GAB3	7.63	3.54E-04	2.08E-02	0.39	1.36	0.90	0.29	0.00	0.00	0.17	0.15
ARL4D	2.69	3.59E-04	2.10E-02	21.00	18.16	17.23	27.36	6.35	9.83	6.01	7.86
CDKN1C	2.98	3.68E-04	2.14E-02	9.73	5.01	7.92	6.35	2.25	2.22	2.06	2.95
PLAGL1	3.29	3.83E-04	2.19E-02	11.73	33.61	12.25	9.34	3.38	1.74	9.68	4.83
TMEM45A	2.19	3.85E-04	2.19E-02	325.13	414.93	435.43	310.99	194.20	111.45	222.09	131.13
NRARP	2.52	3.88E-04	2.20E-02	26.32	21.45	17.40	21.08	9.17	7.52	7.11	8.91
PDE8B	3.41	3.95E-04	2.22E-02	1.93	4.58	3.59	2.85	0.56	0.29	1.46	1.36
RAB31	2.35	3.96E-04	2.22E-02	76.12	87.38	44.76	56.76	20.59	17.35	49.52	25.08
TM4SF19	31.80	4.03E-04	2.25E-02	0.34	0.47	0.20	0.23	0.00	0.00	0.00	0.00
IL6R	3.23	4.17E-04	2.33E-02	5.48	5.72	2.04	2.41	0.96	0.96	1.54	1.13
CHRDL1	5.25	4.34E-04	2.40E-02	1.24	0.57	2.37	1.17	0.16	0.10	0.17	0.53
TSPAN1	2.87	4.44E-04	2.44E-02	68.31	41.61	78.43	82.00	30.64	28.44	12.08	17.98
HBEGF	2.84	4.54E-04	2.49E-02	16.13	11.89	8.00	12.18	5.55	3.57	3.34	3.63
LRRC31	5.11	4.76E-04	2.60E-02	0.54	1.43	1.88	1.02	0.00	0.10	0.34	0.45
LAMA4	2.06	4.88E-04	2.66E-02	263.61	441.03	310.79	297.46	137.83	102.97	220.72	157.94
TF	2.99	5.05E-04	2.74E-02	44.39	44.90	129.62	48.08	23.96	29.02	14.91	17.07
FUT4	3.64	5.23E-04	2.83E-02	2.93	5.15	2.70	1.60	0.32	0.39	1.97	0.68
EGLN3	2.18	5.38E-04	2.90E-02	241.76	155.65	222.00	161.95	96.33	71.93	107.53	76.29
ABCA12	3.42	5.64E-04	2.99E-02	3.94	7.87	3.27	5.84	1.37	0.39	3.17	1.13
C21orf63	2.46	5.63E-04	2.99E-02	18.06	13.51	14.46	14.52	5.63	6.46	6.34	5.51
SRGN	3.68	6.10E-04	3.17E-02	1.47	3.93	1.31	1.97	0.48	0.10	0.86	0.76
CFI	3.75	6.24E-04	3.22E-02	1.00	5.22	4.57	2.26	0.40	0.58	1.29	1.06
G0S2	15.23	6.28E-04	3.23E-02	0.62	0.57	0.00	0.66	0.00	0.00	0.09	0.00
HAPLN3	2.25	6.37E-04	3.26E-02	70.17	71.07	42.31	62.30	20.10	29.70	22.19	32.18
LIPG	2.39	6.41E-04	3.27E-02	42.22	43.19	19.60	23.86	10.94	10.99	19.96	10.65

SMN2	299.20	6.70E-04	3.38E-02	11.27	0.00	0.00	0.00	0.00	0.00	0.00	0.00
CDHR3	8.25	6.74E-04	3.38E-02	0.69	1.07	0.08	0.61	0.00	0.10	0.17	0.00
CREG2	12.09	6.78E-04	3.39E-02	0.15	0.50	0.57	0.29	0.08	0.00	0.00	0.00
FLG	32.86	7.03E-04	3.49E-02	0.31	0.72	0.00	0.29	0.00	0.00	0.00	0.00
TP63	2.67	7.29E-04	3.58E-02	24.24	35.82	12.74	25.97	10.61	4.05	13.45	7.55
BLVRA	2.17	7.26E-04	3.58E-02	56.97	48.26	54.15	49.39	25.57	20.73	23.82	22.96
SERPINB12	31.66	7.42E-04	3.63E-02	0.08	0.64	0.41	0.15	0.00	0.00	0.00	0.00
SPHK1	3.11	7.54E-04	3.68E-02	9.26	7.51	4.82	12.40	2.49	3.66	1.54	2.72
ABCC11	2.85	7.59E-04	3.68E-02	2.70	2.43	2.29	2.33	0.72	0.58	1.03	0.98
LPCAT1	2.11	7.60E-04	3.68E-02	147.44	112.69	80.45	91.41	44.95	52.64	58.09	42.22
RASL10A	3.47	8.09E-04	3.87E-02	1.70	1.64	1.63	1.09	0.32	0.29	0.26	0.76
C12orf75	2.68	8.22E-04	3.93E-02	6.18	7.36	6.62	5.98	2.49	1.06	3.94	2.11
HRASLS5	2.34	8.80E-04	4.13E-02	18.14	21.74	16.42	14.59	4.82	5.79	12.42	7.10
COL13A1	2.72	9.22E-04	4.29E-02	9.80	15.96	4.17	8.90	3.38	3.09	4.28	2.64
SLC2A10	2.27	9.26E-04	4.29E-02	31.11	38.47	31.94	21.01	10.45	7.62	19.45	15.41
FOXO1	2.59	9.78E-04	4.51E-02	3.86	6.15	6.70	5.03	1.85	1.54	3.08	1.74
C3orf36	9.01	1.02E-03	4.65E-02	0.77	0.64	0.25	0.22	0.00	0.00	0.00	0.15
VSNL1	2.46	1.03E-03	4.70E-02	13.74	28.89	18.46	28.38	8.12	5.11	14.82	7.93
CPA4	2.17	1.06E-03	4.79E-02	111.40	98.17	46.15	68.79	35.22	31.62	48.24	30.06
KCNMB4	2.53	1.07E-03	4.83E-02	5.79	5.86	6.13	5.11	2.49	2.02	1.97	2.11
SLC9A2	3.41	1.09E-03	4.90E-02	3.16	13.66	5.96	5.25	0.80	0.87	4.63	1.81
MYH3	3.94	1.11E-03	4.97E-02	0.85	1.98	0.99	1.15	0.00	0.29	0.51	0.40

Genes downregulated in *NR2F1*-high samples

				Treatment samples: *NR2F1*-high				Control samples: *NR2F1*-low			
gene_id	**Fold Change**	**pvalue**	**padj**	**MH01**	**MH02**	**MH03**	**MH04**	**MH05**	**MH06**	**MH07**	**MH08**
IFI27	-12.74	4.33E-18	1.73E-14	17.68	10.87	17.15	19.70	248.18	203.05	116.55	228.11
MAOA	-9.98	4.59E-13	5.63E-10	1.62	2.29	4.41	1.68	22.52	14.56	35.22	26.81
NGFR	-7.47	3.42E-11	2.60E-08	9.88	5.93	9.72	10.80	82.26	72.12	35.47	68.13
TNFAIP2	-4.87	2.80E-09	1.21E-06	86.30	101.89	43.94	50.19	205.70	513.87	320.12	244.13
STAC2	-4.50	3.04E-09	1.27E-06	13.59	17.45	19.11	16.63	89.10	64.31	49.78	79.39
TMEM100	-5.13	5.54E-09	1.95E-06	3.01	6.86	4.49	3.87	18.33	13.50	39.40	20.92
MFI2	-4.94	2.94E-08	8.76E-06	4.88	4.86	4.08	4.38	15.92	30.75	15.34	23.19
ITGB2	-4.12	2.97E-08	8.76E-06	24.66	18.15	33.13	25.39	107.13	115.27	75.30	103.11
WIF1	-6.14	3.41E-08	9.72E-06	0.69	1.14	0.90	0.73	4.74	2.60	6.85	6.65
SLC38A3	-5.44	4.68E-08	1.29E-05	5.71	4.57	9.07	7.95	44.55	40.20	19.95	36.41
FAT3	-8.08	5.16E-08	1.39E-05	0.46	0.79	0.33	0.44	2.89	1.54	7.80	4.38
FAM176A	-3.44	6.36E-08	1.66E-05	37.21	28.38	27.61	24.07	109.76	80.89	95.36	101.59
SPON2	-6.81	7.39E-08	1.87E-05	1.54	0.72	0.57	0.88	3.78	8.87	5.40	6.65

LOC645638	-4.59	7.76E-08	1.93E-05	6.64	4.79	8.58	5.91	34.50	29.41	21.25	28.48
SYT8	-5.80	8.12E-08	1.99E-05	4.94	4.22	1.63	2.48	7.88	19.28	20.57	27.27
KIAA1210	-6.33	1.60E-07	3.71E-05	0.39	1.86	1.80	0.58	5.31	4.63	9.17	9.06
SH3TC1	-4.29	2.07E-07	4.47E-05	11.89	6.22	5.23	6.58	22.92	32.20	30.33	40.18
NES	-4.00	2.22E-07	4.72E-05	8.11	5.93	8.74	8.61	31.20	35.48	25.68	28.70
TNS1	-3.45	4.04E-07	8.16E-05	19.61	13.87	17.56	17.58	51.44	66.72	51.23	59.75
EDN2	-3.87	4.22E-07	8.41E-05	15.59	7.65	12.17	12.40	47.44	47.92	39.76	45.02
CA12	-3.49	5.58E-07	1.08E-04	12.43	13.09	15.19	7.51	36.75	28.83	62.98	37.92
DYX1C1-CCPG1	-8.37	6.00E-07	1.14E-04	0.22	0.72	0.01	0.21	1.20	2.99	2.34	2.88
LINC00284	-3.81	6.36E-07	1.18E-04	3.55	4.00	4.49	3.28	9.81	12.15	22.36	13.97
LOC100505746	-3.85	6.79E-07	1.24E-04	4.47	6.59	2.89	2.55	12.45	13.82	18.86	15.18
LRRC55	-3.32	7.18E-07	1.30E-04	19.84	32.03	25.32	22.03	104.24	67.78	89.37	48.49
ITGAL	-4.51	7.45E-07	1.33E-04	2.47	2.07	4.25	3.21	13.35	13.88	10.28	14.80
UNC5B	-3.96	7.52E-07	1.33E-04	18.99	12.87	12.17	20.86	58.94	88.51	43.70	54.76
NT5E	-4.11	8.20E-07	1.39E-04	12.04	40.26	22.38	17.22	78.24	33.84	155.77	98.57
CD22	-9.82	8.82E-07	1.48E-04	0.23	0.14	0.00	0.15	0.96	1.83	1.03	1.44
ARSA	-3.56	8.91E-07	1.48E-04	10.73	9.72	7.19	6.93	18.41	31.14	30.25	39.43
PRSS21	-9.55	1.01E-06	1.66E-04	0.46	0.07	0.25	0.22	1.61	3.86	1.37	2.57
ATP6V0D2	-4.31	1.50E-06	2.34E-04	2.01	1.86	2.94	1.09	7.00	6.36	12.68	7.80
RNF150	-4.71	1.62E-06	2.46E-04	3.40	2.36	3.10	3.21	11.82	5.40	26.82	14.91
ALDH3B1	-3.60	1.71E-06	2.57E-04	11.33	9.13	10.32	10.38	33.72	43.53	22.99	40.74
ADRB2	-4.25	3.77E-06	5.23E-04	3.24	2.50	1.63	2.19	9.01	13.59	10.71	5.82
LOC100499194	-8.93	3.87E-06	5.33E-04	0.15	0.14	0.08	0.15	0.80	1.83	0.77	1.36
LYNX1	-3.57	4.10E-06	5.55E-04	35.05	23.81	44.27	43.55	134.05	152.62	71.37	142.76
LOC100507218	-8.24	4.35E-06	5.83E-04	0.18	0.15	0.09	0.07	0.75	1.00	1.33	1.09
ATP6V1C2	-2.90	4.74E-06	6.29E-04	24.43	37.34	20.22	13.49	54.92	55.85	91.52	61.93
FRMPD1	-6.46	5.14E-06	6.72E-04	0.15	0.29	0.25	0.15	1.21	1.54	1.54	1.06
CLDN10	-2.90	5.20E-06	6.74E-04	45.08	44.44	65.26	48.51	170.93	138.01	129.90	125.03
RORC	-3.96	6.14E-06	7.70E-04	2.08	1.72	2.12	1.24	6.75	8.39	5.83	6.04
HCP5	-2.76	8.98E-06	1.07E-03	37.52	31.39	32.18	26.04	68.35	102.01	73.94	91.93
CD14	-3.62	1.14E-05	1.30E-03	33.19	23.24	50.23	47.49	111.61	178.58	72.57	175.29
ADAMTS1	-3.40	1.15E-05	1.31E-03	25.63	16.23	12.82	11.16	66.42	28.35	80.80	44.87
GPNMB	-3.12	1.18E-05	1.33E-03	34.12	23.52	49.99	23.71	123.11	76.36	130.24	74.25
C2orf54	-3.30	1.22E-05	1.36E-03	30.03	15.73	20.99	31.22	75.11	97.76	52.87	87.09
CACNG4	-3.92	1.24E-05	1.38E-03	1.54	1.00	1.63	1.82	6.11	5.98	5.23	5.59
FAM3D	-4.10	1.30E-05	1.42E-03	17.45	15.59	40.10	31.00	156.16	96.90	50.21	100.46
KRT38	-3.50	1.36E-05	1.47E-03	1.70	1.57	1.88	2.04	6.27	5.69	6.08	6.42
WFDC2	-5.98	1.42E-05	1.51E-03	0.31	0.43	1.23	0.66	5.55	3.28	1.97	4.00

PRSS22	-3.02	1.47E-05	1.55E-03	11.19	9.94	9.07	8.83	25.49	33.74	22.18	30.60
FGD3	-3.07	1.48E-05	1.56E-03	15.75	10.80	11.76	11.75	31.20	45.37	28.45	42.58
LY6D	-4.94	1.51E-05	1.56E-03	5.40	1.93	1.88	2.26	13.83	18.99	6.17	14.65
GMPR	-3.91	1.50E-05	1.56E-03	3.24	2.15	3.10	3.36	10.78	14.56	6.85	12.09
FAM5B	-3.23	1.62E-05	1.64E-03	8.95	6.15	6.62	7.37	27.10	23.52	19.02	20.24
LEMD1	-2.93	1.61E-05	1.64E-03	26.09	23.24	32.10	22.98	88.53	81.18	52.78	66.17
SNCAIP	-4.61	1.70E-05	1.72E-03	0.23	0.50	0.57	0.66	2.33	1.45	3.08	2.11
TMEM211	-4.95	1.78E-05	1.78E-03	0.46	0.29	0.25	0.44	1.69	1.93	1.63	1.74
SLC28A1	-3.47	1.80E-05	1.80E-03	7.87	6.22	9.72	7.08	33.37	29.02	16.88	21.68
ISM2	-11.28	2.21E-05	2.16E-03	0.15	0.00	0.08	0.07	1.29	0.67	0.43	1.21
H3F3A	-181.88	2.39E-05	2.32E-03	0.00	0.00	0.00	0.00	0.00	0.00	4.14	3.65
FGFBP1	-2.98	2.42E-05	2.33E-03	294.72	234.82	477.58	279.48	1239.39	874.47	759.82	793.32
RGPD8	-312.57	2.49E-05	2.37E-03	0.00	0.00	0.00	0.00	0.00	0.00	13.81	0.87
S100A7	-4.23	2.56E-05	2.43E-03	2.15	12.73	3.02	6.05	19.38	31.24	28.19	13.52
ANKRD1	-3.65	2.61E-05	2.46E-03	6.15	16.87	7.68	6.61	22.58	13.02	67.34	31.72
ATP6V1B1	-2.98	2.97E-05	2.75E-03	209.81	180.00	267.01	282.03	727.90	783.94	388.81	764.52
MUC16	-2.86	3.00E-05	2.75E-03	34.27	29.10	19.28	16.27	35.62	73.08	107.62	64.43
MMP7	-2.38	3.18E-05	2.88E-03	283.06	392.48	311.66	301.87	739.72	546.86	1074.32	627.17
RASD1	-3.50	3.24E-05	2.91E-03	2.16	1.43	1.80	1.75	4.50	7.91	7.03	5.14
SYNM	-2.84	3.70E-05	3.30E-03	8.26	11.51	7.84	7.44	16.48	18.32	41.90	22.13
EVPLL	-4.56	3.73E-05	3.31E-03	0.39	0.29	0.74	0.29	2.03	1.55	1.99	1.98
LOC100506178	-5.26	4.34E-05	3.80E-03	0.85	0.36	0.25	0.22	1.29	1.64	4.20	1.96
C11orf53	-4.45	4.41E-05	3.84E-03	1.54	1.07	1.72	1.46	9.25	6.07	2.74	6.04
LOC100133669	-6.92	4.74E-05	4.02E-03	0.15	0.29	0.00	0.44	1.05	2.41	0.94	1.66
PPP1R1A	-3.93	5.31E-05	4.40E-03	1.62	0.79	2.04	1.39	6.59	4.92	4.11	6.57
UCA1	-3.31	5.34E-05	4.40E-03	4.86	3.07	3.76	2.19	9.65	14.08	9.85	10.50
HSH2D	-3.23	5.87E-05	4.70E-03	4.40	3.00	2.94	2.77	8.60	12.63	8.31	11.10
ZFP57	-2.84	8.80E-05	6.65E-03	6.33	5.15	3.76	4.09	12.71	13.31	13.88	12.92
MOCS1	-3.38	9.87E-05	7.26E-03	2.81	2.04	2.75	2.44	9.43	9.29	4.94	8.47
TMOD1	-2.60	9.89E-05	7.26E-03	12.87	9.82	9.35	9.44	22.98	28.17	31.25	22.84
SFTA1P	-2.74	1.06E-04	7.69E-03	5.48	4.79	5.23	4.96	14.23	12.63	16.02	11.71
IRS4	-3.69	1.31E-04	9.21E-03	0.62	1.00	0.65	0.29	2.17	1.25	3.08	2.64
HSD11B2	-2.48	1.66E-04	1.12E-02	38.36	30.67	28.10	33.04	65.54	92.85	55.78	94.04
PCDH8	-3.03	1.70E-04	1.14E-02	4.63	3.65	5.06	2.70	15.84	9.26	11.48	9.74
PARVG	-3.96	1.74E-04	1.16E-02	0.62	0.50	0.49	0.36	2.65	1.74	1.54	1.51
MAP2K6	-3.53	1.77E-04	1.18E-02	2.39	2.07	3.51	1.75	9.33	2.89	12.08	9.89
RGS2	-3.48	2.16E-04	1.39E-02	1.93	6.29	6.21	3.14	9.01	7.91	32.56	12.46
RASL11A	-3.33	2.18E-04	1.40E-02	1.78	2.50	3.51	2.85	10.53	8.87	4.63	9.14
SLC1A6	-3.26	2.34E-04	1.49E-02	30.41	18.66	62.08	42.75	178.46	126.20	65.03	108.77
TNF	-3.19	2.36E-04	1.50E-02	6.02	5.22	2.70	2.77	9.25	20.34	10.80	9.67

TFCP2L1	-2.50	2.58E-04	1.60E-02	653.04	639.67	327.29	255.53	671.61	964.71	1949.83	1050.83
ADH1C	-4.45	2.95E-04	1.80E-02	1.65	1.79	10.12	2.19	9.65	11.17	26.97	24.83
FBXO32	-2.21	2.97E-04	1.80E-02	54.27	83.16	60.85	50.04	122.95	81.18	211.12	121.08
SLC6A2	-2.25	3.05E-04	1.85E-02	39.83	39.76	37.33	35.82	80.90	98.15	67.08	79.77
KLHL38	-3.20	3.15E-04	1.89E-02	1.24	1.07	1.39	0.73	4.34	2.99	2.91	3.17
ABCG2	-3.36	3.29E-04	1.95E-02	0.69	1.14	1.06	1.17	2.73	1.54	6.00	3.55
RASSF10	-2.65	3.30E-04	1.95E-02	11.81	6.44	5.80	7.22	14.71	18.99	25.36	23.42
LYPD5	-2.24	3.30E-04	1.95E-02	35.05	28.45	26.46	27.14	58.38	71.73	60.42	62.15
CYP24A1	-2.94	3.55E-04	2.08E-02	1.93	4.29	1.39	1.46	4.91	5.50	9.60	5.36
ETV7	-3.03	3.61E-04	2.10E-02	3.09	1.57	3.19	2.63	7.32	9.54	6.25	7.70
CTGF	-2.35	3.71E-04	2.14E-02	140.64	97.68	118.35	106.51	370.45	186.07	271.87	220.11
LOC728978	-5.81	3.81E-04	2.19E-02	0.31	0.00	0.00	0.36	0.72	1.06	0.94	1.36
ANGPT1	-4.55	3.85E-04	2.19E-02	0.15	3.43	2.70	2.12	6.92	3.37	15.77	11.63
OASL	-2.85	3.90E-04	2.21E-02	9.88	6.01	6.13	8.75	20.18	27.96	13.19	22.51
CCL17	-3.51	4.22E-04	2.34E-02	0.23	0.57	0.57	0.58	1.61	1.93	1.80	1.36
UST	-2.41	4.39E-04	2.42E-02	10.50	10.73	10.05	8.75	23.24	14.75	35.90	21.45
FAM107A	-7.05	5.57E-04	2.99E-02	0.08	0.07	0.25	0.07	0.49	1.54	0.34	0.91
LOC100505839	-2.64	5.62E-04	2.99E-02	12.20	7.08	5.47	8.17	20.42	25.16	17.91	20.02
ATP6V0A4	-2.18	5.61E-04	2.99E-02	63.40	50.16	66.75	54.46	120.85	147.20	111.90	113.01
RGS16	-2.92	5.74E-04	3.03E-02	1.16	1.00	0.82	0.88	2.73	3.09	2.74	2.27
PCDH19	-2.69	5.76E-04	3.03E-02	3.47	4.79	5.06	3.50	11.98	5.50	16.79	10.20
ACSS1	-2.42	5.89E-04	3.09E-02	7.95	7.79	8.90	7.30	19.22	19.38	15.85	19.34
LGALS7B	-3.34	6.06E-04	3.17E-02	4.32	1.35	2.61	5.22	11.84	10.62	7.46	14.49
ALDH2	-2.41	6.11E-04	3.17E-02	27.17	21.16	38.55	39.10	84.27	83.10	62.03	64.58
PCSK9	-2.74	6.22E-04	3.22E-02	14.67	6.86	10.29	8.39	30.07	33.65	17.91	23.34
MX2	-2.99	6.29E-04	3.23E-02	6.72	4.15	1.88	2.48	6.52	12.73	12.85	12.32
RINL	-2.68	6.50E-04	3.30E-02	12.44	7.94	10.86	11.89	37.55	27.09	16.54	28.55
SLC16A3	-2.56	6.70E-04	3.38E-02	31.03	16.30	31.69	28.82	77.84	71.83	40.69	74.32
PGLYRP4	-3.63	6.75E-04	3.38E-02	0.85	0.79	1.14	0.95	4.26	4.63	1.54	2.19
ARL4C	-2.06	7.03E-04	3.49E-02	131.38	117.12	91.15	90.45	166.53	160.82	331.67	222.06
ARSG	-2.73	7.13E-04	3.53E-02	2.01	1.57	1.72	1.46	4.50	4.82	3.77	4.53
LOC100128054	-8.88	7.27E-04	3.58E-02	0.08	0.07	0.08	0.00	0.26	0.48	1.35	0.30
OSR1	-2.67	7.72E-04	3.73E-02	4.32	3.22	3.43	5.11	9.33	10.12	8.14	14.20
SARM1	-2.20	7.81E-04	3.76E-02	16.92	17.53	14.23	15.61	33.00	36.99	30.00	34.75
MIP	-8.80	7.85E-04	3.77E-02	0.00	0.14	0.00	0.00	0.56	0.39	0.43	0.15
LHX4	-2.93	8.31E-04	3.95E-02	1.73	1.87	0.79	0.58	3.27	2.38	4.42	3.91
FLJ41200	-3.62	8.36E-04	3.97E-02	1.31	6.22	5.64	3.06	10.05	3.95	32.99	12.84
CHADL	-3.33	8.57E-04	4.06E-02	1.64	0.79	0.66	1.33	2.52	3.11	2.77	6.11
C10orf116	-2.39	8.74E-04	4.12E-02	18.12	12.99	12.74	13.86	27.26	39.33	23.36	41.84

LRRN2	-2.73	8.79E-04	4.13E-02	6.72	3.36	5.55	5.76	14.62	17.35	9.85	14.43
TNNI2	-2.69	9.00E-04	4.21E-02	3.47	6.65	3.84	5.03	4.74	14.08	18.85	12.69
PDE5A	-3.32	9.26E-04	4.29E-02	2.70	18.38	11.68	7.66	16.73	13.02	71.43	32.93
ITGA9	-2.14	9.53E-04	4.40E-02	18.56	16.45	13.80	17.14	32.73	29.41	40.96	34.82
PDE4B	-2.23	1.01E-03	4.62E-02	18.99	40.04	25.65	19.04	44.71	32.88	96.99	51.89
C1orf64	-14.00	1.02E-03	4.65E-02	0.08	0.00	0.00	0.00	0.24	0.10	0.43	0.76
LGALS7	-4.79	1.06E-03	4.79E-02	0.32	0.15	0.00	0.55	1.03	1.34	0.77	1.75
TNNC1	-3.19	1.11E-03	4.97E-02	3.16	2.50	5.15	4.23	16.57	12.05	4.37	12.01

An adjusted p-value cut-off of <0.05 was considered statistically significant. Treatment and control sample values are normalized counts per million (CPM) values obtained from EdgeR analysis.

S4 Table: Functional annotation clustering of upregulated gene set using DAVID - top 25 clusters.

Due to its size, this table is only available in the electronic version of this thesis.

S5 Table: Functional annotation clustering of downregulated gene set using DAVID - top 25 clusters.

Due to its size, this table is only available in the electronic version of this thesis.

Supplemental Methods – Integer linear program (IP) approach

The IP methods outlined below were developed through a collaboration with Dr. Mark Craven and Sid Kiblawi in the Department of Computer Sciences and Department of Biostatistics and Medical Informatics.

IP notation and variables

The NR2F1 signaling subnetwork was inferred by solving two integer linear programs (IP). The first of which determined how many of the DE nodes (genes) could be reached within the given background. The second IP constructed the subnetwork using the information generated from the first. We encoded the relevance of each node, edge, and candidate paths as binary variables. Inference was performed by selecting a combination of paths that satisfied the constraints listed below and optimized the objective functions.

Notation

We denoted the set of nodes as $\mathcal{N}$ where each node, *n*, represented either a protein or a target gene. $\mathcal{N}^S$ was a single element set containing the source node, NR2F1. $\mathcal{N}^T$ was the set of DE genes. $\mathcal{N}^{T*}$ was a subset of $\mathcal{N}^T$ that represented the number of target genes that could actually be reached. $\mathcal{N}^I$ represented the remaining nodes that were either gene regulatory proteins or intermediary factors that were present in our background network but were not source nodes or targets.

The complete set of edges, $\mathcal{E}$, was the union of subsets $\mathcal{E}^U$ and $\mathcal{E}^D$ which represented the undirected (protein-protein interactions) and directed (regulatory interactions) respectively. We denoted an edge between two nodes n_i and n_j as $e = (n_i, n_j)$. $\mathcal{E}(\mathrm{n})$ referred to the edges that touched a particular node n. $\mathcal{N}(\mathrm{e})$ returned the nodes that were involved in that edge.

The set of paths was denoted as $\mathcal{P}$, with each path represented as *p*. All paths in the set were source-target paths. $\mathcal{E}$(p) returned all edges in a given path. $\mathcal{N}(\mathrm{p})$ returned all nodes that were contained within a specific path.

Variables

The relevance of a path *p* was represented with a binary variable σ_p, which took the value of 1 if the particular path was predicted to be included, and 0 otherwise. The predicted relevance of an edge was represented by the binary variable x_e, which took a value of 1 if the edge was in at least one relevant path, and 0 otherwise. The predicted relevance of a node was represented by the binary variable y_n. The variable received a value of 1 if it was present in any relevant paths, and 0 otherwise.

IP constraints and objective functions

Determining relevant target nodes

The first integer linear program determined how many of the DE genes could actually be reached through the background network satisfying the following constraints:

All edges in a relevant path were relevant. For an edge, *e,* to be relevant (x_e=1) there must have been at least one relevant path that contained it ($\sigma_p = 1$). Therefore, a relevant path *p* must have been composed of relevant edges, *e*.

$$\forall e \in \mathcal{E} \quad x_e \leq \sum\nolimits_{p \in \mathcal{P}(e)} \sigma_p$$

$$\forall p \in \mathcal{P} \qquad \forall e \in \mathcal{E}\,(p) \qquad \sigma_p \leq x_e$$

All nodes in a relevant edge were relevant. A node, *n,* was relevant if it was connected to a relevant edge, *e,* (*where* x_e=1). Each node, *n,* for a relevant edge, *e,* must have been relevant (y_n=1).

$$\forall n \in \mathcal{N} \quad y_n \leq \sum_{e \in \mathcal{E}(n)} x_e$$

$$\forall e \in \mathcal{E} \quad \forall n \in \mathcal{N}(e) \quad x_e \leq y_n$$

All differentially expressed genes must have come from a regulatory element. In order for a gene to have been DE, at some point in a path there was a change to one of the gene's regulators. We wanted all paths to end with a regulatory element acting on a DE gene.

$$\forall n \in \mathcal{N}^T \quad y_n \leq \sum_{e \in \mathcal{E}^D(n)} x_e$$

No intermediary factors could come from a regulatory element edge. Since most of our intermediary factors and gene regulatory proteins were not part of the DE gene set, it was safe to assume that these genes were not changing their expression.

$$\forall n \in \mathcal{N}^I \quad 0 = \sum_{e \in \mathcal{E}^D(n)} x_e$$

Objective function. In order to determine which of the target nodes could actually be reached following these constraints we used the following objective function:

$$\max\left(|\mathcal{N}^{T*}| = \sum_{n\in \mathcal{N}^{T}} y_n\right)$$

The set $\mathcal{N}^{T*}$ was composed of all genes that were determined to be relevant based on the constraints and the objective function above. This set was what we used in the second IP.

Determining intermediary nodes

The second objective function we used contained all of the same constraints in the first IP, but by replacing $\mathcal{N}^{T}$ with $\mathcal{N}^{T*}$, an additional constraint was added.

There must have been at least one path leading to a target node. In order to include as many target genes in the subnetwork as possible, we forced the subnetwork to include at least one relevant path where a target gene was also relevant.

$$\forall n \in \mathcal{N}^{T*} \qquad 1 \leq \sum_{p\in\mathcal{P}(n)} \sigma_p$$

Objective functions. Our goal was to infer a subnetwork that minimized redundancy while still containing all of the possible DE genes. Therefore, we wanted the smallest number of gene regulatory proteins and intermediary factors as possible in our network.

$$\min \sum_{n\in \mathcal{N}} y_n$$

The objective function above gave us the size of the network, but there was the possibility of multiple solutions that satisfied this objective function. To improve the plausibility of our network,

intermediary factors and gene regulatory proteins were selected that had the highest relevance to the literature based on their RNA-Seq normalized counts per million (CPM) values, $w(n)$. In order to further reduce the chance of multiple solutions, we selected the potential nodes that had the highest GADGET score, $g(n)$, from the subset of nodes that achieved the highest RNA-Seq score, $\mathcal{N}^{I*}$.

$$\max\left(|N^{I*}| = \sum_{n \in \mathcal{N}^{I}} y_n w(n)\right)$$

$$max \sum_{n \in \mathcal{N}^{I*}} y_n g(n)$$

Finally, as multiple paths may have been used to reach the target nodes, and we did not know whether one or many gene regulatory proteins were acting on the DE genes, the edges between nodes selected for our subnetwork were maximized. This was true for protein-protein interactions within the network as well.

$$max \sum_{e \in \mathcal{E}} x_e$$

References

1. American Cancer Society. Breast Cancer Facts & Figures 2015-2016. Atlanta: 2015. Available: http://www.cancer.org/acs/groups/content/@research/documents/document/acspc-046381.pdf

2. Berry DA, Cronin KA, Plevritis SK, Fryback DG, Clarke L, Zelen M, et al. Effect of Screening and Adjuvant Therapy on Mortality from Breast Cancer. N Engl J Med 2005;353:1784–92. doi:10.1056/NEJMoa050518.

3. (EBCTCG) EBCTCG. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet 2005;365:1687–717. doi:10.1016/S0140-6736(05)66544-0.

4. Saurel CA, Patel TA, Perez EA. Changes to adjuvant systemic therapy in breast cancer: a decade in review. Clin Breast Cancer 2010;10:196–208. doi:10.3816/CBC.2010.n.027.

5. Liedtke C, Mazouni C, Hess KR, André F, Tordai A, Mejia JA, et al. Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. J Clin Oncol 2008;26:1275–81. doi:10.1200/JCO.2007.14.4147.

6. Hsu LC, Kennan WS, Shepel LA, Jacob HJ, Szpirer C, Szpirer J, et al. Genetic identification of Mcs-1, a rat mammary carcinoma suppressor gene. Cancer Res 1994;54:2765–70.

7. Shepel LA, Lan H, Haag JD, Brasic GM, Gheen ME, Simon JS, et al. Genetic identification of multiple loci that control breast cancer susceptibility in the rat. Genetics 1998;149:289–99.

8. Lan H, Kendziorski CM, Haag JD, Shepel LA, Newton MA, Gould MN. Genetic loci controlling breast cancer susceptibility in the Wistar-Kyoto rat. Genetics 2001;157:331–9.

9. Haag JD, Shepel LA, Kolman BD, Monson DM, Benton ME, Watts KT, et al. Congenic rats reveal three independent Copenhagen alleles within the Mcs1 quantitative trait locus that confer resistance to mammary cancer. Cancer Res 2003;63:5808–12.

10. Samuelson DJ, Aperavich BA, Haag JD, Gould MN. Fine mapping reveals multiple loci and a possible epistatic interaction within the mammary carcinoma susceptibility quantitative trait locus, Mcs5. Cancer Res 2005;65:9637–42. doi:10.1158/0008-5472.CAN-05-1498.

11. Samuelson DJ, Hesselson SE, Aperavich BA, Zan Y, Haag JD, Trentham-Dietz A, et al. Rat Mcs5a is a compound quantitative trait locus with orthologous human loci that associate with breast cancer risk. Proc Natl Acad Sci U S A 2007;104:6299–304. doi:10.1073/pnas.0701687104.

12. Smits BMG, Sharma D, Samuelson DJ, Woditschka S, Mau B, Haag JD, et al. The non-protein coding breast cancer susceptibility locus Mcs5a acts in a non-mammary cell-autonomous fashion through the immune system and modulates T-cell homeostasis and functions. Breast Cancer Res 2011;13:R81. doi:10.1186/bcr2933.

13. Smits BMG, Traun BD, Devries TL, Tran A, Samuelson D, Haag JD, et al. An insulator loop resides between the synthetically interacting elements of the human/rat conserved breast cancer susceptibility locus MCS5A/Mcs5a. Nucleic Acids Res 2012;40:132–47. doi:10.1093/nar/gkr610.

14. Zumwalde NA, Haag JD, Sharma D, Mirrielees JA, Wilke LG, Gould MN, et al. Analysis of immune cells from human mammary ductal epithelium reveals Vδ2+ T cells that efficiently target breast carcinoma cells in the presence of bisphosphonate. Cancer Prev Res (Phila) 2016. doi:10.1158/1940-6207.CAPR-15-0370-T.

15. Smits BMG, Haag JD, Rissman AI, Sharma D, Tran A, Schoenborn AA, et al. The Gene Desert Mammary Carcinoma Susceptibility Locus Mcs1a Regulates Nr2f1 Modifying Mammary Epithelial Cell Differentiation and Proliferation. PLoS Genet 2013;9. doi:10.1371/journal.pgen.1003549.

16. Dent R, Trudeau M, Pritchard KI, Hanna WM, Kahn HK, Sawka CA, et al. Triple-negative breast cancer: clinical features and patterns of recurrence. Clin Cancer Res 2007;13:4429–34. doi:10.1158/1078-0432.CCR-06-3045.

17. Morris GJ, Naidu S, Topham AK, Guiles F, Xu Y, McCue P, et al. Differences in breast carcinoma characteristics in newly diagnosed African-American and Caucasian patients: a single-institution compilation compared with the National Cancer Institute's Surveillance, Epidemiology, and End Results database. Cancer 2007;110:876–84. doi:10.1002/cncr.22836.

18. Dai X, Li T, Bai Z, Yang Y, Liu X, Zhan J, et al. Breast cancer intrinsic subtype classification, clinical use and future trends. Am J Cancer Res 2015;5:2929–43.

19. Pereira FA, Qiu Y, Tsai MJ, Tsai SY. Chicken ovalbumin upstream promoter transcription factor (COUP-TF): Expression during mouse embryogenesis. J Steroid Biochem Mol Biol 1995;53:503–8. doi:10.1016/0960-0760(95)00097-J.

20. Tsai SY, Tsai M-J. Chick Ovalbumin Upstream Promoter-Transcription Factors (COUP-TFs): Coming of Age. Endocr Rev 1997;18:229–40.

21. Pipaón C, Tsai SY, Tsai MJ. COUP-TF upregulates NGFI-A gene expression through an Sp1 binding site. Mol Cell Biol 1999;19:2734–45.

22. Hall RK, Sladek FM, Granner DK. The orphan receptors COUP-TF and HNF-4 serve as accessory factors required for induction of phosphoenolpyruvate carboxykinase gene transcription by glucocorticoids. Proc Natl Acad Sci U S A 1995;92:412–6.

23. Power SC, Cereghini S. Positive regulation of the vHNF1 promoter by the orphan receptors COUP-TF1/Ear3 and COUP-TFII/Arp1. Mol Cell Biol 1996;16:778–91.

24. Qiu Y, Cooney AJ, Kuratani S, DeMayo FJ, Tsai SY, Tsai MJ. Spatiotemporal expression patterns of chicken ovalbumin upstream promoter-transcription factors in the developing mouse central nervous system: evidence for a role in segmental patterning of the diencephalon. Proc Natl Acad Sci U S A 1994;91:4451–5.

25. Qiu Y, Pereira FA, DeMayo FJ, Lydon JP, Tsai SY, Tsai MJ. Null mutation of mCOUP-TFI results in

defects in morphogenesis of the glossopharyngeal ganglion, axonal projection, and arborization. Genes Dev 1997;11:1925–37. doi:10.1101/gad.11.15.1925.

26. Lin B, Chen GQ, Xiao D, Kolluri SK, Cao X, Su H, et al. Orphan receptor COUP-TF is required for induction of retinoic acid receptor beta, growth inhibition, and apoptosis by retinoic acid in cancer cells. Mol Cell Biol 2000;20:957–70.

27. Kim RS, Avivar-Valderas A, Estrada Y, Bragado P, Sosa MS, Aguirre-Ghiso JA, et al. Dormancy signatures and metastasis in estrogen receptor positive and negative breast cancer. PLoS One 2012;7:e35569. doi:10.1371/journal.pone.0035569.

28. Hillen W, Gatz C, Altschmied L, Schollmeier K, Meier I. Control of expression of the Tn10-encoded tetracycline resistance genes. Equilibrium and kinetic investigation of the regulatory reactions. J Mol Biol 1983;169:707–21.

29. Hillen W, Berens C. Mechanisms underlying expression of Tn10 encoded tetracycline resistance. Annu Rev Microbiol 1994;48:345–69. doi:10.1146/annurev.mi.48.100194.002021.

30. Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol 2003;4:P3.

31. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009;4:44–57. doi:10.1038/nprot.2008.211.

32. Bernardo GM, Keri RA. FOXA1: a transcription factor with parallel functions in development and cancer. Biosci Rep 2012;32:113–30. doi:10.1042/BSR20110046.

33. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498–504. doi:10.1101/gr.1239303.

34. Kittler R, Zhou J, Hua S, Ma L, Liu Y, Pendleton E, et al. A Comprehensive Nuclear Receptor Network for Breast Cancer Cells. Cell Rep 2013;3:538–51. doi:10.1016/j.celrep.2013.01.004.

35. Bernardo GM, Lozada KL, Miedler JD, Harburg G, Hewitt SC, Mosley JD, et al. FOXA1 is an essential determinant of ERalpha expression and mammary ductal morphogenesis. Development 2010;137:2045–54. doi:10.1242/dev.043299.

36. Thorat MA, Marchio C, Morimiya A, Savage K, Nakshatri H, Reis-Filho JS, et al. Forkhead box A1 expression in breast cancer is associated with luminal subtype and good prognosis. J Clin Pathol 2008;61:327–32. doi:10.1136/jcp.2007.052431.

37. Albergaria A, Paredes J, Sousa B, Milanezi F, Carneiro V, Bastos J, et al. Expression of FOXA1 and GATA-3 in breast cancer: the prognostic significance in hormone receptor-negative tumours. Breast Cancer Res 2009;11:R40. doi:10.1186/bcr2327.

38. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature 2000;406:747–52. doi:10.1038/35021093.

39. Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001;98:10869–74. doi:10.1073/pnas.191367098.

40. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest 2011;121:2750–67. doi:10.1172/JCI45014.

41. Marcotte R, Sayad A, Brown KR, Sanchez-Garcia F, Reimand J, Haider M, et al. Functional Genomic Landscape of Human Breast Cancer Drivers, Vulnerabilities, and Resistance. Cell 2016;164:293–309. doi:10.1016/j.cell.2015.11.062.

42. Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell 2006;10:515–27. doi:10.1016/j.ccr.2006.10.008.

43. Scully S, Yan W, Bentley B, Cao QJ, Shao R. Inhibitory activity of YKL-40 in mammary epithelial cell differentiation and polarization induced by lactogenic hormones: a role in mammary tissue involution. PLoS One 2011;6:e25819. doi:10.1371/journal.pone.0025819.

44. Shao R, Cao QJ, Arenas RB, Bigelow C, Bentley B, Yan W. Breast cancer expression of YKL-40 correlates with tumour grade, poor differentiation, and other cancer markers. Br J Cancer 2011;105:1203–9. doi:10.1038/bjc.2011.347.

45. Kang EJ, Jung H, Woo OH, Park KH, Woo SU, Yang DS, et al. YKL-40 expression could be a poor prognostic marker in the breast cancer tissue. Tumour Biol 2014;35:277–86. doi:10.1007/s13277-013-1036-0.

46. Roslind A, Knoop AS, Jensen M-B, Johansen JS, Nielsen DL, Price PA, et al. YKL-40 protein expression is not a prognostic marker in patients with primary breast cancer. Breast Cancer Res Treat 2008;112:275–85. doi:10.1007/s10549-007-9870-7.

47. Linton JM, Martin GR, Reichardt LF. The ECM protein nephronectin promotes kidney development via integrin alpha8beta1-mediated stimulation of Gdnf expression. Development 2007;134:2501–9. doi:10.1242/dev.005033.

48. Kahai S, Lee S-C, Seth A, Yang BB. Nephronectin promotes osteoblast differentiation via the epidermal growth factor-like repeats. FEBS Lett 2010;584:233–8. doi:10.1016/j.febslet.2009.11.077.

49. Li S, Xie Y, Zhang W, Gao J, Wang M, Zheng G, et al. Interferon alpha-inducible protein 27 promotes epithelial-mesenchymal transition and induces ovarian tumorigenicity and stemness. J Surg Res 2015;193:255–64. doi:10.1016/j.jss.2014.06.055.

50. Xu S, Adisetiyo H, Tamura S, Grande F, Garofalo A, Roy-Burman P, et al. Dual inhibition of survivin and MAOA synergistically impairs growth of PTEN-negative prostate cancer. Br J Cancer 2015;113:242–51. doi:10.1038/bjc.2015.228.

51. Tsang JYS, Wong KHY, Lai MWH, Lacambra MD, Ko C-W, Chan SK, et al. Nerve growth factor receptor (NGFR): a potential marker for specific molecular subtypes of breast cancer. J Clin Pathol 2013;66:291–6. doi:10.1136/jclinpath-2012-201027.

52. Jia L, Zhou Z, Liang H, Wu J, Shi P, Li F, et al. KLF5 promotes breast cancer proliferation, migration and invasion in part by upregulating the transcription of TNFAIP2. Oncogene 2015. doi:10.1038/onc.2015.263.

53. Reis-Filho JS, Steele D, Di Palma S, Jones RL, Savage K, James M, et al. Distribution and significance of nerve growth factor receptor (NGFR/p75NTR) in normal, benign and malignant breast tissue. Mod Pathol 2006;19:307–19. doi:10.1038/modpathol.3800542.

54. Forgac M. Vacuolar ATPases: rotary proton pumps in physiology and pathophysiology. Nat Rev Mol Cell Biol 2007;8:917–29. doi:10.1038/nrm2272.

55. Hinton A, Sennoune SR, Bond S, Fang M, Reuveni M, Sahagian GG, et al. Function of a subunit isoforms of the V-ATPase in pH homeostasis and in vitro invasion of MDA-MB231 human breast cancer cells. J Biol Chem 2009;284:16400–8. doi:10.1074/jbc.M901201200.

56. Feng S, Zhu G, McConnell M, Deng L, Zhao Q, Wu M, et al. Silencing of atp6v1c1 prevents breast cancer growth and bone metastasis. Int J Biol Sci 2013;9:853–62. doi:10.7150/ijbs.6030.

57. Capecci J, Forgac M. The function of vacuolar ATPase (V-ATPase) a subunit isoforms in invasiveness of MCF10a and MCF10CA1a human breast cancer cells. J Biol Chem 2013;288:32731–41. doi:10.1074/jbc.M113.503771.

58. Sennoune SR, Bakunts K, Martínez GM, Chua-Tuan JL, Kebir Y, Attaya MN, et al. Vacuolar H+-ATPase in human breast cancer cells with distinct metastatic potential: distribution and functional activity. Am J Physiol Cell Physiol 2004;286:C1443–52. doi:10.1152/ajpcell.00407.2003.

59. Ohta T, Arakawa H, Futagami F, Fushida S, Kitagawa H, Kayahara M, et al. Bafilomycin A1 induces apoptosis in the human pancreatic cancer cell line Capan-1. J Pathol 1998;185:324–30. doi:10.1002/(SICI)1096-9896(199807)185:3<324::AID-PATH72>3.0.CO;2-9.

60. Kruse SW, Suino-Powell K, Zhou XE, Kretschman JE, Reynolds R, Vonrhein C, et al. Identification of COUP-TFII orphan nuclear receptor as a retinoic acid-activated receptor. PLoS Biol 2008;6:e227. doi:10.1371/journal.pbio.0060227.

61. Veillet AL, Haag JD, Remfert JL, Meilahn AL, Samuelson DJ, Gould MN. Mcs5c: A mammary carcinoma susceptibility locus located in a gene desert that associates with tenascin c expression. Cancer Prev Res 2011;4:97–106. doi:10.1158/1940-6207.CAPR-10-0187.

62. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 2011;12:323. doi:10.1186/1471-2105-12-323.

63. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010;26:139–40.

doi:10.1093/bioinformatics/btp616.

64. Chasman D, Ho Y-H, Berry DB, Nemec CM, MacGilvray ME, Hose J, et al. Pathway connectivity and signaling coordination in the yeast stress-activated signaling network. Mol Syst Biol 2014;10:759.

65. Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, et al. Human Protein Reference Database--2009 update. Nucleic Acids Res 2009;37:D767–72. doi:10.1093/nar/gkn892.

66. Stark C, Breitkreutz B-J, Reguly T, Boucher L, Breitkreutz A, Tyers M. BioGRID: a general repository for interaction datasets. Nucleic Acids Res 2006;34:D535–9. doi:10.1093/nar/gkj109.

67. Schaefer MH, Fontaine J-F, Vinayagam A, Porras P, Wanker EE, Andrade-Navarro MA. HIPPIE: Integrating protein interaction networks with experiment based quality scores. PLoS One 2012;7:e31826. doi:10.1371/journal.pone.0031826.

68. Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M. An integrated encyclopedia of DNA elements in the human genome. Nature 2012;489:57–74. doi:10.1038/nature11247.

69. Poon H, Quirk C, DeZiel C, Heckerman D. Literome: PubMed-scale genomic knowledge base in the cloud. Bioinformatics 2014;30:2840–2. doi:10.1093/bioinformatics/btu383.

70. Poon H, Toutanova K, Quirk C. Distant supervision for cancer pathway extraction from text. Pac Symp Biocomput 2015:120–31.

Chapter 4 – Conclusions and Future Directions

The *Mcs5c* locus:

***Mcs5c* acts within the mammary gland to influence the expression of Pregnancy-associated plasma protein A (*Pappa*) in an age-specific manner during the immature (iWOS) and adolescent (aWOS) windows of susceptibility.**

Mammary gland transplant experiments identified a significant donor effect on carcinoma multiplicity, indicating that *Mcs5c* acts in a mammary gland-autonomous manner. Expression analysis of nearby genes found *Pappa* to be differentially regulated between *Mcs5c* resistant and susceptible rats. This differential regulation occurred in an age-specific manner, with 4 week old susceptible rats having reduced and 6-9 week old susceptible rats having increased levels of *Pappa* expression in mammary epithelial cells (MECs). This time period corresponds to the broad birth-to-young-adult WOS (byaWOS) observed in humans, and the age-specific differences in *Pappa* expression changes suggested the existence of the immature and adolescent sub-windows which differentially alter *Mcs5c* activity.

The *Mcs5c* locus regulates *Pappa* expression in MECs through a physical interaction with an intronic region of *Pappa* that is age- and tissue-dependent.

We hypothesized that *Mcs5c* contained a regulatory element responsible for mediating *Pappa* expression differences observed in MECs. Chromosome conformation capture (3C) experiments were used to identify a physical interaction between the two regions that would be indicative of such regulation. Through 3C analysis, we identified an 8.5kb region within *Mcs5c*, termed the temporal control element (TCE), that interacts with *Pappa*'s first intron in an age-dependent and MEC-specific manner. This interaction was observed in both *Mcs5c* resistant and susceptible rats, but was specific to

time periods falling within the aWOS (6 and 7 week) and was not observed during the iWOS (4 weeks) or in adult rats (12 weeks). A reduction of TCE copies in a rat mammary carcinoma cell line (LA7) through CRISPR/Cas9 gene editing resulted in reduced *Mcs5c*-*Pappa* looping and correlated with a reduction in *Pappa* expression, suggesting that the TCE is indeed relevant to *Pappa* regulation.

Allele-dependent differential methylation of *Pappa* CpG island (CGI) shore sites may contribute to *Mcs5c* regulation of *Pappa* expression.

The intronic region of *Pappa* that is targeted by TCE looping resides in a CGI shore region, located approximately 2kb away from a conserved CGI spanning exon 1. Bisulfite pyrosequencing was used to investigate the methylation levels of CpG sites near this region, as shore methylation has been associated with expression differences. We identified 6 sites within the *Pappa* intron that are differentially methylated in MECs between *Mcs5c* resistant and susceptible rats specifically during the aWOS (6 & 7 weeks). *Mcs5c* susceptible rats had decreased methylation levels compared to resistant rats and, during this time period, methylation levels were negatively correlated with *Pappa* expression levels, suggesting that *Pappa* CGI shore methylation is an additional mechanism by which *Mcs5c* regulates *Pappa* gene expression.

This work has identified, to our knowledge, the first genetic mechanism underlying the WOS phenomenon, and highlights *Pappa* as a novel factor during this critical developmental period. Many questions, however, still remain. Of primary importance is the creation of a *Mcs5c* TCE knockout rat. TCE knockdown in LA7 cells resulted in a reduction in *Pappa* expression, suggesting that this region acts as an enhancer of *Pappa*. I, therefore, would propose the targeting of the TCE in the susceptible Wistar-Furth (WF) rat strain, and hypothesize that this would reduce *Pappa* expression in MECs during the aWOS compared to wild-type WF rats, resulting in a mammary carcinoma resistance phenotype that

mimics *Mcs5c* resistant congenic rats. This knockout model would also provide the opportunity to address another unanswered question, namely, whether *Mcs5c-Pappa* looping is the driving force behind differential shore methylation. If this is the case, we would expect to see a difference in shore methylation between TCE knockout WF and wild-type WF rats. Alternatively, if differential methylation drives *Mcs5c-Pappa* looping through some other non-TCE-dependent mechanism, methylation would be unaffected. If this were the case, we would hypothesize that carcinoma multiplicity in TCE knockout rats may not mimic *Mcs5c* resistant congenics and may either show no difference from wild-type rats or may show an intermediary phenotype, as an absence in looping, in general, may still decrease *Pappa* expression. In this instance, a *Pappa* CGI shore knockout rat would clarify the dependency of looping on shore methylation.

Our results suggested the presence of at least two sub-windows (iWOS and aWOS) in the rat within the corresponding human byaWOS. *Mcs5c* activity appears to heavily depending on age, with differences in activity observed during the iWOS and aWOS. We have speculated that these window-specific differences in function could be explained by interactions with proteins specific to these developmental time points. It is known that a great deal of gland development and differentiation takes place during this age range [1,2], and these differences in cell composition of the mammary gland may result in significant proteomic differences. A thorough investigation of cell-type differences and the proteomic landscape of MECs from these developmental time periods through flow cytometry and mass spectrometry would provide insight into differentially expressed proteins that may be involved in the differential activity of *Mcs5c* during the iWOS and aWOS. More targeted studies could also be initiated to identify proteins involved in *Mcs5c*-mediated regulation of *Pappa* specifically during the aWOS. Specifically, CTCF binding sites identified in rat liver tissue [3] within the TCE suggest that this protein may be involved in the looping interaction. I would like to verify that this interaction takes place in MECs

through ChIP-qPCR. However, this approach could not be used to identify other factors that bind this region. To do so, I propose using a technique developed by Nordhoff *et al* [4] that uses matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) to identify proteins bound to a specific DNA probe. This would allow for the identification of protein complexes bound to regions within the TCE or the targeted intronic *Pappa* region. Additionally, through the use of WF or WKy allele-specific probes, we could investigate whether strain-specific sequence variants affect protein binding or protein complex formation that may contribute to differential *Pappa* expression.

In addition to *Pappa*'s role during the aWOS, expression analysis I have performed on DMBA-induced mammary carcinomas suggests that *Pappa* may also play a role during tumor progression. Carcinomas from *Mcs5c* susceptible rats showed a 59% increase in *Pappa* expression compared to tumors from resistant rats (p-value = 0.0017, n = 17 and 19; Fig 1). To identify the relevance of *PAPP-A* expression in human breast cancers, I utilized the Oncomine database (www.oncomine.com, Thermo Fisher Scientific) to examine *PAPP-A* levels in publically available human breast cancer microarray data stratified by various clinical classifications. This analysis was similar to the one performed by Smits *et al* [5] for the *Mcs1a* locus. My analysis revealed that *PAPP-A* expression was increased in ER-, HER2+, and triple-negative (TN) breast tumors, but was not associated with tumor grade or PR status (Fig 2). In this way, *PAPP-A* seemed to be associated with more aggressive tumor types, and we speculated that it may represent a novel therapeutic strategy for TNBC. Work by others has shown that increased PAPP-A activity enhances the growth of ovarian and lung cancer cell lines [6,7], and we sought to test if this is true for TNBC cells. An overexpression strategy identical to that described for *NR2F1* in Chapter 3 was employed; however, overexpression of *PAPP-A* in MDA-MB-468 cells failed to increase carcinoma growth *in vivo* (data not shown). I hypothesize that this may be due to the presence of an inhibitory protein, such as Stanniocalcin-2 (STC2) [8], that diminishes the effects of *PAPP-A* overexpression.

Alternatively, the MDA-MB-468 cell line may not be dependent on the IGF-signaling pathway for cell growth and survival. To address the first hypothesis, I purpose to examine *STC2* levels in cells with and without *PAPP-A* induction, as perhaps compensatory overexpression of *STC2* was initiated. To address the second hypothesis, *PAPP-A* expression could be reduced in a TNBC cell line with high endogenous expression, and its effect on cell growth *in vivo* could be measured, with the hypothesis that this should reduce carcinoma growth.

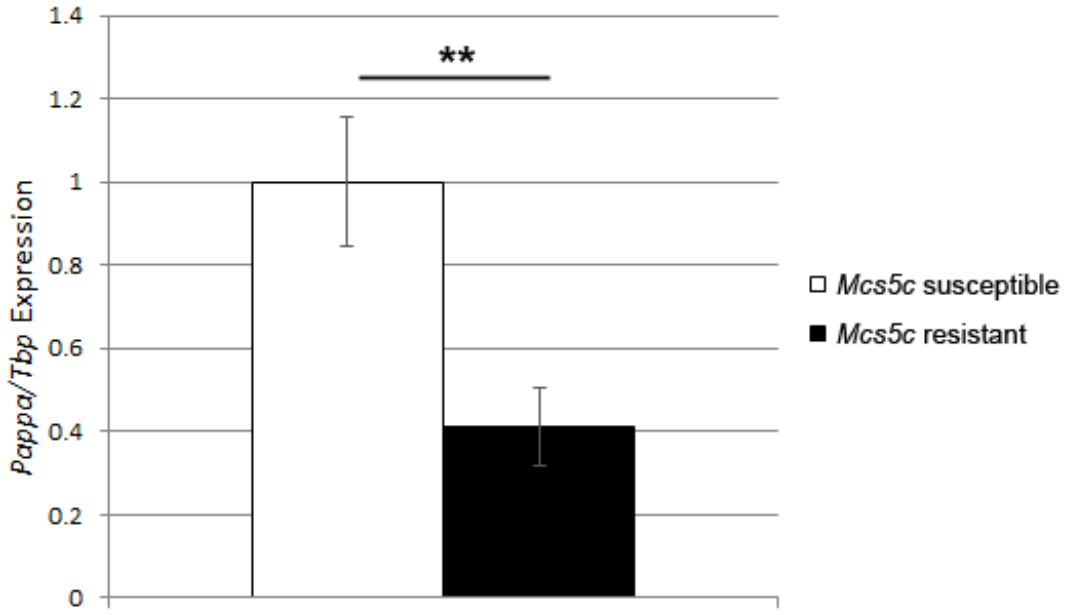


Fig 1: *Pappa* expression is increased in *Mcs5c* susceptible mammary carcinomas. *Pappa* expression was examined in mammary carcinomas from *Mcs5c* susceptible (n = 17) and *Mcs5c* resistant (n = 19) rats 15 weeks after DMBA administration. Gene expression relative to *Mcs5c* susceptible levels was determined via qPCR and standardized to *Tbp* expression. P-values were obtained using the non-parametric Mann-Whitney U test (*, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$).

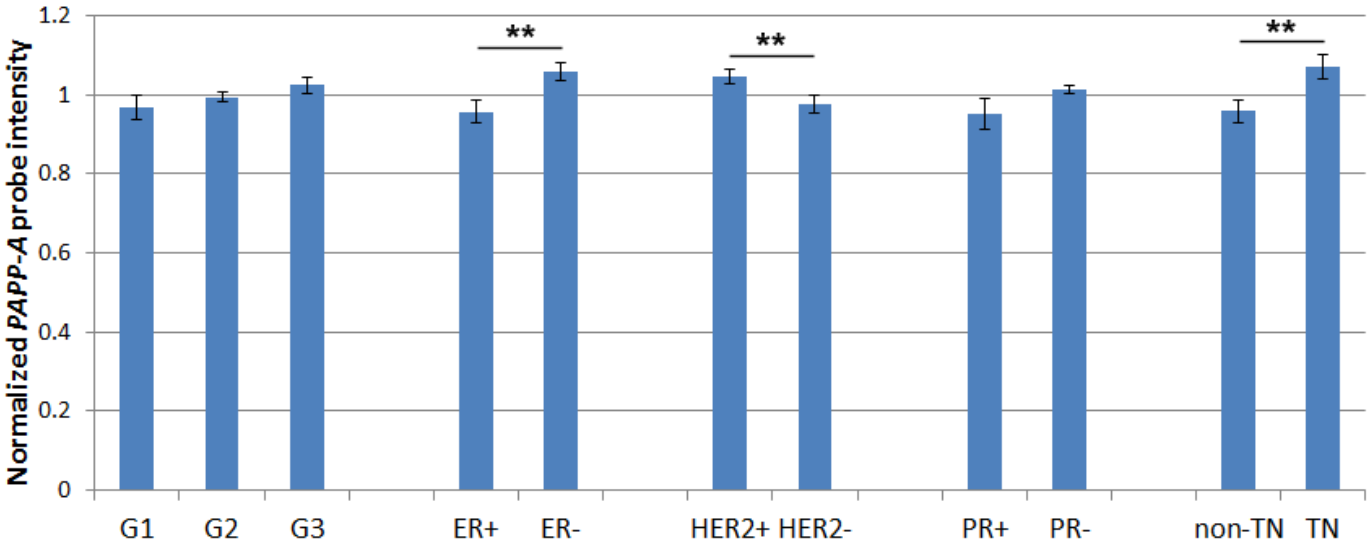


Fig 2: Oncomine analysis of *PAPP-A* expression in human breast tumors. Microarray datasets with >120 samples were analyzed for expression differences based on various clinical subtypes (G = tumor grade, ER = estrogen receptor, HER2 = human epidermal growth factor receptor 2, PR = progesterone receptor, TN = triple-negative). The number of studies analyzed is listed below each category, along with the number of tumors included in the studies. Statistics were performed using the Kruskal-Wallis rank test in R (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$).

Additional future directions focus on the use of the *Mcs5c* congenic rat model, or its target gene *Pappa*, in a prevention and therapeutic setting. I believe that this model will be useful for functionally characterizing how environmental factors may affect breast cancer risk by influencing genetic x WOS (G x WOS) interactions. There is growing concern over the impact adolescent exposure to a broad range of environmental factors, including early-life stress (ELS), diet, phytoestrogens and other chemicals, may have on long-term breast cancer risk [9]. The effect aWOS-specific exposure of these factors may have on mammary cancer risk could be investigated in the *Mcs5c* model, and exposure-specific alterations in *Mcs5c* mechanistic activity could be investigated. In this way, we could begin to understand molecularly how early exposure to environmental risk factors affects a specific pathway related to breast cancer risk. Understanding the interaction of environmental factors on G x WOS interactions may be crucial for the identification of window-specific risk variants. To date, no breast cancer-associated SNP has been identified in *Mcs5c* or the *Pappa* gene through GWAS. However, this may be due to the requirement of an initiating environmental exposure during childhood that then affects the risk of a select group of risk-

variant harboring women. *Mcs5c* or *Pappa*-associated SNPs may therefore be missed in GWAS that do not account for this type of exposure. This thesis work showing the effect that WOS can have on genetic risk factor function will hopefully highlight the need for these types of interactions to be considered in the context of risk assessment.

The *Mcs1a* locus:

Overexpression of the *Mcs1a* target gene *Nr2f1/NR2F1* results in decreased TNBC cell growth *in vivo*.

The human TNBC cell line MDA-MB-468 was identified as having low endogenous *NR2F1* expression levels. Inducible overexpression of *NR2F1* in these cells resulted in a decrease in final carcinoma volume and growth rate, suggesting that the NR2F1 signaling pathway is a relevant therapeutic target for TNBC therapy.

Many differentially expressed genes resulting from *NR2F1* overexpression in TNBC cells are associated with a luminal-like molecular profile. This is consistent with *Mcs1a*-associated promotion of a luminal phenotype in MECs and may explain the therapeutic benefits of *NR2F1* expression in TNBC.

FOXA1 is a well-characterized driver of luminal breast cancers and a critical component of normal luminal breast development. Its upregulation following *NR2F1* overexpression, and the downregulation of triple-negative/basal-associated genes, supports earlier work correlating *Mcs1a* activity to luminal MEC differentiation. As luminal breast cancers tend to be less aggressive than TNBC, it further suggests that the promotion of a luminal-like profile may explain *NR2F1*-associated TNBC tumor reduction.

Network modeling represents a useful tool in the identification of regulatory pathways of therapeutic interest.

Our modeling strategy incorporated numerous regulatory interaction data to identify possible pathways by which NR2F1 is mediating differentially expressed gene changes. Through this modeling, we were able to identify FOXA1, RXRA, RARA, and FOS as important regulatory nodes. FOXA1 and the retinoic acid pathways, in particular, represent therapeutic areas that may be targeted through combination treatment strategies. These strategies can be assessed in the future using our *NR2F1*-inducible model.

In this work, I have demonstrated the therapeutic relevance of the *Mcs1a* target gene *NR2F1* in TNBC. The analysis of differentially expressed (DE) genes resulting from *NR2F1* overexpression and NR2F1-based regulatory mapping has identified a number of genes, pathways, and strategies of future interest for targeted and combination therapies. Most intriguing of these include targeting of the vacuolar-ATPase complex, manipulation of the retinoic acid signaling pathway, and the promotion of a luminal molecular profile, or "luminalization", of triple-negative cells via NR2F1 and FOXA1 signaling. The tetracycline (tet)-inducible TNBC cell line system engineered during the course of this work represents a valuable *in vitro* model to begin to test these therapeutic strategies. However, its ability to recapitulate *in vivo* results needs to be tested. Initial *in vitro* gene expression experiments show that 48 hours of exposure to tet-containing media was not capable of recapitulating the majority of expression changes seen at necropsy in xenograft carcinomas (Fig 3; compare to Chapter 3 – Fig 3). Specifically, only *NR2F1* and *CHI3L1* show significant tet-dependent expression differences. I purpose to repeat these experiments with a longer time period of tet exposure, and hypothesize that this will result in a better recapitulation of *in vivo* expression changes. Additionally, I would like to verify that protein level

changes can be detected for DE genes of interest, particularly for secreted DE genes, as they would represent a readout of NR2F1-signalling that could be scaled up for high-throughput screening purposes.

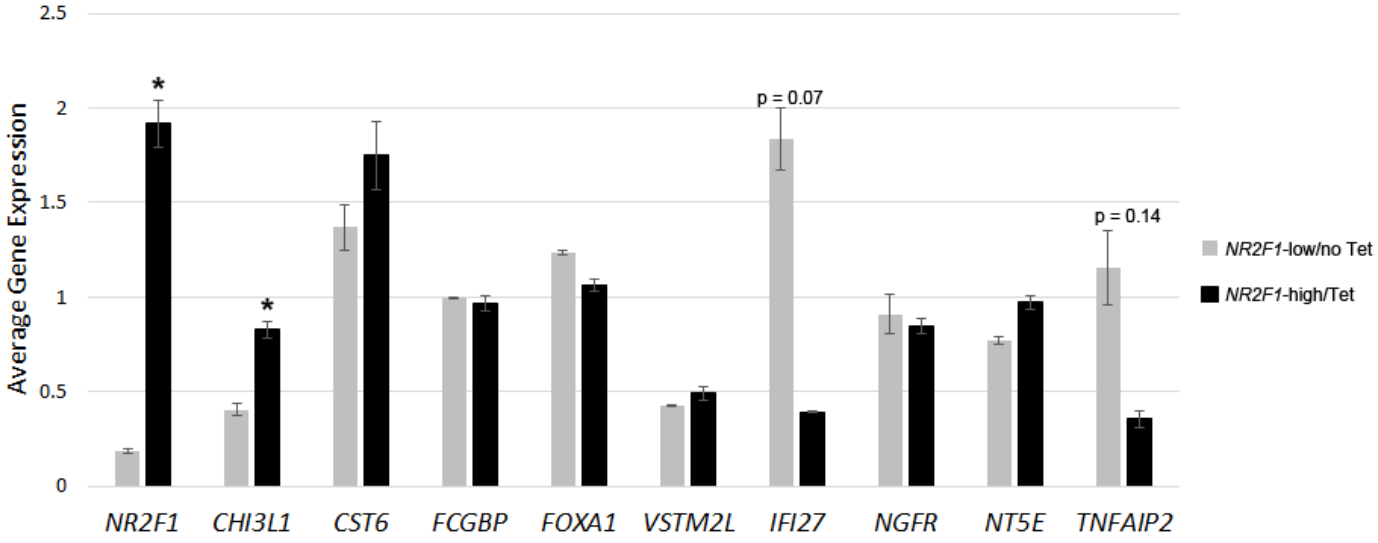


Fig 3: qPCR validation of DE genes *in vitro*. The same set of DE genes selected for verification through qPRC *in vivo* were analyzed *in vitro*. Two separate plates of cells were exposed to media containing 1µg/mL tetracycline (*NR2F1*-high/Tet, n = 2) for 48 hours or mock treatment (*NR2F1*-low/no Tet, n =2). Cell were collected and RNA extracted for qPCR expression analysis. Relative gene expression levels were standardized to the reference gene *PPIA*, and standard error bars are shown. P-values were obtained using a two-sided t-test (*, P < 0.05; **, P < 0.01; ***, P < 0.001).

If *NR2F1*-mediated expression changes could be modeled *in vitro*, it would make for a simplified system with which to test out therapeutic candidates before initiating *in vivo* studies. If not, these types of studies would be undertaken directly through xenograft experiments that would mimic those presented in Chapter 3. I would specifically like to test out whether overexpression of *NR2F1* in TNBC cells has sensitized cells to treatment with existing therapeutic agents. *In vitro*, this would entail treating cells with a combination of tet (for a pre-determined length of time that would initiate expression changes observed *in vivo*) followed by treatment with retinoic acid (RA). The anti-cancer effects of RA have been shown by Lin *et al* [10] to be mediated by NR2F1- and RARA-dependent signaling, and our network modeling suggests that RARA is a central regulator of numerous *NR2F1*-dependent DE genes. I hypothesize that *NR2F1* overexpression would make cells more sensitive to RA treatment and may be an example of a possible combination treatment strategy for TNBC. Additionally, work performed by

Marcotte and colleagues [11] that correlated gene essentiality with drug sensitivity data from Daemen *et al* [12], suggests additional cancer drugs that *NR2F1*-high cells may be sensitive to due to *NR2F1*-mediated expression changes. In this analysis, a gene was considered positively correlated to drug X if it was essential to the survival of a cell line and that cell line was also sensitive to drug X. This would suggest that the gene in question may be a direct target or in the targeted pathway of drug X, and high expression would be ideal with use of such a drug. A gene was considered negatively correlated to drug X if it was essential to the survival of a cell line, but that cell line was resistant to drug X. This would suggest that the gene may be important for drug resistance, and you would want low expression of the gene before using drug X. I examined the correlation between gene essentiality of top DE genes of interest and drug resistance data (Table 1). An initial analysis suggests that the PI3K/mTOR class of drugs may be more effective in *NR2F1*-high 468 cells, as we have increased expression of genes that are positively correlated with drug sensitivity (and may represent drug-targeted pathways) and decreased the expression of negatively correlated genes (that may confer resistance). In general, these combination therapies would be exciting treatment strategies to test out in either an *in vitro* or *in vivo* model.

Table 1: Correlation of gene essentiality and drug sensitivity for *NR2F1* -dependent DE genes

Upreg genes	**Class 1 (PI3K/mTOR)**	**Class 2 (EGFR/MEK/ERK)**	**Class 3 (AURORA/CDK)**	**Class 4 (Mixed)**	**Class 5 (DSB/Taxanes)**
CHI3L1	POS	-	-	-	NEG
FCGBP	-	POS	POS	-	-
FOXA1	POS	NEG	NEG	-	NEG
SERPINA5	POS	-	-	POS	-
NPNT	POS	-	NEG	-	NEG
Downreg genes					
IFI27	-	-	POS	-	POS
MAOA	NEG	-	POS	POS	-
NGFR	NEG	-	-	-	-
MFI2	NEG	-	-	NEG	POS
ITGB2	-	-	NEG	POS	-

POS = positive correlation between gene and drug sensitivity; NEG = negative correlation between gene and drug sensitivity

Additionally, I would like to continue to work on refining the NR2F1-based network model. Currently, a major limitation of our approach is that much of the protein-DNA interactions were obtained from an ER+ breast cancer cell line [13]. Although the literature-based weighting method used for node integration (GADGET) was TNBC-specific and therefore helped to downplay ER-dependent interactions, ideally, I would like to perform ChIP-seq experiments in our *NR2F1*-high 468 cells. From our initial network model, NR2F1, FOXA1, RARA, FOS, and RXRA appear to be central gene regulatory proteins whose binding sites following *NR2F1*-overexpression would be of interest. Binding site verification for these critical factors in a TNBC cell line would be essential for refinement of the model. Altogether, this thesis work highlights the relevance of the NR2F1 signaling pathway as a target in the treatment of TNBC. As this clinical class currently lacks effective targeted therapies, the identification of new combination strategies and targets is a welcome addition in the fight against this particularly aggressive form of the disease.

References

1. Russo IH, Russo J. Developmental stage of the rat mammary gland as determinant of its susceptibility to 7,12-dimethylbenz[a]anthracene. J Natl Cancer Inst 1978;61:1439–49.

2. Russo J, Wilgus G, Russo IH. Susceptibility of the Mammary Gland to Carcinogenesis. Am J Pathol 1979;96:721–33.

3. Schmidt D, Schwalie PC, Wilson MD, Ballester B, Gonçalves A, Kutter C, et al. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. Cell 2012;148:335–48. doi:10.1016/j.cell.2011.11.058.

4. Nordhoff E, Krogsdam AM, Jorgensen HF, Kallipolitis BH, Clark BF, Roepstorff P, et al. Rapid identification of DNA-binding proteins by mass spectrometry. Nat Biotechnol 1999;17:884–8. doi:10.1038/12873.

5. Smits BMG, Haag JD, Rissman AI, Sharma D, Tran A, Schoenborn AA, et al. The Gene Desert Mammary Carcinoma Susceptibility Locus Mcs1a Regulates Nr2f1 Modifying Mammary Epithelial Cell Differentiation and Proliferation. PLoS Genet 2013;9. doi:10.1371/journal.pgen.1003549.

6. Boldt HB, Conover CA. Overexpression of pregnancy-associated plasma protein-A in ovarian

cancer cells promotes tumor growth in vivo. Endocrinology 2011;152:1470–8. doi:10.1210/en.2010-1095.

7. Pan H, Hanada S, Zhao J, Mao L, Ma MZQ. Protein Secretion Is Required for Pregnancy-Associated Plasma Protein-A to Promote Lung Cancer Growth In Vivo. PLoS One 2012;7. doi:10.1371/journal.pone.0048799.

8. Jepsen MR, Kløverpris S, Mikkelsen JH, Pedersen JH, Füchtbauer E-M, Laursen LS, et al. Stanniocalcin-2 inhibits mammalian growth by proteolytic inhibition of the insulin-like growth factor axis. J Biol Chem 2015;290:3430–9. doi:10.1074/jbc.M114.611665.

9. IBCERCC. Breast Cancer and the Environment. Prioritizing Prevention. 2013. Available: http://goo.gl/N9AUre.

10. Lin B, Chen GQ, Xiao D, Kolluri SK, Cao X, Su H, et al. Orphan receptor COUP-TF is required for induction of retinoic acid receptor beta, growth inhibition, and apoptosis by retinoic acid in cancer cells. Mol Cell Biol 2000;20:957–70.

11. Marcotte R, Sayad A, Brown KR, Sanchez-Garcia F, Reimand J, Haider M, et al. Functional Genomic Landscape of Human Breast Cancer Drivers, Vulnerabilities, and Resistance. Cell 2016;164:293–309. doi:10.1016/j.cell.2015.11.062.

12. Daemen A, Griffith OL, Heiser LM, Wang NJ, Enache OM, Sanborn Z, et al. Modeling precision treatment of breast cancer. Genome Biol 2013;14:R110. doi:10.1186/gb-2013-14-10-r110.

13. Kittler R, Zhou J, Hua S, Ma L, Liu Y, Pendleton E, et al. A Comprehensive Nuclear Receptor Network for Breast Cancer Cells. Cell Rep 2013;3:538–51. doi:10.1016/j.celrep.2013.01.004.