5-2013

Genetic mapping and mechanism of action of rat mammary carcinoma susceptibility quantitative trait locus Mcs1b.

Aaron D. denDekker
University of Louisville

GENETIC MAPPING AND MECHANISM OF ACTION OF RAT MAMMARY CARCINOMA SUSCEPTIBILITY QUANTITATIVE TRAIT LOCUS *MCS1B*

By

Aaron D. denDekker
B.S. University of Louisville, 2001
M.S. University of Louisville, 2012

A Dissertation
Submitted to the Faculty of the
School of Medicine of the University of Louisville
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

Department of Biochemistry and Molecular Biology
University of Louisville
Louisville, KY

May 2013

GENETIC MAPPING AND MECHANISM OF ACTION OF RAT MAMMARY CARCINOMA SUSCEPTIBILITY QUANTITATIVE TRAIT LOCUS *MCS1B*

By

Aaron David denDekker
B.S. University of Louisville, 2001
M.S. University of Louisville, 2012

A Dissertation Approved on

January 9, 2013

by the following Dissertation Committee:

David Samuelson, Dissertation Director

Ronald Gregg

Carolyn M. Klinge

David Hein

Russell A. Prough

DEDICATION

This dissertation is dedicated to my friend and colleague

Michael Wayne Gordon

who was taken from this world too soon.

Vita mutatur, non tollitur.

ACKNOWLEDGEMENTS

I would like to express gratitude to all of the people who helped me along my way, both academically and personally. I would like to thank my mentor, Dr. David Samuelson, for his patience and guidance. As his first graduate student there were some growing pains for us both, but his door was always open and I've learned so much from him over the past few years. I want to thank my dissertation committee, Drs. Carolyn Klinge, Russell Prough, Ronald Gregg, and David Hein, for their support and assistance. I would like to extend a special thanks to Ron, as well as Dr. Jill Suttles, for giving me the opportunity to work in their respective labs prior to entering graduate school. The skills and lessons I learned from them helped prepare me for what was to come. I would like to thank the department of Biochemistry and Molecular Biology and its faculty for affording me the opportunity to pursue this degree. I would like to especially thank Drs. William Dean, Steven Ellis, and Barbara Clark for their friendship and stewardship through the years. You were all also such excellent teachers and had a huge impact on my growth and development as a scientist as well as a person.

I want to thank Joe Moore for his friendship and support. Without him graduate school would not have been the same. I want to thank Traci Kruer for being my veep, helping me format my dissertation, and just listening and being a good friend... I really appreciate everything you've done for me. I want to thank Alan Kerr and Huy Le for their thought-provoking and informed discussions and for just being great friends. I want to give special recognition to Mike Gordon. Mike, you were so bright and such a good

person and you had such potential to do great things in the world. I am lucky to have had you as such a close friend and confidant. I miss you. You will always live in our hearts. I would also like to thank all of my other fellow graduate students, former and current, for their support and willingness to listen to my presentations.

Last, but not least, I want to thank my lovely wife, Tara, for her unwavering support and encouragement through everything. She was working full-time and working on her own graduate degree in nursing, but still found the strength and the words to help keep me going when I didn't think I could go any further. I love her and my wonderful children, Paul and Bowie, more than I can express. I could not have done any of this without all of them. They make it all so much more worth it.

ABSTRACT

GENETIC MAPPING AND MECHANISM OF ACTION OF RAT MAMMARY CARCINOMA SUSCEPTIBILITY QUANTITATIVE TRAIT LOCUS *MCS1B*

Aaron D. denDekker

January 9, 2013

Breast cancer is a complex disease that involves genetic, epigenetic, and environmental components. High and moderate penetrant genes have been identified that affect risk to developing breast cancer; however, these risk alleles are present in a small percentage of breast cancer cases. Low penetrant modifier genes have risk-associated alleles that are common in the population. Although these genes have lower penetrance, it is expected that the majority of genetic risk to developing breast cancer is controlled by common genetic variation. Studying mechanisms of common genetic variants on breast cancer risk is difficult due to their small individual effects and overlapping contribution of other risk factors; thus, animal models are commonly used. The rat mammary carcinoma susceptibility quantitative trait locus (QTL) *Mcs1b* was identified between mammary carcinoma-resistant Copenhagen (COP) and susceptible Wistar Furth (WF) rats on chromosome *2*. This rat QTL is an ortholog of a human breast cancer-associated locus identified on human chromosome *5q*; therefore, the rat *Mcs1b* model can be used to identify mechanisms and causative factors contributing to breast cancer risk associated with human breast cancer-associated locus *5q*.

The goal of the work presented in this dissertation is to identify quality candidate breast cancer risk genetic elements associated with the rat *Mcs1b* locus. This project utilized a well-defined rat mammary carcinogenesis system and congenic rat model to fine map and characterize the rat *Mcs1b* locus. My studies reduced the number of candidate genes by narrowing the rat *Mcs1b* locus from a 13 megabase (Mb) to a 1 Mb containing nine annotated transcripts. I determined that *Mcs1b*-conferred mammary carcinoma resistance is being controlled by a cell type within the mammary gland. This is an important finding because mammary carcinogenesis is dependent on both mammary gland-extrinsic and -intrinsic factors. I also found that the transcript *Mier3* is differentially expressed between resistant and susceptible rat mammary glands with or without carcinogen exposure providing genetic evidence that *Mier3* is a strong mammary carcinoma susceptibility gene. Taken together, these results provide insight into the mechanism by which *Mier3* controls mammary carcinogenesis and implicate human *MIER3* as a potential target for breast cancer prevention.

TABLE OF CONTENTS

PAGE

LIST OF TABLES

TABLE

LIST OF FIGURES

FIGURE

CHAPTER I

GENES, ENVIRONMENT, AND BREAST CANCER

General Introduction

Breast Cancer Statistics

Breast cancer is the leading cause of cancer death in the United States in women aged 20-59 and is the second most diagnosed cancer in women in the US (Siegel et al., 2012). Additionally, with the exception to skin cancer, breast cancer is the most common cancer diagnosed in women worldwide (Mathers et al., 2008). In the US, it was estimated that near 230,000 new cases of invasive breast cancer were diagnosed and that almost 40,000 women died of breast cancer in 2012 (American Cancer Society, 2011). Breast cancer deaths in women rose 0.4% per year from 1975 to 1990. However, in recent years, early detection, increased awareness, better treatments, and decreased use of hormone replacement therapies have led to a decrease in breast cancer death rates with a 2.2% decrease per year from 1990 to 2007 (American Cancer Society, 2011). Still, breast cancer remains a major health concern for many women and current risk estimates are that 1 out of 8 women in the US will develop breast cancer (Altekruse SF, 2010).

Breast cancer also affects men; however, male breast cancer only accounts for approximately 1% of breast cancers diagnosed in the US (American Cancer Society, 2011). Due to its rarity, much less is known about the male form of breast cancer. Incidence of male breast cancer has risen from 1 in 100,000 men being diagnosed with

breast cancer in the late 1970's to 1.2 in 100,000 from 2000-2004 (Onami et al., 2010). However, breast cancer death rates in men have fallen 3.3% since 2000; likely due to increased awareness and better diagnoses (American Cancer Society, 2011; Anderson and Devesa, 2005).

Breast Cancer Risk

Breast cancer is a complex disease and the risk of developing it has environmental, genetic, and epigenetic components. The most important factors affecting risk to developing breast cancer are age and female gender (American Cancer Society, 2011). As noted, approximately 12% (1 in 8) of US women are predicted to be diagnosed with breast cancer based on lifetime risk, and this probability increases as a woman gets older (Altekruse SF, 2010; American Cancer Society, 2011). This estimate is based on epidemiology of population incidence. However, these estimates may not be accurate for many individual women. Risk for an individual may be higher or lower depending on various risk factors including age, family history and reproductive history as well as other heritable and non-heritable factors (Gail et al., 1989).

Radiation exposure is one of the most potent exogenous factors known to increase chances for developing breast cancer (Land et al., 2003; Ronckers et al., 2005). Although exposure to high doses of ionizing radiation is less common, many women are exposed to low-dose radiation through mammograms and routine low-dose exposure has been associated with an increase in breast cancer risk especially in women already with an increased familial risk (Pijpe et al., 2012). Additionally, other non-heritable risk factors have been identified e.g., living near nuclear power sites (chronic radiation exposure) and

shift work cycles (prolonged exposure to light at night) but these risks aren't fully understood (Boice et al., 2003; Bonde et al., 2012; Hill et al., 2011).

Inherited genetic susceptibility to cancer is now considered an established fact (Fletcher and Houlston, 2010). Mutations in several high-risk genes have been identified, e.g. *BRCA1* and *BRCA2*, that increase a woman's chances of developing breast cancer by 51-75% and 33-54%, respectively (Antoniou et al., 2003). Screening for mutations in these genes is, under certain circumstances, more routinely conducted; however, mutations in these genes account for approximately 3-5% of female breast cancer cases, leaving the majority of alleles contributing to breast cancer susceptibility unknown (Campeau et al., 2008; Narod and Salmena, 2011). Conversely, common genetic modifiers have a small independent effect but can act cumulatively to exert a greater influence on disease development (Jostins and Barrett, 2011; Pharoah et al., 2008). Therefore, the majority of genetic risk to breast cancer development is likely attributable to common genetic variation (Fletcher and Houlston, 2010; Lee et al., 2011).

Predicting Breast Cancer Risk

Accurate prediction of an individual's breast cancer risk is vital to developing better prevention and treatment strategies. The ability to predict the development of disease in an individual has been useful in improving strategies for breast cancer prevention; e.g. increased frequency of mammograms for high risk breast cancer patients (Jostins and Barrett, 2011). However, breast cancer etiology is complex, which makes accurate individual risk assessment problematic for many women. Individual risk for developing breast cancer is currently assessed using the Gail model. Gail and colleagues

developed a risk-assessment model based on a set of risk factors: age at menarche, age at first live birth, number of previous biopsies, and number of first-degree relatives with breast cancer (Costantino et al., 1999; Gail et al., 1989). An interactive tool has been developed by the National Cancer Institute (NCI) and the National Surgical Adjuvant Breast and Bowel Project (NSABP) based on the Gail model. This tool is used by physicians to assess risk for an individual with a limited family history (http://www.cancer.gov/bcrisktool/).

The Gail model provides an efficient method to project the likelihood to develop breast cancer for most individual women; however, it has limitations. First, the Gail model only accounts for first degree relatives with breast cancer potentially underestimating risk in the 50% of families with cancer in the paternal lineage (Euhus et al., 2002; Evans and Howell, 2007). Also, the Gail model does not take into account the age of onset of breast cancer of the affected relative, thereby, possibly overestimating risk in women with an affected relative who developed breast cancer late in life. Last, although it has been modified to account for ethnic background, the Gail model focuses primarily on non-genetic risk factors. Heritable factors are recognized to have significant roles in complex disease risk and this is underscored by the importance of accounting for family history in risk assessment (Bevier et al., 2011; Costantino et al., 1999; Gail et al., 1989; Lalloo and Evans, 2012).

At present, the effects of risk-predisposing genes are difficult to evaluate; thus, a better understanding of genetics and the molecular mechanisms that influence breast cancer susceptibility is necessary to better predict an individual's risk for developing breast cancer. Several low- to moderate-penetrance breast cancer risk alleles have been

identified (Easton et al., 2007; Turnbull et al., 2010). Relative risk conferred by alleles at individual loci is small, but risk alleles are hypothesized to act multiplicatively. It has been estimated that risk of developing breast cancer is approximately six times as great among women carrying 14 risk alleles as among those carrying no risk alleles at these loci (Pharoah et al., 2008). Although there is little clinical use for single, low-penetrance genes, the cumulative effects of these alleles may be useful to separate high risk individual women from those at lower risk (Pharoah et al., 2008). As previously mentioned, the genetic contribution to breast cancer development is not fully understood. However, new technologies and approaches are available that will allow for the discovery of common genetic risk alleles and it is believed that assessing individual breast cancer risk based on genetic factors will be achieved in the near future (Jostins and Barrett, 2011).

Breast Cancer Risk Factors: Breast Cancer as a Complex Disease

Estrogen Exposure and Breast Cancer

As stated earlier, the most widely recognized risk factors to breast cancer development are female gender and increasing age (American Cancer Society, 2011). Women develop breast cancer at a rate one hundred times that of men and these rates increase as women age. This is interpreted as a representation of accumulated exposure to ovarian hormones in the form of estrogens and progesterone (Pike et al., 1993). It is thought that women exposed longer to estrogen due to early menarche, late menopause or hormone replacement therapy exhibit an increase in breast cancer risk (Kelsey et al., 1993; Pike et al., 1993). It was first discovered in 1896 that oophrectomy could

effectively regress breast carcinomas (Beatson, 1896). Later observations indicated that women who experienced early menarche or late menopause exhibited higher incidence of breast cancer (Kelsey et al., 1993). Also, women who had their first child early (before age 18) had a lower incidence of breast cancer than women who had children later in life (35 years of age) (Kelsey et al., 1993; MacMahon et al., 1970; Pike et al., 1983). More recently, it has been shown that oophrectomy in women before 35 years of age reduces breast cancer risk by 75% (Kronenberg and Williams, 2008). Taken together, this suggests a pivotal role for endogenous estrogens in the development of breast cancer.

Estrogen exposure is now considered one of the most important factors determining breast cancer risk. Estrogens primarily function to control the estrous cycle but have roles in a variety of other processes. There are three forms of estrogen produced in the body: estrone (E1), estradiol (E2), and estriol (E3). These are made and secreted primarily by the ovary but can also be made in other tissues. Of these E2 is the most potent as E1 and E3 bind estrogen receptors α and β (ERα and ERβ) with lower affinity compared to E2. However, E1 and E3 can be converted to E2. Significantly, E2 is secreted by the breast epithelium and adipose and serum E2 levels have been correlated with an increase in breast cancer incidence (Kronenberg and Williams, 2008). During the 1940s, hormone replacement therapy (HRT) was instituted into clinical practice for the treatment of menopausal symptoms by administration of exogenous estrogens. Subsequent studies of groups of women receiving HRT demonstrated that the relative risk (RR) of developing breast cancer for women receiving therapy was elevated (RR=2.0) compared to that of the general population (RR=1.3) (Hoover et al., 1976). Breast cancer risk due to HRT was debated for many years citing contradictory studies;

however, studies conducted by the Women's Health Initiative (WHI) and the Million Women Study (MWS) indicate that continued HRT increases breast cancer risk (Beral, 2003; Rossouw et al., 2002). Data from the MWS showed the effect of combined progesterone-estrogen therapy was greater (RR=2.00, [95% Confidence Interval (CI) 1.88-2.12], P<0.0001) compared to estrogen alone (1.30 [1.21-1.40], P<0.0001) when compared to women never receiving HRT (Beral, 2003). Women in the WHI randomized control study receiving combined estrogen-progesterone therapy also had an increase incidence in breast cancer compared to those receiving a placebo (Hazard Ratio (HR)=1.62 [95% CI 1.00-1.59], P<0.05) (Rossouw et al., 2002).

Estrogens have been extensively studied to understand the action by which they influence breast cancer. Estrogens promote rapid proliferation of mammary epithelial cells, which increases the probability that mutations will become fixed and propagated, thereby promoting tumor formation (Preston-Martin et al., 1993). The "canonical" estrogen signaling pathway occurs through estrogens binding ERα and ERβ. Estrogens diffuse passively through cell and nuclear membranes and bind to ERα and ERβ (Kronenberg and Williams, 2008). Once an estrogen receptor is bound to an estrogen ligand, it undergoes a conformational change and binds to specific DNA sequences called estrogen response elements (EREs) to drive transcription of target genes. ERE-bound ERs interact with basal transcription factors and co-activator proteins, which stabilize basal transcription factor binding and initiate transcription (Klinge, 2000). The products of these genes act to promote cell growth and differentiation.

ERα and ERβ exhibit similar binding affinities for E2, the predominate estrogen in premenopausal women, and both ER subtypes have been shown to bind to EREs

similarly (Kuiper et al., 1997; Paech et al., 1997). Both receptors form homodimers in response to ligand binding; however, studies have shown that they can also form heterodimers *in vitro* and *in vivo*, and that ERβ reduces ERα transcriptional activity (Hall and McDonnell, 1999; Pettersson et al., 2000; Pettersson et al., 1997). Also, recent findings have shown that ERβ can recruit a corepressor complex to the ERα gene promoter resulting in reduced expression of ERα (Bartella et al., 2012). This implies that ERβ is acting as an internal regulator of ERα activity. Indeed, ERα expression is often higher compared to ERβ in invasive mammary tumors (Leygue et al., 1998).

ERs are the active components controlling estrogen signaling and have, thus, become popular targets for breast cancer treatment. There are several classes of hormone treatment options for breast cancer therapy. Selective estrogen-receptor modulators (SERMs), such as tamoxifen and raloxifene, are synthetic agonist/antagonists of the ER. In the mammary gland, both of these drugs bind ER as antagonists to prevent ER-mediated transcription. Alternatively, tamoxifen is an ER agonist in uterus and bone while raloxifene acts as an agonist in bone only (Dutertre and Smith, 2000). Other chemotherapeutics also act on the ER, such as fulvestrant, an ER antagonist that abolishes estrogen-specific gene transcription by degrading the ER (Flemming et al., 2009). Aromatase inhibitors (AI), on the other hand, operate indirectly by inhibiting the activity of aromatase, the enzyme responsible for converting androgens into estrogens, thereby reducing estrogen availability for ER binding (Mokbel, 2002).

Tamoxifen is the most widely used treatment for ER-positive breast cancers in pre- and post-menopausal women. Although tamoxifen is an effective therapeutic, it has some disadvantages and limitations. For one, tamoxifen has been shown to increase the

chance of endometrial cancer by acting as an agonist in the uterus and endometrium (Fisher et al., 1994; Gottardis et al., 1988). Also, many breast tumors are unresponsive to tamoxifen as they have lost the ability to express ERα (Ring and Dowsett, 2004). One of the biggest problems clinicians face when treating with tamoxifen is that many patients develop resistance to the drug and relapse (Cui et al., 2012; Osborne et al., 2005). Often, tumors in these patients continue to express a functionally normal ER yet they grow independently of estrogen action. Amplification and over-expression of the growth factor receptor HER2 is thought to be a major mechanism contributing to endocrine resistance in many cases (Osborne et al., 2005). Additionally, it is believed that tamoxifen resistance in some breast cancers may is due to changes in expression of co-activators and co-repressors (Dobrzycka et al., 2003). For example, the co-repressor Metastatic tumor antigen 1 (MTA1) interacts directly with ER and histone deacetylases (HDACs) to inhibit ERE transcriptional activity and promote hormone-independent growth. Further, MTA1 over-expression correlates with a reduced response to tamoxifen (O'Malley and Kumar, 2009). In addition, mis-expression of many of these co-regulators is associated with many cancer types. It is not clear whether this dysregulation is a cause or consequence of the pathology; however, co-regulators of ER action appear to be important factors in breast cancer pathogenesis.

Estrogens also function to promote breast carcinogenesis independent of the canonical ER-signaling pathway. Several mechanisms have been identified that are independent of estrogen receptor activity. For example, Liehr *et al.* have shown that 4-hydroxylated estrogen metabolites play a central role in the genotoxic activity of estrogen via generation of free radicals, which correlates with increased cancer risk (Liehr, 1990;

Liehr and Roy, 1990). Second, estrogen metabolites, catechol estrogen quinones, are formed via cytochrome P450 activation. These electrophilic quinine metabolites form depurinating adducts on DNA thereby inducing point mutations (Cavalieri et al., 2002). Normally, metabolism of these catechol estrogens by *O*-methylation, glutathionation, glucuronidation, or sulfation renders these inactive; however, studies have shown that when levels of catechol estrogen metabolites are increased they cannot be sufficiently metabolically inactivated, and this associates with an increase in breast cancer incidence (Rogan et al., 2003). In addition, Barrett and colleagues demonstrated that diethylstilbestrol (DES), a synthetic estrogen previously used for the treatment of menopause, interferes with microtubule organization at low doses (Tsutsui et al., 1983). Their data indicate that DES does not prevent cell division but, rather, interferes with microtubule organization sufficiently to cause non-dysjunction at mitosis resulting in aneuploidy (Tsutsui et al., 1983). Therefore, endogenous estrogens may also be affecting carcinogenesis through this pathway.

Various genetic modifiers have been identified that affect estrogen action in the context of breast cancer susceptibility. Common variants have been identified in, *ESR2*, the gene encoding ERβ (Maguire et al., 2005). Gene targeting association studies along with *in vitro* studies suggest these variants may be involved in increasing risk for breast cancer. Interestingly, an ultra-rapid metabolizing allele for *CYP2C19*, *CYP2C19*17*, has been identified that associates with a lower risk for breast cancer (Justenhoven et al., 2009). Additionally, deletion alleles of *CYP2C19*, were identified at an increased frequency in familial breast cancer cases from a cohort of Northern Finnish women (Pylkas et al., 2012). *CYP2C19* is a gene encoding an enzyme involved in the catabolism

of E2 implying that increased catabolism of E2 reduces estrogen levels and subsequently reduces breast cancer risk. Correspondingly, haploinsufficiency of *CYP2C19* may retard estrogen catabolism and, thereby, increase risk for breast cancer. What is more, copy number variants (CNVs) have recently been discovered in familial breast cancer cases that affect genes in estrogen signaling pathways (Pylkas et al., 2012). Taken together, the effects of estrogen signaling in breast cancer development are very complex. Unraveling these mechanisms will be useful in determining the role of estrogens in breast cancer and may lead to better diagnoses and treatments.

Progesterone and Breast Cancer

Progesterone is another steroid hormone involved in mammary gland development (Kronenberg and Williams, 2008). Like estrogens, progesterone passively diffuses into the cell and binds to either of its cognate receptors, A or B (PR-A or PR-B, respectively) and is able to activate gene transcription (Kronenberg and Williams, 2008). It is believed that estrogens and progesterone act in concert to promote ductal branching and development of mammary terminal end buds (TEBs), while estrogen is responsible for overall growth of mammary ducts emanating from the nipple (Ruan et al., 2005; Singletary and McNary, 1992). Although progesterone is associated with mammary duct outgrowth, studies have also shown that PR-deficient animals develop normal mammary glands (Kleinberg et al., 1990; Lydon et al., 1995). This suggests that progesterone may be involved in ductal morphogenesis by acting through an alternative mechanism. Although the mechanism by which progesterone is acting is not fully understood, it is

held that progesterone plays a role in promoting mammary gland development (Lydon et al., 1995; Ruan et al., 2005).

Progesterone is also known to attenuate estrogen action *via* three separate mechanisms (Kronenberg and Williams, 2008). First, progesterone can reduce synthesis of ERα (Hsueh et al., 1976; Tseng and Gurpide, 1975). Second, progesterone can activate transcription of the gene coding for 17β-hydroxysteroid dehydrogenase (17β-HSD), an enzyme involved in the enzymatic inactivation of E2 to the weakly estrogenic E1 (Casey et al., 1994). Third, progesterone induces transcription of estrogen sulfotransferases; enzymes that catalyze the conjugation of sulfate groups to E2 and E1 rendering the estrogens inactive (Falany and Falany, 1996). Therefore, although progesterone promotes mammary gland development, it could play a role in reducing the deleterious effects of estrogen action.

Based on these observations, one may expect progesterone to reduce the pro-tumorigenic effects of estrogen. Not surprisingly, breast tumors that express ER and PR are much more sensitive to endocrine chemotherapies compared to ER- and PR-negative tumors (American Cancer Society, 2011). However, the role of progesterone in breast cancer risk is confounding. The WHI and the MWS Study found that women taking estrogen plus progestin replacement therapy exhibited higher breast cancer incidence than those taking estrogen alone (Chlebowski et al., 2010; Chlebowski et al., 2003; Travis et al., 2010). Additionally, postmenopausal women with prior hysterectomy that received equine conjugated estrogen alone had a lower incidence of breast cancer compared to the placebo group (Anderson et al., 2012). This implicates progesterone with a role in promoting breast cancer development.

The mechanism by which progesterone and PR affect mammary cell proliferation and cancer development is not fully understood. *In vitro* studies have shown that progesterone induces cell proliferation by activating protein kinases such as MAPK, Akt/PI3K, and c-Src (Boonyaratanakornkit et al., 2001; Migliaccio et al., 1998; Saitoh et al., 2005); however, the role this plays in human (i.e. whole animal) physiology is not clear and further studies are needed to determine how it effects mammary carcinogenesis. What is more, PR expression is driven by ERα but can also be expressed independently of ERα (Horwitz et al., 1982; Lange, 2008). Therefore, it is difficult to ascertain the effects of progesterone alone from those of estrogen (Lange, 2008). Regardless, while the distinct role of progesterone and PR in breast carcinogenesis is unclear, breast tumor PR status is routinely used as a prognostic indicator for response to adjuvant chemotherapy and PR-targeted therapies are in use to treat breast cancer (Lange, 2008).

Exogenous Environmental Effects on Breast Cancer Development

As described, breast cancer is influenced by both exogenous and endogenous environmental chemical carcinogens. The prime example of a physical carcinogen is irradiation. Less common than estrogen exposure, ionizing radiation is the strongest known exogenous environmental factor affecting breast cancer development (Ronckers et al., 2005). This is based on breast cancer incidence rates in atomic bomb survivors and patients receiving prolonged radiotherapy (Land et al., 2003; Ronckers et al., 2005). The effects of living near nuclear power plants and exposure to low-level ionizing radiation sources remains a concern, but has not been extensively studied (Boice et al., 2003). Unrelated to radiation, there has been increasing interest in "light at night" as a risk factor

since exposure to light at night due to night shift work correlates with both disruption in melatonin secretion and increased breast cancer incidence (Bonde et al., 2012; Hill et al., 2011). Also, exposure to radiofrequency/microwave radiation emissions (RF) from wireless communications has become a concern due to the increasing use of mobile communication devices. This is thought to play a role in numerous pathologies including breast cancer development; however, these areas are relatively new and are not fully understood (Hardell and Sage, 2008).

Many different classes of exogenous chemical carcinogens have been identified that associate with breast cancer risk. Some of these act though the estrogen signaling pathway. Xenoestrogens, such as polychlorinated biphenyls (PCBs), bisphenol A (BPA) and phthalates are structurally different from endogenous estrogens but are able to bind ERs and activate mitotic events (Darbre and Charles, 2010; Fernandez and Russo, 2010; Kester et al., 2000; Tsutsui et al., 2000). Phytoestrogens, such as genestein, are natural plant-derived xenostrogens. They are structurally similar to endogenous mammalian estrogens and are able to act in a similar manner through the canonical signaling pathway (Martin et al., 1978). Yet, reported effects on phytoestrogens on breast cancer risk are conflicting due to studies that have shown protective as well as detrimental effects (Darbre and Charles, 2010; Martin et al., 1978; Safe, 1997; Zhang and Chen, 2011).

Other chemical carcinogens are able to act to promote breast cancer development through estrogen-independent means. Some of the most prominent are polycyclic aromatic hydrocarbons (PAHs) and their nitrated metabolites (nitro-PAHs), which cause genotoxic effects (Mersch-Sundermann et al., 1993). For example, PAHs such as benzo(a)pyrene (B[a]P) and 7,12-dimethylbenz(a)anthracene (DMBA) are metabolized to

form reactive diol epoxide intermediates that directly bind DNA and induce mutations (Todorovic et al., 1997). B[a]P is one of the most well-studied PAHs and it is found in many places because it is a byproduct of incomplete combustion of organic matter, e.g., cigarette smoke, car exhaust, and industrial smoke stack exhaust (Lawther and Waller, 1976; Phillips, 1999). DMBA is a synthetic organ-specific PAH also used as a laboratory carcinogen (Christou et al., 1995; Miyata et al., 2001). DMBA induces expression of *CYP1B1* in a mammary gland-specific manner (Christou et al., 1987). Thus, CYP1B1, is the primary P450 enzyme metabolizing DMBA in the mammary gland. CYP1B1, along with epoxide hydrolase, converts DMBA to its carcinogenic diol-epoxide form, which makes it ideal to induce carcinogenesis in the mammary gland (Christou et al., 1987; Christou et al., 1995; Miyata et al., 1999). DMBA is routinely used as an experimental mutagen for research purposes because it models PAH-induced carcinogenesis (Modi et al., 2012).

Lifestyle-based Factors and Breast Cancer Risk

Many different lifestyle-based environmental factors are also contributors to breast cancer risk. Tobacco smoke has long been accepted as modifiable lifestyle risk factor for many different cancer types although there is limited evidence for its role in breast cancer (American Cancer Society, 2012). Obesity and alcohol consumption are two other factors that have been identified as risk predictors for developing breast cancer (American Cancer Society, 2011; Carmichael, 2006; Hankinson et al., 1995; Pelucchi et al., 2011).

In post-menopausal women, obesity is positively associated with an increased incidence of breast cancer and has also been shown to confer a poorer prognosis (Carmichael, 2006; Hankinson et al., 1995). However, studies on pre-menopausal women have had conflicting results as some data suggest an inverse relationship between weight and breast cancer outcome. Regardless, weight gain and obesity associate with poor outcomes in both pre- and post-menopausal women (Carmichael, 2006). Adipocytes exhibit increased aromatase expression, the enzyme responsible for converting androgens to estrogen (Bulun et al., 2012; Santen et al., 2009). Thus, it has been postulated that increased estrogen is produced in the breast adipose tissue thereby promoting breast cancer development (Bulun et al., 2012).

Alcohol consumption increases endogenous estrogen levels and positively correlates with breast density and breast cancer incidence (Boyd et al., 1995; Hankinson et al., 1995). Although alcohol consumption associates with increased incidence of breast and other types of cancers, there are also data that suggest beneficial effects from alcohol, especially in red wine (de Lorimier, 2000). However, the mechanisms for this have not been extensively studied and are currently inconclusive.

Additionally, the consumption of "well-done" or charred meat has been identified as a risk factor to a variety of cancers, including the breast (Zheng and Lee, 2009). Heterocyclic amines (HCAs) are the most abundant mutagen found in overcooked meat. Studies have identified associations between high-temperature cooked meat intake, HCA exposure and breast cancer incidence (De Stefani et al., 1997; Sinha et al., 2000; Zheng et al., 1998; Zheng and Lee, 2009). Women who consistently eat overcooked meat have a 4.6-fold higher RR (95% CI=1.36-15.70) of developing breast cancer compared to

women who consume rare or medium-done meat. (Zheng et al., 1998). Further studies found increased levels of 2-amino-1-methyl-6-phenylimidazo(4,5-b)pyridine (PhIP), the most common HCA found in over-cooked meat, to correlate with an increase in breast cancer incidence; however, there was no significant effect associated with other HCAs (Sinha et al., 2000). In addition to HCAs, overcooked meat may also contain other mutagens such as PAHs including the ubiquitous B[a]P (Zheng and Lee, 2009). Other studies suggest that breast cancer risk associated with HCA exposure may be modified by common polymorphisms in genes coding for enzymes responsible for metabolizing HCAs, thereby strengthening the role of genetics in breast cancer risk (Deitz et al., 2000; Zheng et al., 1999; Zheng et al., 2002).

Breast Cancer Genetics and Susceptibility Genes

The etiology of breast cancer is driven by multiple components that include environmental factors, physiological host factors, and inherited genetic components. Genes influencing complex diseases are inherited according to Mendelian principles. However, while Mendelian disorders are monogenic, complex diseases are controlled by multiple alleles (Badano and Katsanis, 2002). Genes associated with disease provide a genetic predisposition to development of the disease; however, the outcome is determined by gene-gene interactions and gene-environment interactions (Hunter, 2005; Marian, 2012). At present, the interplay between hereditary and environmental factors is not fully understood (Hunter, 2005). Although it remains difficult to quantify the magnitude of the effect of each of the components on breast cancer etiology, as outlined above, progress is being made to begin to estimate the contribution of the genetic components.

Women with a family history of breast cancer have an increased breast cancer risk (American Cancer Society, 2011; Bevier et al., 2011; Collaborative Group on Hormonal Factors in Breast Cancer, 2001). RR increases with increasing numbers of first degree relatives diagnosed with breast cancer with RR=1.80 (99% CI=1.69 - 1.91), 2.93 (2.36 - 3.64) and 3.90 (2.03 - 7.49) for one, two, or three affected first degree relatives, respectively (Collaborative Group on Hormonal Factors in Breast Cancer, 2001). Having an affected second-degree relative increases risk to a lesser degree, with RR=1.27 (95% CI=1.09 - 1.47) and 1.26 (1.05 - 1.50) for either a maternal or paternal grandmother affected, respectively. Further, there is a RR=1.60 (1.24 - 2.07) for two affected second degree female relatives (Bevier et al., 2011). Recent epidemiological studies have revealed that having a brother diagnosed with breast cancer increases a woman's RR to 2.48 (95% CI=1.44 - 4.27), which is more than having an affected sister (RR=1.87, 1.80-1.95). This may suggest that male breast cancer has a higher genetic basis than female breast cancer (Bevier et al., 2011). Clearly, the inherited component affecting breast cancer risk is significant. Moreover, estimates based on breast cancer studies in twins suggest that the inherited genetic component in the etiology of breast cancer accounts for at least 30% of risk (Lichtenstein et al., 2000). However, Peto (2000) contested that this is an underestimate and that the actual contribution is much higher .

The inherited genetic component of breast cancer in a population consists of both highly penetrant genes at a low frequency and those genes that occur at a high frequency, but have a low penetrance. High risk alleles have been identified in three genes; *BRCA1*, *BRCA2* and *TP53* (Lalloo and Evans, 2012). Mutations in these genes are highly penetrant, conferring a 40-85% increase in lifetime risk for developing breast cancer.

Additionally, moderately penetrant alleles have been identified in genes such as *PALB2*, *BRIP1*, *ATM*, and *CHEK2* (Meijers-Heijboer et al., 2002; Rahman et al., 2007; Seal et al., 2006; Vahteristo et al., 2002). Variants in these moderately-penetrant alleles confer a 20-40% increase in breast cancer susceptibility (Lalloo and Evans, 2012).

Many mutations and variants in highly- and moderately-penetrant genes are characteristic of specific ethnic backgrounds; therefore, frequencies of these alleles vary across populations due to founder effects and population sizes. For example, mutations in *CHEK2* exist in ~1% of Dutch, Finnish and Ashkanazi Jewish populations, while *PALB2* mutations have been identified in Finnish and French-Canadian populations (Erkko et al., 2007; Foulkes et al., 2007; Nevanlinna and Bartek, 2006). In addition, *BRCA1/2* mutations have been identified in Ashkanazi Jewish women at frequencies 10-50 times higher than those in the general population (Neuhausen et al., 1996; Oddoux et al., 1996; Roa et al., 1996; Struewing et al., 1995). Although variation in these highly- and moderately-penetrant alleles contribute significantly to breast cancer susceptibility, the total population frequency is low with combined frequencies of approximately 0.4% for *BRCA1/BRCA2/TP53* and <0.6% for *PALB2/BRI1P/ATM/CHEK2* (Lalloo and Evans, 2012). Furthermore, less than half of familial breast cancer patients exhibit predisposing mutations in these genes (Ford et al., 1995; Ford et al., 1998; Smith et al., 2006).

Population-based estimates indicate mutations in high-penetrance genes account for 25% or less of the heritable component of breast cancer susceptibility (Easton, 1999). This suggests that the majority of heritable risk of developing breast cancer is attributable to additive, dominant, and interactive effects of low-penetrance genes. However, the number and properties of these genes are not known. Much of the work performed to

identify low-penetrance breast cancer genes has focused primarily on case-control studies targeting potential candidate genes involved in cancer related pathways (de Jong et al., 2002; Nathanson and Weber, 2001). Unfortunately, results of these studies have yet to be validated in larger population-based studies (Wooster and Weber, 2003).

Although useful, risk models that include family history, such as the Gail model, are limited in power. It has been calculated, accounting for no genetic factors, that 62% of breast cancer cases can be predicted to occur in 50% of the population at high risk and 15% can be predicted in the 10% of the population at highest risk. However, if all of the low-penetrance alleles were known, the ability to predict breast cancer in 50% of the population at high risk would increase from 62% to ~90%. This would present the opportunity to detect breast cancer earlier (Pharoah et al., 2002). Therefore, identifying and characterizing common, low-penetrance breast cancer genes will strengthen breast cancer screening and prevention programs.

Common Human Genetic Variation and Breast Cancer

Genetic variation between individuals occurs at a myriad of sites across the human genome. These variants fall into two broad classes based on their nucleotide composition: single nucleotide polymorphisms (SNPs) and structural variants (Frazer et al., 2009). As the name implies, SNPs are variants at a single base position. To date, more than 12 million SNPs have been catalogued (2005). Structural variants are insertion-deletions (indels), inversions, block substitutions, and copy number variants (Frazer et al., 2009). The ability to effectively detect structural variants has been lacking;

therefore, much is still not known about the frequency of these variants and their association with human disease. However, current estimates point toward structural variants accounting for 20% of all genetic variation (Frazer et al., 2009).

Human genetic variants are defined by their minor allele frequency (MAF), which is, simply, the lowest allele frequency that a particular variant is observed in a population. Common variants are defined as having a MAF of 1% or higher. SNPs are observed more often than structural variants among individuals and it is estimated that approximately 7 million SNPs have a MAF of 5% or higher with the rest being between 1% and 5% (Barrett and Cardon, 2006; Frazer et al., 2009). These SNPs can be used to mark common variation that is hypothesized to underlie genetic susceptibility to developing breast cancer (Fletcher and Houlston, 2010; Frazer et al., 2009; Lee et al., 2011).

Identifying common genetic variants that contribute to disease requires the ability to screen and analyze thousands of variants in a large cohort of individuals in diseased and non-diseased populations. With the advent of high throughput genotyping technologies, population-based genome-wide association studies (GWAS), which assess thousands of SNPs in thousands of individuals, is possible. In another technological advancement, next-generation sequencing techniques that allow for sequencing of large regions of genomic DNA in large sample populations can be incorporated into GWASs. By comparing the genotypes of diseased to non-diseased members of a population, novel genetic determinants may be identified that associate with disease risk. GWASs are now employed to identify novel common genetic variants that associate with an increased susceptibility or resistance to developing breast cancer. Since 2008, 23 breast cancer risk

GWAS reports have been published using libraries of common polymorphisms to assay genetic variation in human populations (Hindorff et al., 2009). Several common SNPs have been identified in GWAS that associate with increased incidence of breast cancer and these SNPs are located in regions containing novel potential breast cancer susceptibility genes (Easton et al., 2007; Turnbull et al., 2010). Particularly, a GWAS by Easton *et al.* identified five independent loci that showed a strong correlation with increased breast cancer incidence ($P<10^{-7}$) (Easton et al., 2007). Of these five loci, four contain plausible potential causative genes, i.e. *FGFR2*, *MAP3K1*, *TNRC9*, and *LSP1*. Although these genes are plausible candidates, further study is necessary to confirm genes contributing to differences in susceptibility.

Further, additional SNPs are often in linkage disequilibrium (LD) with variants that show a positive association to risk. Put differently, particular SNP alleles will frequently be found together in LD and one of these SNPs can be genotyped to "tag" the other SNPs in LD. As a consequence, SNPs used in genome-wide screens may only mark a region carrying an allele associated with a disease phenotype while the causative SNP may be unknown. In other words, the causal variant may not be the one tested in the GWAS. Moreover, tagging SNPs tested in a GWAS may also be in LD with common structural variants, the majority of which have not yet been identified. This underscores one of the limitations of GWAS: there is no simple way of moving beyond statistical association to understanding the functional relationship between a genetic locus and a complex disease phenotype (Frazer et al., 2009).

Another limitation of GWAS is the prospective lack of sufficient statistical power. Many GWAS have been conducted using sample sizes of 2,000 to 5,000 individuals and

have the statistical power to identify common variants an odds ratio (OR) of 1.5 or higher (Altshuler et al., 2008). To detect variants with an OR of 1.1, studies would require 60,000 individuals to be tested. Although, some of the common genetic variants already identified have per allele OR of <1.30 (Ahmed et al., 2009; Antoniou et al., 2010; Cox et al., 2007; Easton et al., 2007; Garcia-Closas et al., 2008b; Milne et al., 2009; Stacey et al., 2008; Thomas et al., 2009; Turnbull et al., 2010) it is likely that many low-penetrance alleles may have been missed due to the small affect on susceptibility. Additionally, novel or rare variants may also be missed by GWAS. Thus, although GWAS have been useful in identifying some common genetic variants associated with breast cancer susceptibility, new approaches are needed to effectively study common, low-penetrant alleles and their effect on disease susceptibility.

Gene-Environment Interactions

Further complicating studies of breast cancer susceptibility is the combined action of risk-predisposing genes and environmental factors. Aside from studies on genetic variants in detoxifying metabolic enzymes, relatively little is known about the majority of genetic risk alleles and their potential interactions with the environment (Masson et al., 2005). Recent studies have implicated potential interactions between *FGFR2* variants and hormone replacement therapy; however, these data are inconsistent (Prentice et al., 2009; Travis et al., 2010). Additionally, a study was published on potential interactions between ten environmental risk factors (age at menarche, parity, age at first birth, breastfeeding, menopausal status, age at natural menopause, hormone replacement therapy, body-mass index, height, alcohol consumption) and twelve breast cancer

susceptibility-associated SNPs (Travis et al., 2010). No strong interaction was identified between any of the 120 possibilities. However, the majority of women in this study were post-menopausal, and therefore, these data may miss gene-environment interactions occurring in younger women. Also, moderate modulations may be occurring that are not being detected. Detecting a gene-environment interaction requires a four-fold larger sample size than does a main effect; thus, for a RR=~1.5 thousands of cases and controls or tens of thousands of GWAS cases would need to be assessed (Thomas, 2010). This study examined 7610 women with breast cancer and 10,196 controls; although a large sample size, the authors acknowledge it may be insufficient to identify moderate effects (Travis et al., 2010). Furthermore, these studies are somewhat limited in that they exclude certain environmental components; i.e. pollutants such as BPA, PHCs and PCBs. Exposure to these environmental pollutants is becoming more common and it is relevant to assess the effect of exogenous chemical exposure in the environment on breast cancer-predisposing alleles.

Animal Models of Breast Cancer Susceptibility

Mouse Models for Breast Cancer Research

Human studies provide the most direct way of studying breast cancer susceptibility. However, different approaches are needed to identify the effects of low-penetrant alleles. Human cell lines are routinely used to study breast cancer development, progression and metastasis (Burdall et al., 2003). They have a number of advantages; they are easy to handle and represent a potential unlimited self-replicating source that can be grown in almost infinite quantities. In addition, they exhibit a

relatively high degree of homogeneity and are easily replaced from frozen stocks if lost through contamination. There are disadvantages though, as cell lines are prone to genotypic and phenotypic drift due to continual culture. This occurs as subpopulations may arise over time by the selection of specific, more rapidly growing clones within a population leading to phenotypic changes (Bahia et al., 2002; Osborne et al., 1987). In addition to cell growth rate, changes have been observed in hormone receptor content, karyotype and clonogenicity, despite the cells appearing morphologically identical. However, the most important weakness for susceptibility studies is that cell lines are usually derived from tumors and have adapted to growth in culture. Although cell culture tries to create a close-to-physiology milieu by adding appropriate amounts of salt, glucose, amino acids, vitamins, and serum, the lack of tissue architecture and heterogeneous population of cell types often abolishes cell-cell interaction, secretion, and other functions based on tissue context (Pan et al., 2009). As stated, cells in culture are prone to genotypic and phenotypic drifting. Thereby cell lines can lose tissue-specific functions and acquire a molecular phenotype quite different from cells *in vivo*. Thus, cell lines are limited in scope and cannot fully replicate the disease phenotype. Although they can be used to study specific molecular targets in transformed cells they are not an effective model to study breast cancer susceptibility.

Use of animal models has been an effective approach for studying human diseases. Various model organisms have been successfully used to study aspects of different human diseases *in vivo*. The most widely used animal models of breast cancer are the mouse and the rat. Mouse models have the benefit that more genetic manipulation techniques exist for them compared to the rat. Many transgenic models have been

developed taking advantage of the long terminal repeat (LTR) of the mouse mammary tumor virus (MMTV) (Hutchinson and Muller, 2000). The MMTV-LTR is active in the mammary gland and was shown to be able to promote expression of genes in a mammary gland-specific manner (Cardiff and Kenney, 2007). The MMTV-Polyoma virus middle T antigen MMTV/PyV mT mouse model is one of the most widely used experimental animal models used to study mammary tumor development and metastasis. In this model, the mT antigen derived from PyV was placed under the transcriptional control of the MMTV-LTR (Guy et al., 1992a). The mT antigen was identified to induce multifocal tumors in mammary glands through activation of various signaling molecules such as Src family kinases and phosphatidylinositol 3' kinase (PI3K) in mice (Guy et al., 1992a). This mouse model develops polyclonal tumors aggressively within 7-8 weeks with a high degree of lung metastasis (Marcotte and Muller, 2008). The MMTV/PyV mT mouse model is often used in combination with gene knockouts to examine the influence they may have during mammary carcinogenesis or metastasis.

Another widely used transgenic mouse model is the *HER2/neu* or *ErbB2* mouse. *HER2* codes for a tyrosine kinase growth factor receptor and is the human ortholog of the rat *neu* oncogene that was identified to increase neuroblastoma development (Coussens et al., 1985; Schechter et al., 1985). Since it was identified in humans, *HER2* has been identified to be overexpressed in 20-30% of human breast cancers and has become a prime therapy target for treating breast cancer (Wang et al., 2000). The mouse ortholog of *HER2/neu* is *ErbB2*. *ErbB2* was identified to be an oncogene in studies using transgenic mice that overexpressed activated *neu* under the control of an MMTV promoter. While mammary epithelial expression of activated *neu* is sufficient to induce

mammary tumorigenesis, few activating mutations in *HER2* have been identified in humans, suggesting it exerts oncogenic effects through overexpression of the wild-type *Her2* receptor (Lemoine et al., 1990). The *ErbB2* transgenic mouse model of breast cancer was developed with a transgene carrying the wild-type *ErbB2* proto-oncogene under the control of the MMTV promoter (Guy et al., 1992b). Overexpression of wild-type *ErbB2* results in multifocal tumors, but they occur with a longer latency compared to the MMTV-PyV mouse model of breast cancer. Further, MMTV-*ErbB2*-induced mammary tumors are less metastatic than MMTV-PvV-induced tumors (Guy et al., 1992b). Regardless, the *ErbB2* transgenic mouse model has been used extensively to examine the role of the *ErbB2* proto-oncogene in mammary tumor development. In addition, the MMTV system has been extensively used to study other proto-oncogenes, e.g., c-*myc* and *cyclin D1* (Stewart et al., 1984; Wang et al., 1994). Overall, the use of the MMTV transgenic mouse system has led to a greater understanding of the genetic machinery of mammary carcinogenesis.

In addition to transgenic mouse models, gene knockout mice are regularly used and these have been combined with 'knock-in' and conditional tissue-specific gene targeting technologies allowing for a wide range of approaches to study human disease (Hutchinson and Muller, 2000; Maddison and Clarke, 2005). These studies involve deleting or inserting specific genes of interest into a targeted region using targeted recombination. Moreover, the development of the Cre-Lox system has allowed for conditional and/or temporal deletion of target genes (Maddison and Clarke, 2005). With this method, specific regions of DNA, e.g. whole genes, exons, promoters, are flanked by specific pieces of DNA termed loxP sites oriented in the same direction. When crossed

with mice carrying the Cre recombinase gene, Cre expression carries out loxP-specific recombination in a topoisomerase-like manner; thereby cleaving out the intervening sequence between the two loxP sites (Maddison and Clarke, 2005). When Cre is under the control of a tissue- or development-specific promoter, the knockout is directed to specific sites and/or at specific points in development. Also, inducible systems have been developed to control expression of genes when administered a specific inducer. Gossen and Bujard developed the tetracycline (tet)-dependent transcriptional activation system allowing spatial and temporal control of effector gene expression through the use of a tissue-specific transactivator (Gossen and Bujard, 1992). When an effector gene is under the control of the tet system, expression is activated only when the mouse is administered tetracycline in the diet. Therefore, gene expression can be easily turned on or off. Moreover, Cre can be placed within a tet-dependent system making it inducible (Gossen and Bujard, 1992; Maddison and Clarke, 2005). This way, the knockout only occurs once the mouse has been administered tetracycline. However, the Cre-Lox system has been found to have "off target" effects where Cre acts on lox-like sites causing inadvertent deletion (Maddison and Clarke, 2005). Regardless, the Cre-Lox and the tet-inducible systems have been useful in dissecting pathways of genes involved in a host of diseases.

A genetic "tool box" has been developed for the mouse making it a useful model to study the genetic and molecular aspects of human disease. The mouse has been instrumental in identifying mechanisms of specific genes involved in mammary carcinogenesis and this has increased our knowledge of these genes in human breast cancer. There are additional models and technologies not mentioned that are being

employed to study many human diseases including breast cancer. Furthermore, new technologies for using the mouse continue to be developed.

Rat Models of Breast Cancer Susceptibility

Although mice are widely used to study mammary carcinogenesis, there are some key differences between mice and humans in the pathology of mammary and breast tumors that each species develops. First, ~50% of breast carcinomas arising in humans are ERα positive and thus are hormonally responsive while mouse mammary tumors are almost exclusively ERα negative and hormone independent (Marcotte and Muller, 2008). Also, human mammary gland tissue is primarily made up of connective tissue, while mouse mammary gland stroma consists mainly of adipocytes (Marcotte and Muller, 2008). Thus, mouse model studies on mammary carcinogenesis are lacking in their ability to translate to human breast cancer.

The laboratory rat is preferable for the study of breast cancer susceptibility because the pathology of rat mammary cancer is more similar to human breast cancer. Although mouse models clearly have merit in understanding human disease, rats are more similar to humans than mice in their normal physiology and pathogenesis. Both rats and humans develop tumors that arise in epithelial cells lining the mammary gland duct have similar histopathology, and both develop the same proportion of hormone-dependent and endocrine therapy, e.g., tamoxifen, -responsive carcinomas (Gould, 1995). This is contrary to the mouse where most mammary tumors are hormonally refractive and many tumors are of endothelial cell origin (Gould, 1995). Additionally, while a majority of mammary tumors that form in mice are induced by MMTV, rats and humans do not form mammary tumors with a known viral etiology (Gould, 1995). Finally, and most

importantly for mammary gland susceptibility studies, rats and humans both have a natural genetic variation in mammary and breast carcinoma susceptibility (Dunning and Curtis, 1952; Dunning et al., 1947; Gould, 1995; Isaacs, 1986; Isaacs, 1988). These similarities make the rat an ideal model to study natural variation in development of mammary carcinogenesis.

Various rat models have been developed to study breast cancer susceptibility. Age and gender are the most widely accepted risk factors affecting breast cancer development and this has been attributed to ovarian hormone exposure. Dunning *et al.* discovered that the ACI rat strain was susceptible to estrogen-induced mammary cancers (Dunning and Curtis, 1952; Dunning et al., 1947; Dunning et al., 1953; Shull et al., 1997). Conversely, Copenhagen (COP) rats are almost completely resistant to developing estrogen-induced mammary tumors. In this regard, ACI and COP rats are commonly used together to study estrogen action in the rat. Specifically, ACI female X COP male and COP female X ACI female intercrosses were performed to generate F1a and F1b progeny, respectively. F1 progeny were administered E2 to induce carcinogenesis and tested for mammary carcinoma susceptibility. Data indicated no significant difference in tumor incidence between F1a, F1b or homozygous ACI rats. However, latency to appearance of the first E2-induced mammary tumor was significantly prolonged in both F1 populations compared to ACI homozygous rats (Shull et al., 2001). Further, siblings from F1a and F1b progeny were mated to each other to generate F2a and F2b progeny, respectively, and F1a and F1b males were mated back to ACI females to generate backcrosses a (BCa) and BCb progeny, respectively. Using these lines a locus was mapped on chromosome *5* that conferred susceptibility to estrogen-induced mammary

carcinoma development. This method was further utilized to identify multiple quantitative trait loci (QTLs) in the ACI rat responsible for estrogen-induced mammary carcinoma susceptibility (Gould et al., 2004; Kurz et al., 2008; Shull et al., 2001).

Other rat lines have varying propensities for developing exogenous non-estrogen carcinogen-induced mammary carcinomas. The Wistar Furth (WF) rat strain was identified to be susceptible to developing spontaneous, radiation-, oncogene-, and carcinogen-induced mammary carcinomas (Gould, 1986; Moore et al., 1983). Conversely, the COP rat strain is susceptible to ionizing radiation and oncogene-induced mammary carcinogenesis, but is almost completely resistant to developing chemical carcinogen-induced mammary tumors (Isaacs, 1988). Studies using ionizing radiation on (WF X COP) F1 rats detected random allelic imbalances throughout the genome leading to development of mammary tumors with no preferential loss for either the WF or COP parental alleles (Haag et al., 1996). This indicates that neither the WF nor COP allele confer resistance to radiation-induced mammary carcinogenesis. Additionally, both WF and COP females treated with a v-H-*ras* oncogene-containing retrovirus directly infused into the mammary gland develop similar tumor incidences (Wang et al., 1991). As previously mentioned, COP rats are resistant to estrogen-induced tumors; however, hormonal promotion enhanced the penetrance of the *ras* oncogene on mammary tumor formation in both COP and WF female rats in a similar manner (Wang et al., 1991). Although the COP and WF rats showed similar tumor incidence from *ras* oncogene induction, COP tumors were more differentiated and less invasive than WF tumors (Wang et al., 1991). This suggests that although oncogene activation *in situ* was able to

ultimately overcome the resistance phenotype, COP rats still exhibit a less malignant phenotype compared to WF rats.

Mammary Carcinoma Susceptibility Quantitative Trait Loci

Mammary Carcinoma Susceptibility Alleles

The rat is not as easily genetically manipulated as the mouse. Until recently, chemical or random-insertional mutagenesis screens had been employed in the rat to identify causative genes for specific phenotypes (Aitman et al., 2008). Now, transgenic rat models are available and it is possible to generate knockout lines (Aitman et al., 2008; Geurts et al., 2009). However, these approaches are biased towards particular genes of interest or involve disrupting the genetic architecture by random insertion of foreign DNA into the chromosome. Although these methods may be used for validating or studying the effects of specific genes, they are not suited for forward genetics studies focused on identifying genotypes responsible for disease susceptibility phenotypes. Instead, congenic rat lines have been developed that allow one to study mammary carcinoma susceptibility in a "natural" context. Congenic lines take advantage of varying susceptibility phenotypes between different rat strains. Congenic rat lines for mammary carcinoma susceptibility studies are developed by introgressing alleles from a resistant strain onto a susceptible strain by first identifying an allele of interest in an F1 population and then continuously backcrossing to a recipient strain. By genotyping and selecting for the donor allele of interest, the donor allele is retained and introgressed onto the genotype of the recipient background strain (Figure 1) (Markel et al., 1997). Lines with different segments of the donor allele can be tested to map a genetic locus to define the allele of interest.

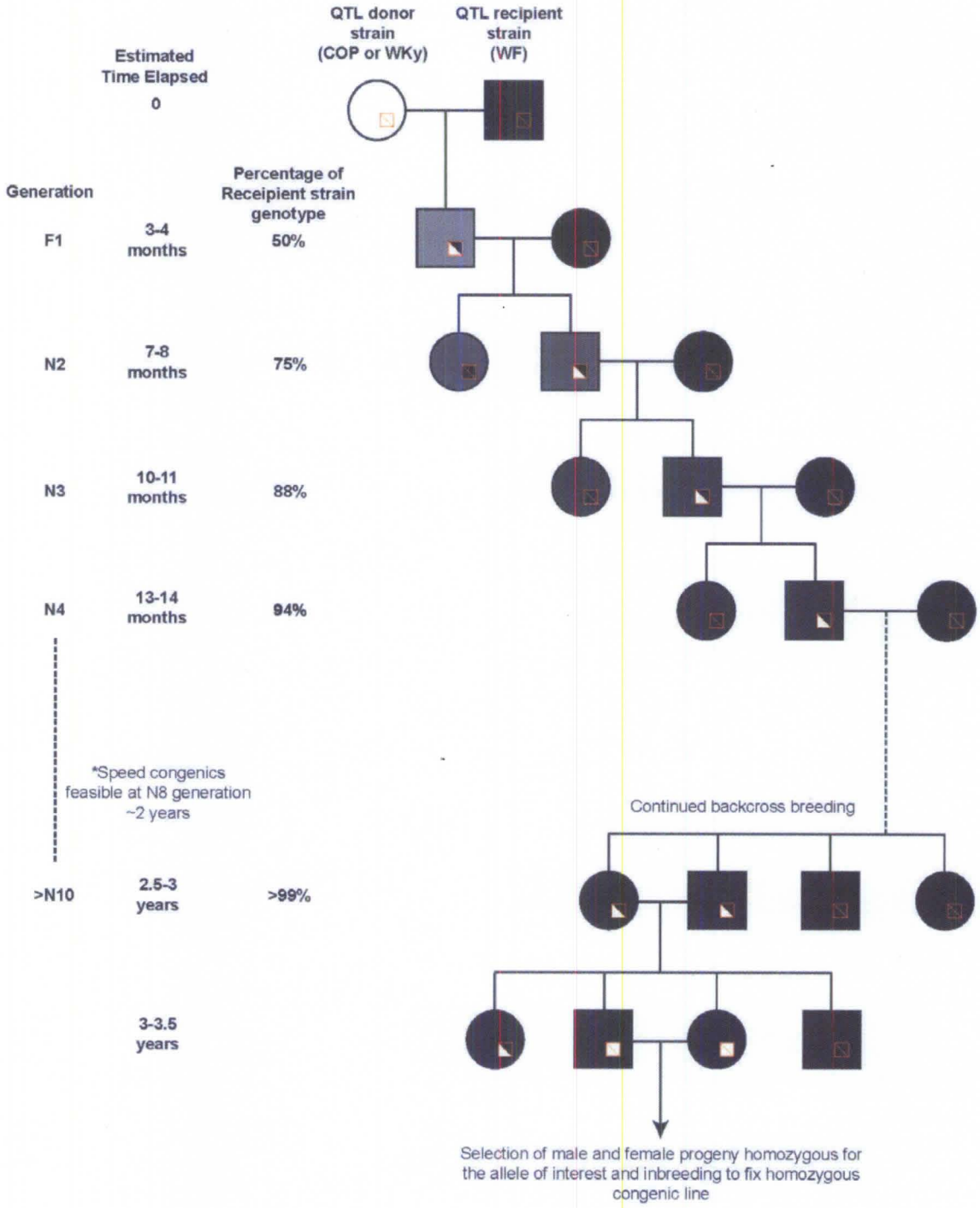

QTL donor strain (COP or WKy)
QTL recipient strain (WF)
Estimated Time Elapsed
0
Generation
Percentage of Receipient strain genotype
F1
3-4 months
50%
N2
7-8 months
75%
N3
10-11 months
88%
N4
13-14 months
94%
*Speed congenics feasible at N8 generation ~2 years
>N10
2.5-3 years
>99%
3-3.5 years
Continued backcross breeding
Selection of male and female progeny homozygous for the allele of interest and inbreeding to fix homozygous congenic line

Figure 1: Congenic Breeding Scheme. Homozygous susceptible male rats (WF) are bred to female resistant rats (COP/WKy/BN) to create an F1 heterozygous population. F1 progeny are genotyped for the allele of interest. F1 males containing the allele of interest are backcrossed to WF females to produce the N2 population. At each subsequent generation the litters are genotyped and males carrying the allele of interest are backcrossed again to WF females. At the N8 generation, speed congenics are potentially viable to use for mapping as the background genotype is >99%. However, backcrossing to N10 is common. At N10 (or N8) heterozygous male and female littermates are inbred and the offspring are genotyped to identify pups homozygous for the target allele. These homozygotes are inbred continuously to fix the line for the allele of interest. ○, ■ and ● represent resistant female, susceptible male, and susceptible female rats, respectively. The gray shade of a circle or square depicts the genotype percentage of the background genome in heterozygous offspring; i.e. as the percentage of the background genome becomes more WF, the shade becomes darker gray. ■, ▨, and ◥ inside a circle or square represent the target allele when it is homozygous WF, homozygous COP, and heterozygous, respectively.

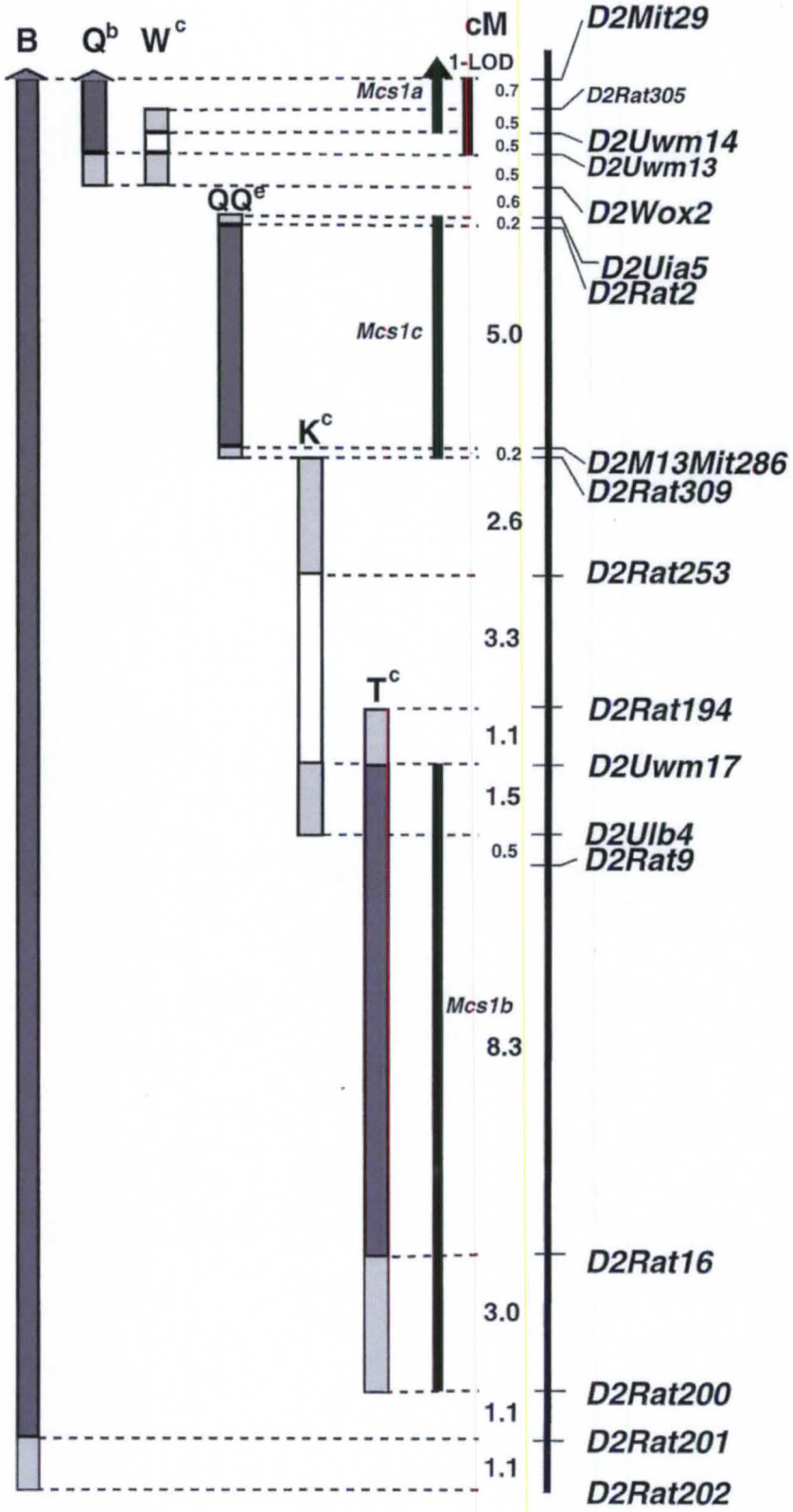
Work by our group employs congenic strains developed with resistant alleles COP and Wistar Kyoto (WKy) donor strains on a WF background denoted as WF.COP and WF.WKy congenic lines, respectively. In contrast to WF rats, COP and WKy rats have been shown to be almost completely resistant to developing mammary carcinomas induced by the chemical carcinogens DMBA and *N*-Nitroso-*N*-methylurea (NMU) (Haag et al., 1992; Isaacs, 1986).

WF.WKy and WF.COP have been used to identify novel independent loci affecting mammary carcinoma susceptibility (Cotroneo et al., 2006; Haag et al., 2003; Lan et al., 2001; Samuelson et al., 2003). These QTLs have been denoted as mammary carcinoma susceptibility (*Mcs*) loci. The first *Mcs* locus, *Mcs1*, was found in a DMBA-induced carcinogenesis study on female progeny from a (WF X COP)F1 X WF backcross (Hsu et al., 1994). Since then, WF.COP and WF.WKy congenic lines have been used to identify other *Mcs* loci: *Mcs2*, *Mcs3*, and *Mcs4*, all found in WF.COP congenics, and *Mcs5*, *Mcs6*, *Mcs7* and *Mcs8*, found in WF.WKy congenics (Lan et al., 2001; Shepel et al., 1998). Each strain, COP or WKy, has three QTLs that increase and one that decreases susceptibility to developing DMBA-induced mammary tumors. Only one QTL in each strain overlaps with a QTL at the same genetic locus in the other strain. These are COP *Mcs2* and WKy *Mcs6*. This illustrates the genetic diversity of mammary carcinoma genetic susceptibility as both of these strains are highly resistant to developing DMBA-induced mammary carcinomas.

Mammary Carcinoma Susceptibility Locus 1b (Mcs1b)

As stated above, (WF X COP)F1 X WF backcrosses were used to identify the *Mcs1* locus. Rats were divided into groups based on susceptibility: resistant (R), undetermined (U), and susceptible (S). Rats from R and S groups were genotyped using microsatellite markers and the locus conferring resistance was determined to be on the proximal end of chromosome *2* (Hsu et al., 1994). Resistant congenic F1 rats were backcrossed to WF rats and progeny were genotyped to identify shorter regions of the interval to fine map the *Mcs1* locus. Results of phenotyping showed that the *Mcs1* locus contained 3 independent loci that each reduced DMBA-induced tumor incidence by ~60% compared to homozygous WF controls (Haag et al., 2003). These were termed *Mcs1a*, *Mcs1b*, and *Mcs1c* (Figure 2). Interestingly, one of these loci, *Mcs1b*, contains a region orthologous to a human locus identified in a GWAS of breast cancer susceptibility (Easton et al., 2007).

Easton *et al.* (2007) reported the identification and validation of five novel breast cancer susceptibility loci in a 3-stage breast cancer GWAS (Easton et al., 2007). One of these SNPs, *rs889312*, has MAF of 0.38 with an allelic OR of 1.13 (95% CI=1.10 - 1.16) and associates with an increase in breast cancer incidence ($P=7 \times 10^{-20}$). Additionally, SNP *rs889312* localizes to a region on human chromosome *5* and is in LD with 6 other SNPs contained within a 280 kb haplotype block (Figure 3)(Barrett et al., 2005). Importantly, this human haplotype block is orthologous to a region within the rat *Mcs1b* locus. There are three transcripts annotated to the human locus delineated by the 280 kb haplotype block; *MAP3K1*, *MIER3*, and *C5ORF35* (Figure 3). These transcripts are also annotated to the rat *Mcs1b* locus. In addition there are 7 more transcripts annotated that lie nearby, but outside the 280 kb haplotype block on the human locus, that are also



Haag *et al. Cancer Research*, 2003

Figure 2: Chromosome *2* genetic map of *Mcs1*-congenic and -recombinant rat lines. Rat lines are designated with *capital letters*. ■, indicates the presence of two COP alleles for congenic lines resulting in reduced mammary tumor development (one to four carcinomas/rat); □ indicates the presence of two COP alleles for congenic lines incapable of conferring resistance to tumor development (six to eight carcinomas/rat). ▩, indicates areas of unknown genotype because of recombination. Three independent regions of chromosome *2* were capable of conferring resistance to mammary cancer development in COP-homozygous congenic rats. These genomic regions are shown as *Mcs1a*, *Mcs1b*, and *Mcs1c*. The chromosome markers used to identify the congenic rats are listed to the *far right* of the figure with genetic distances in centiMorgans (cM) to the *left* of the marker names. The *Mcs1* QTL 1-LOD interval is also shown between *D2Mit29* and *D2Uwm13*.

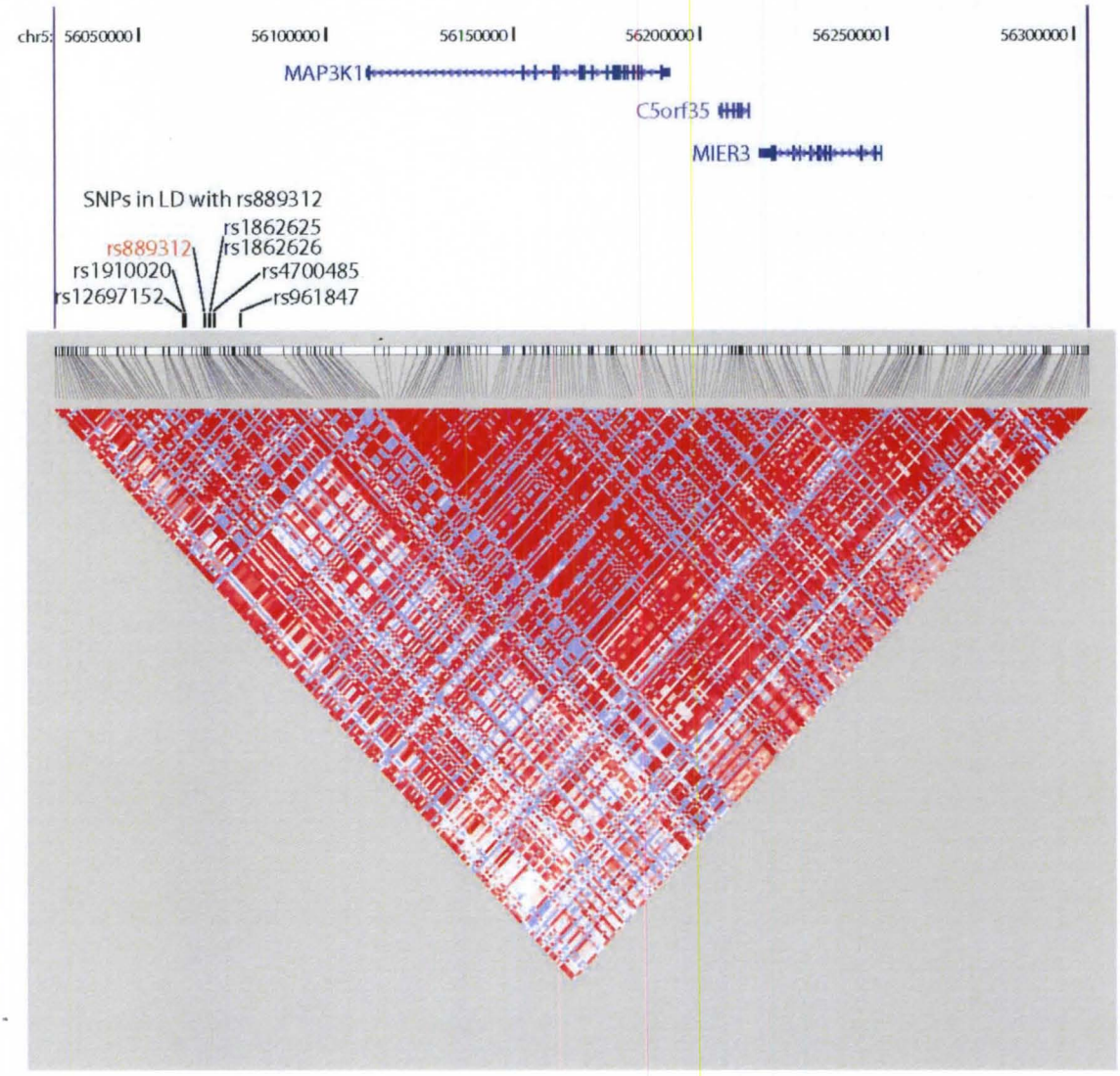


Figure 3: 280 Kb haplotype block containing breast cancer-associated SNP *rs889312*. Transcripts are shown in blue with exons designated as vertical bars. Above, the location of the tagging SNP *rs889312* is shown in red, other SNPs in LD with *rs889312* are shown in black. Below, pairwise r^2 values for LD are shown; light blue represents lower disequilibrium, white is intermediate, and red represents strong disequilibrium. An r^2 threshold of 0.80 was used to determine SNPs in LD in this block.

contained in the rat *Mcs1b* locus; *GPBP1*, *IL31RA*, *IL6ST*, *DDX4*, *ANKRD55*, *ACTBL2*, and *SLC38A9*. *MAP3K1* has been described as the most likely candidate gene due to its role as a protein kinase and involvement in promoting cell growth and proliferation; still, any of these genes could be involved in affecting breast cancer susceptibility in humans. Furthermore, any of the SNPs or combination of variants in LD with *rs889312* could be causative to affect breast cancer susceptibility. Therefore, each SNP must be studied to identify the causative variant. The original study identifying this human locus consisted of an initial two-stage GWAS on 4,398 breast cancer cases and 4,316 controls, followed by a third stage testing 21,860 cases and 22,578 controls from 22 studies (Easton et al., 2007). Testing individual SNPs in LD with *rs889312* in this population would offer a better view of the genetic composition associated with this phenotype; however, this provides no insight into their function. Studying the molecular mechanisms by which these SNPs are operating in a human population is not realistic. Therefore, the WF.COP *Mcs1b* congenic rat line is a good model to study this locus in the context of breast cancer susceptibility.

There are over 50 transcripts annotated within the rat *Mcs1b* locus. It is anticipated that the causative gene is conserved between humans and rats which reduces the number of viable gene candidates to be tested. Nonetheless, all of the transcripts annotated to the current rat *Mcs1b* locus must be treated as potentially causative; thus, all genes must be tested. It is useful, then, to attempt to reduce the number of genes to be tested by narrowing the current rat *Mcs1b* locus by means of positional mapping. The rat genome has been sequenced and many microsatellite and SNP markers are available to effectively map *Mcs1b* to a narrower interval. Testing genes in the rat may identify the

causative gene involved at the human locus. Related to this, the *Mcs1b* rat model can also be used to study the cellular and physiological phenotypes affecting mammary carcinogenesis associated with this locus. On the whole, a *Mcs1b* rat genetic model could help determine mechanisms controlling human breast cancer that are associated with the locus demarcated by SNP *rs889312* and expand our current understanding of breast cancer etiology.

Dissertation Overview

Overall Goal

The work presented in this dissertation is focused on using congenic WF.COP *Mcs1b* rat lines to study mammary carcinoma susceptibility. I made use of the varying propensity for developing mammary tumors between WF.COP *Mcs1b* congenic female rats and homozygous WF female rats. The overall goal was to narrow the genetic interval and to identify potential gene candidates. Mapping the *Mcs1b* locus to a shorter interval reduces the number of potential gene candidates and focuses on identifying the causative gene(s). Identification of the causative gene(s) could lead to development of novel genetic tests for better diagnosis as well as novel targets for prevention and treatment of breast cancer. In addition, I began to identify the cellular mechanism by which rat *Mcs1b* affects mammary carcinogenesis and tumor development. By elucidating the means by which this susceptibility locus is working, we will increase the current understanding of breast cancer etiology.

Hypothesis and Research Aims

I hypothesized that the rat *Mcs1b* locus contains genetic elements controlling molecular and cellular phenotypes that translate into mammary carcinoma susceptibility phenotypes. By completing aim one; I successfully delimited the rat *Mcs1b* locus to a region of 1.1 Mb containing three genes. In aim two, I identified that the mammary carcinogenesis susceptibility allele is acting in a mammary gland cell-autonomous manner. In aim three, I show that the transcript *Mier3* is a candidate modifier gene controlling mammary carcinogenesis based on its differential transcript expression between susceptible and resistant rat strains. These data will help us better understand breast cancer etiology and may lead to better diagnoses and treatments of human breast cancer.

CHAPTER II

RAT *MCS1B* IS WITHIN A 1 MB REGION OF RAT CHROMOSOME *2*

Introduction

The rat *Mcs1* locus was originally identified on the proximal end of chromosome *2* in (WF x COP)F1 x WF backcrosses using DMBA to induce mammary carcinomas (Hsu et al., 1994). Previously, the *Mcs1* locus was physically confirmed and positionally mapped using congenic lines containing unique segments of the *Mcs1* predicted QTL interval. Female rats from these lines were tested using the same DMBA carcinogenesis protocol and revealed that the locus contained three independent susceptibility loci that conferred resistance to developing mammary carcinomas. These QTLs were termed *Mcs1a*, *Mcs1b* and *Mcs1c* (Haag et al., 2003)(Figure 2).

The *Mcs1b* locus is of particular interest because it is orthologous to a human locus associated with increased breast cancer incidence that was identified in a GWAS (Easton et al., 2007). The human breast cancer risk-associated allele was tagged with SNP *rs889312* which lies within a 280 kb LD block on human chromosome *5*. This LD block contains three annotated transcripts; *MAP3K1*, *MIER3*, and *C5ORF35* (Figure 3). The rat ortholog to this locus is contained in the *Mcs1b* congenic line T interval, which was mapped to a region delimited from marker *D2Uwm17* to *D2Mco42* (Chr*2*: 32051319 - 45248161). This locus contains over 50 annotated transcripts, including *Map3k1*, *Mier3*, and *C5orf35*. Mapping rat *Mcs1b* to a narrower region will reduce the number of

candidate genetic elements to be functionally tested and minimize the work required for comparative genetic studies.

Many QTLs are multigenic or affected by more than one genetic element. For that reason, two or more elements may potentially act in concert to affect susceptibility. Narrowing the interval may provide insight into the function of this locus. Accordingly, I attempted to map the *Mcs1b* locus to a tighter interval using a congenic mapping approach previously used to segregate and identify three independent loci contained within the original *Mcs1* locus (Haag et al., 2003). This approach will reduce the genetic elements to be functionally tested as well as potentially identify complex genetic interactions or additional risk alleles within the line T *Mcs1b* interval.

The advantages of the rat over a mouse model have been noted. However, there are other rat-centric methods that could be employed, e.g., transgenic rat models, rather than the congenic approach used in these studies. Many genes are annotated to the rat *Mcs1b* locus, as it is currently defined, and these could be tested in a transgenic rat model by introducing the genes as transgenes under the control of an artificial promoter. However, this approach would be extremely labor-intensive requiring the development of over 50 transgenic lines. Further, gene regulatory elements cannot be cloned into traditional transgenic cassettes. This is critical, since we do not know the transcript profile of the hypothesized candidate gene. In this regard, bacterial artificial chromosome (BAC) transgenic rats may be useful. BAC transgenics allow for up to 400 kb of foreign DNA to be cloned so that all the regulatory elements may be included. For this reason, BAC transgenes are expressed in a more spatially, temporally, and physiologically accurate manner. The current *Mcs1b* locus is delineated to a ~15 Mb

interval containing 50 transcripts; thus, much like traditional transgenics, the BAC transgene method would also require many lines to be developed to tile across the region. Moreover, if there are multiple components involved, using a BAC transgenic may "hide" these effects. Another shortcoming of this method is that the BAC transgene is incorporated randomly in the genome. Although regulatory elements and surrounding DNA are included, this technique disregards possible long range effects, e.g., distal enhancers and long range *cis* regulatory elements, thereby altering the context.

Although laborious and time-consuming, the congenic rat model is the preferred approach. It is powerful in that it is unbiased with regard to the cause of action, i.e., the heritable element(s) controlling the susceptibility phenotype may be non-protein coding regulatory elements that control expression or processing, causative elements may be in *cis* of secondary element, or they may be multigenic and require multiple factors. Overall, congenics allow one to study the action of the *Mcs1b* locus in its native physiologically-relevant environment. For these reasons, I continued to use the previously-described congenic approach for these studies.

Until now, the rat *Mcs1b* region has been genotyped using microsatellite markers. The genome sequence for the rat is available and groups are attempting to develop SNP maps of the rat genome (Nijman et al., 2008; Saar et al., 2008). However, the annotated sequence is based on the Brown Norway (BN) strain, and the COP and WF genomes were not included in SNP and haplotype studies; therefore, there are no well-documented polymorphisms between the COP and WF lines in the *Mcs1b* region. Identifying new polymorphisms between the WF and COP strains will provide better markers to more accurately map the *Mcs1b* locus. Moreover, within the 280 kb haplotype block

containing the human *Mcs1b* ortholog, SNP *rs889312* is in LD with 6 other annotated SNPs: *rs12697152*, *rs1910020*, *rs1862625*, *rs1862626*, *rs4700485*, and *rs961847* (Figure 3). Any of the SNPs lying in the 280 kb LD block could be causative. However, it is not clear which of these SNPs, if any, is the causative variant responsible for conferring breast cancer susceptibility. Rat orthologs to the 7 human SNPs lying within the LD block are not known. Consequently, identifying novel polymorphisms between the WF and COP rat lines could potentially lead to identifying the causative SNP controlling mammary carcinoma susceptibility and comparative genetics could be used to accurately identify the causative SNP in the human genome.

Overall, identifying novel polymorphisms between the WF and COP strains will add to the current knowledge of these strains and can be used for the mapping studies described herein. These SNPs are potentially useful in that they may shed light on the causative element responsible for the difference in mammary carcinoma susceptibility between these two rat lines and identify the causative variant in the human genome. Fine-mapping this region in the rat will reduce the number of potential gene candidates that must be functionally evaluated. As stated, this approach is powerful in that it is unbiased with regards to the cause of action, i.e., the heritable element(s) controlling the susceptibility phenotype may be a regulatory element(s) controlling expression or processing of a transcript. In addition, these elements may lie within or outside of the protein coding region of the gene. Indeed, all of the SNPs in the human breast cancer-associated haplotype block lie outside of the transcribed regions of any of the annotated transcripts. Furthermore, this approach may uncover complex genetic interactions and/or potential additional mammary carcinoma susceptibility loci that lie within the region.

Design and Methods

Resequencing

WF and COP rats were euthanized and splenectomized using a protocol approved by the University of Louisville Institutional Animal Care and Use Committee (IACUC). Spleens were frozen until DNA was isolated. DNA was isolated using a Gentra Tissue DNA Extraction Kit (Gentra, Minneapolis, MN) and PCR amplified using AccuPrime Taq polymerase (Life Technologies, Grand Island, NY) at specific regions defined in an article by Cuppen and colleagues (Nijman et al., 2008) or compiled in the STAR Consortium SNP database (http://www.snp-star.eu/). These PCR products were purified using a QIAquick PCR Purification Kit (QIAGEN, Valencia, CA) and sequenced using the BigDye Terminator v3.1 Cycle Sequencing Kit (Life Technologies). Sequencing reaction products were purified with 5 μl Agencourt AMPure XP beads and 80% ethanol. Beads were washed once with 80% ethanol and subsequently eluted in molecular biology grade water. Sequence products were submitted for analysis to the University of Louisville Center for Genetics & Molecular Medicine DNA Core using an ABI PRISM 3130XL Sequence Detection System (Life Technologies).

Additionally, random regions across the *Mcs1b* locus were selected to be amplified and sequenced to identify polymorphisms. Using the UCSC genome browser, regions exhibiting stretches of di- and tri-nucleotide repeats were selected and primers were designed using Primer3 (Rozen and Skaletsky, 2000) oligonucleotide design software. Primers were obtained from Integrated DNA Technologies (IDT, Coralville, IA). Using these primers, genomic DNA (gDNA) was PCR amplified and run on high-resolution agarose gels to identify potential microsatellite polymorphisms between WF

and COP sequence. Additionally, long spans were selected randomly across the *Mcs1b* locus and PCR amplified and sequenced to identify potential SNPs between WF and COP sequence. Sequence reads were analyzed using DNAStar Sequence Analysis software (DNAStar, Madison, WI).

Congenic Strain Breeding

The congenic breeding method was detailed above (Figure 1). The rat WF.COP *Mcs1b* congenic lines T and B were used to generate new recombinant lines and potentially isolate the SNP(s) and/or gene(s) involved in the breast cancer susceptibility phenotype. These lines, containing varying pieces of the rat *Mcs1b* locus, will be tested using the DMBA mammary carcinoma susceptibility assay. The line T *Mcs1b* locus was delineated to ~13 Mb interval at *D2Uwm17:g2UL2-30* (Chr*2*:32051320-44932309) and confers a reduction in mammary tumor development (Figure 2) (Haag et al., 2003). Line B was delineated a Chr*2* region from marker *D2Mit29* to marker *D2Rat201* Figure 2. In addition to *Mcs1b*, the Line B COP interval contains *Mcs1a* and *Mcs1c*; however, the recombinant lines at the distal end may be used to narrow the *Mcs1b* locus.

Inbred WF male and female rats were obtained from Harlan Sprague Dawley, Inc. (Indianapolis, IN). All housing and breeding was performed in the Research Resources Center Animal Facility at the University of Louisville under protocols approved by the University of Louisville Animal Care and Use Committee. Line T male and female rats at the N10 generation or beyond were bred with inbred homozygous WF female or male rats, respectively, at 12 weeks of age. Progeny were genotyped for unique recombinant intervals of the *Mcs1b* locus. Novel recombinant rats were backcrossed with inbred

homozygous WF rats to obtain heterozygous progeny. Heterozygous male and female progeny were inbred to obtain progeny homozygous for the unique recombinant allele.

WF.BN-*RN02* animals were generated in a similar manner by initially breeding BN female rats (Harlan Sprague Dawley, Indianapolis, IN) to a WF male rat to obtain an F1 generation. An F1 generation male was backcrossed to inbred (homozygous) WF female rats and progeny were screened for the *Mcs1b* locus using the markers mentioned. WF.BN animals were bred out to the N6 generation. Mammary carcinoma susceptibility phenotype studies were performed as described.

Genotype Analysis

Progeny from crosses between congenic and homozygous WF rats were genotyped using an approved protocol. A tail clipping of each pup was taken at 1-2 weeks of age. DNA was extracted from tail clip samples using a Gentra Tissue DNA Extraction Kit, diluted and PCR amplified using primers for informative microsatellite markers. Multiple markers were tested spanning the length of the original line T locus from *D2Rat194* to *D2Rat201* (Table 1). Microsatellite markers were analyzed by PCR amplification using the GeneAmp Fast PCR system (Life Technologies) and separation on 3% high-resolution agarose gels in TBE buffer. SNP markers were analyzed by either PCR amplification followed by sequence analysis by the University of Louisville Center for Genetics & Molecular Medicine DNA Core.

Phenotype Analysis

At 50-55 days of age nulliparous female homozygous WF and female congenic rats were administered DMBA (Acros, Pittsburgh, PA) at 65 mg/kg body weight by gavage. DMBA was suspended in sesame oil at a concentration of 20 mg/ml, heated in

Table 1:

Informative microsattelite markers used to narrow Mcs1b locus to 1.8 Mb interval

Samuelson Lab ID	Public ID	Position*	Forward Primer Sequence	Reverse Primer Sequence
D2Rat194	D2Rat194	29237360-29237535	TAATTGCAACAGGTCAGGGC	GAGTGGATTTGAGAGCAGCTG
D2Uwm17	D2Uwm17	32051319-3205163	AAGCTACAATGCCTAGCAAC	CCAACAGGACTTTAGTCATTG
D2Mco43	D2Mco43	32836549-32836763	AACCACTTTTAGAATGTTAATCAG	CGATCCTTCATGGGGCTAACACT
D2Got11	D2Got11	33838802-33839075	CCTGGTCTCTGTCTCTGTCTCA	TCCTTTAGCCTTCCTTTTGG
D2Rat195	D2Rat195	33957192-33957523	TTGCTGTTTCTAGTATGTGCAGG	CCCATGCACACAAGTATGAA
D2Rat12	D2Rat12	39101957-39102109	CCAGTCCCTCAGAAGGAACA	GCAACCACATTTTCAGAATTGA
D2Mgh2	D2Mgh2	39829003-39829053	GAAATGGGGAGTCAGAGAAGG	TTTCTTGTTTACCTCTGTCTGGG
D2Rat199	D2Rat199	41032096-41032201	TCAGGTATCTCCTATGGGGG	GAG CGC TCA TTG CTC TCT CT
D2Rat142	D2Rat142	42318356-42318464	CACAAATGCATGTGTGCCTT	CAAAGCCTTTGATTGTGCAA
D2Rat16	D2Rat16	43376467-43376635	CTGCATGTGTTAAATCATTAGTCA	ACTTCACGATCCAGTCTGGG
g2UL1-5	12324219†	43485572-43485628	AGACAATCCCCCACAGACAT	GAGAAGGTGCATGTTCCAAA
g2UL2-27	12324060†	44195286-44195382	TAAATGTGGTTTCCTTTGCT	TCAACGTAGCTGAAATTGTG
g2UL2-29	NA	44325512-44325564	CATAACAGCAAGAAGCATCA	GGAAAAGAACAACTGTTTGG
D2Mco42	D2Mco42	45247893-45248161	GAGGAGTATATTAGTTTGGGCTG	ATGGGCTGGCTAGTGAGAAAGTT
g2UL2-30	12324220†	44932096-44932309	ATTCAATTCCAACAATCCTC	CATTTTCAAGCCTTACAGGT
D2Rat200	D2Rat200	48762858-48762979	AGGGTGGTTTGAAGCCAGTT	CAGGATTGAACAGCAAGCAG
D2Rat201	D2Rat201	49691463-49691646	GCAACCACAAAAGGAGAAGG	GCTAACTAGAATGCATTTCAAAATT
D2Rat202	D2Rat202	51821000-51821161	TGGCTTAGCATAATCTCAGCA	CGCCCAGCTCACATTAATTT

Abbreviations: WF, Wistar Furth; COP, Copenhagen

Genotypes determined using 3% high-resolution agarose gel

*Position based on *Rattus norvegicus* Chromosome *2*, genome bulid version 3.4

†Marker IDs in NCBI database

boiling water to dissolve and cooled to room temperature. For each congenic line, 15-25 female rats were used for phenotype analysis. At 15 weeks post DMBA administration, rats were euthanized and the total number of mammary carcinomas measuring≥3x3 mm in diameter were counted per rat. Spleens were removed to confirm genotype. Statview (SAS Institute, Cary, NC) statistical analysis software was utilized for analysis. Data are presented as means ± SD. Nonparametric Mann-Whitney tests were performed for analysis within congenic groups.

Haplotype Block Analysis

Human haplotype blocks and SNPs in LD are based on the *GRCh37/hg19* human genome assembly and were identified using the LD and Tagger functions in Haploview v.4.2 (www.broad.mit.edu/mpg/haploview/), respectively (Barrett et al., 2005). Sequences for human haplotype blocks were identified using the UCSC Genome Browser (genome.ucsc.edu/). Syntenic analysis was performed using the *Convert* function of the browser to determine the rat ortholog for the human haplotype block associated with each SNP (Kent et al., 2002).

Results

Line T Backcrosses Result in Novel WF.COP Mcs1b Congenic Rat Strains/Lines

Mapping the *Mcs1b* QTL to a shorter syntenic interval reduces future comparative genetics work and abbreviates the list of potential candidate genes to be tested. To fine map the locus five congenic lines were generated by backcrossing *Mcs1b* line T and line B (Figure 2) to homozygous WF rats. These were termed F3, U2, W2, N3, and I4

(Figure 4). Each of these lines contain a unique COP rat Chr*2* segment of the *Mcs1b* candidate region from *D2Uwm17:D2Rat201* (Chr*2*: 32051319 - 49691646) on a susceptible WF genetic background (Table 2, Figure 4). Lines W2 and U2 contained a COP allele spanning from marker *D2Uwm17* to marker *D2Rat142*. Line N3 contained a COP allele spanning from marker *D2Uwm17* to marker *g2UL1-5*. Line I4 contained a COP allele spanning from marker *g2Uwm65-18* to marker *bUwm15-3*.

When a unique recombinant animal was identified from a backcross, it was bred to a homozygous WF animal to generate additional heterozygous rats. Progeny were genotyped and pups lacking the recombinant allele were not used. Once enough male and female rats carrying the allele of interest had been generated they were inbred to produce homozygous rats. These homozygous rats were inbred from here forth to expand and maintain the line.

Mammary Carcinoma Susceptibility Phenotypes of Congenic Lines F3, U2, W2, I4 and N3 Shorten Mcs1b to a 2 Mb Interval

Mcs phenotypes were determined using tumor multiplicity at 15 weeks following DMBA induction of mammary carcinogenesis. As anticipated, the lines exhibited varying propensities to developing DMBA-induced mammary carcinomas. Rats homozygous for the F3, U2, and W2 COP allele developed 9.6 ± 4.1 (N=32), 5.7 ± 3.9(N=6) and 6.0 ± 1.8 (N=18) tumors per rat, respectively, while littermate homozygous WF control rats developed 8.8 ± 3.5 (N=32), 6.3 ± 2.2 (N=12) and 5.9 ± 3.2 (N=9), respectively (Table 2). These numbers suggest that there is no difference between rats carrying the U2 and W2 COP alleles and rats having the WF allele in this interval.

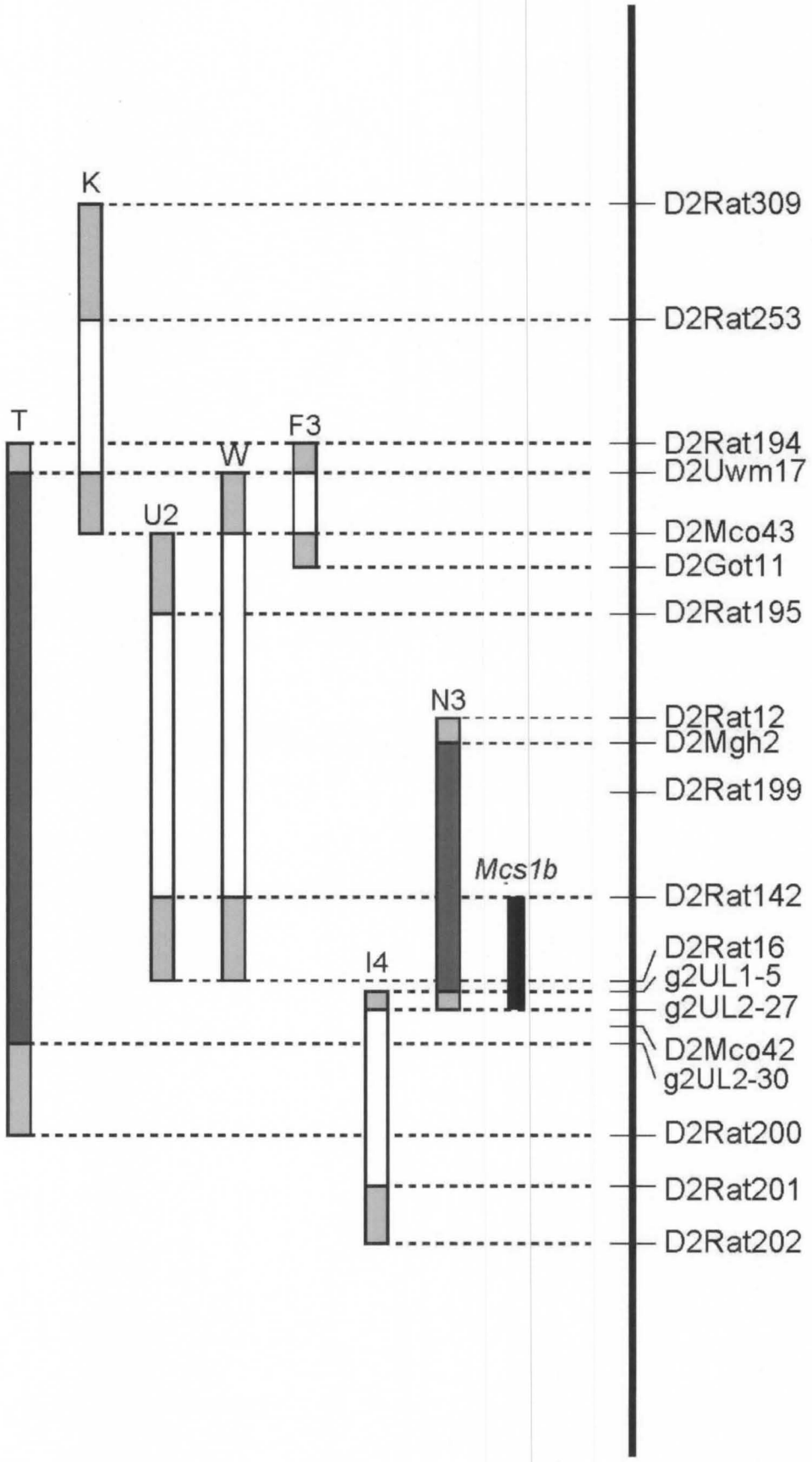

K
T
F3
W
U2
N3
I4
Mcs1b
D2Rat309
D2Rat253
D2Rat194
D2Uwm17
D2Mco43
D2Got11
D2Rat195
D2Rat12
D2Mgh2
D2Rat199
D2Rat142
D2Rat16
g2UL1-5
g2UL2-27
D2Mco42
g2UL2-30
D2Rat200
D2Rat201
D2Rat202

Figure 4: Rat Chr *2* map of WF.COP lines delimiting *Mcs1b* to 1.8 Mb. Markers used to genotype WF.COP congenics are listed in relative positions on the *y*-axis. Lines are labeled with letter–number combinations and designated with filled dark-gray bars to indicate *Mcs1b*-resistant alleles. Lines that are drawn with unfilled bars represent COP intervals incapable of conferring decreased susceptibility or resistance to mammary carcinoma development. The filled light-gray bars at ends of each congenic segment are intervals of unknown genotype. Lines T and K are shown again for reference.

Table 2:

Mammary carcinoma multiplicity phenotypes (mean mammary carcinomas per rat ± SD) by genotype for WF.COP Chr 2 congenic lines used to map Mcs1b to 0.75 Mb

WF.COP *Chr2* region	Line	COP/COP (COP/WF)	N	WF/WF	N	P value*
D2Uwm17/g2UL2-30	T†	3.5 ± 2.2	21	8.3 ± 3.3	19	0.001
	T†	(7.6 ± 3.4)	18			NS
D2Uwm17/D2Ulb4	F3	9.6 ± 4.1	32	8.8 ± 3.5	32	0.8433
D2Mgh2/g2UL1-5	N3	3.4 ± 2.0	25	7.8 ± 3.1	25	0.0001
	N3	(5.5 ± 3.6)	15			0.0413
D2Ulb4/ENSRNOSNP274	W2	6.0 ± 1.8	18	5.9 ± 3.2	9	0.8498
D2Rat116/ENSRNOSNP27	U2	5.7 ± 3.9	6	6.3 ± 3.3	12	0.8866
g2UL2-27/D2Rat201	I4	9.3 ± 3.0	19	7.9 ± 3.7	13	0.247
D2Rat16/g2UL2-30	1b-11	5.9 ± 2.2	12	6.4 ± 1.9	8	0.4179

Abbreviations: WF, Wistar Furth; COP, Copenhagen; *Chr*, chromosome

*P values from Mann Whitney tests

†Line T phenotype published previously by Haag *et al. Cancer Research*, 63:5808-5812, 2003

Similarly, female rats homozygous for the I4 COP allele developed 9.3 ± 3.0 (N=19) tumors per rat compared to 7.9 ± 3.7 (N=13) for homozygous WF littermate controls again suggesting that the congenic I4 rats have the same mammary carcinoma susceptibility phenotype as homozygous WF rats. Conversely, female rats from line N3 that were homozygous for the N3 COP allele developed 3.4 ± 2.0 tumors (N=25) per rat while rats homozygous for the WF allele developed 7.9 ± 3.7 (N=13) tumors per rat. This is a ~56% reduction in tumor multiplicity for animals bearing the N3 COP allele and is similar to the ~58% reduction exhibited by rats carrying the line T COP allele. Additionally, line N3 heterozygotes, only carrying one N3 COP allele, developed 5.5 ± 3.6 (N=15) tumors per animal suggesting that the N3 COP resistance allele exhibits no dominance over the WF susceptible allele.

I was unable to define a precise distal end to the *Mcs1b* interval. Comparison of microsatellite DNA and published rat SNPs located in the 0.66 Mb of genomic sequence between the distal and proximal ends of lines N3 and I4 yielded no genetic variation between resistant COP and susceptible WF alleles (Tables 3 and 4). It is possible that polymorphisms between WF and COP exist in this 0.66 Mb region and have yet to be found; therefore, there remains a potential area of recombination from marker *g2UL1-5* to marker *g2Uwm65-18*. Taken together these data delineate the *Mcs1b* locus to a 1.7 Mb region spanning from marker *D2Rat142* at position *chr2:*42318464 to marker *g2UL2-29* at *chr2:* 44325564.

Targeted Resequencing of the WF and COP Mcs1b Locus Reveals New Polymorphic Markers and Tightens the Mcs1b Interval

Table 3:

Primer Sequences for Microsatellites in 0.66 Mb region of Mcs1b locus

Samuelson Lab ID	Public ID	Ensembl ID	Forward Primer Sequence	Reverse Primer Sequence	Position*	Variant
AU046380	AU046380	AU046380	GCCACCATTGTTATCTGACACA	CTCAGTGTGAGACCATGGTTCA	43313765-43313989	WF=COP†
g2UL2-22	12324055		CAAGCAACTGAGGGAGATAC	AGAGGCAATCAATTTGAAGA	43531384-43531546	WF=COP†
g2UL2-13	12324048		GTGTATGTTTAGGGGGTGAA	CCCAAAGATATGGTGAAGAA	43536766-43536909	WF=COP†
g2UL2-23	12324056		GGACACATGAGCCAGTATTT	CTTGGAGACGACTGAAACTC	43541442-43541637	WF=COP†
g2UL2-1	12324044		TACACTTGAGCAAGGACACA	ACAAGCCTGTTGTTGTAGGT	43598558-43598706	WF=COP†
D2RAT267	D2RAT267		CGCAGGAGAAACGCCTTATA	CCTCAGTTTTCATCAGCCTG	43605867-43605989	WF=COP†
D2Rat254	D2Rat254		AGCATGACCAAGACATTCCA	CGCCATGGAGAGAGATGCTA	43624817-43625068	WF=COP†
g2UL2-14	12324049		CGCTCTCTCTCTCATACACAC	CCTAAGACACTGGCTGAGAC	43659242-43659460	WF=COP†
g2UL2-2	12324054		GGAGGTTCATTATGTTGTGG	CTTTCTCTTCTCTCCCCATT	43697612-43697849	WF=COP†
g2UL2-15	12324050		TTGGACCTAACACCTAGCAT	ATAGGTCGAGATGGAAAACA	43704760-43704927	WF=COP†
g2UL2-16	12324051		GCCAGATGTAGTGGCATATT	GTTTGACTTTTGGGACAAAT	43753048-43753283	WF=COP†
g2UL2-3	12324061		GCAGGCAAAACATCTATACAC	ACTATGGGTAGAAACGCAAA	43751704-43751942	WF=COP†
g2UL2-24	12324057		CTTCCATACCTGACACACCT	AACATGACAGCTTTGGAAAT	43840761-43841043	WF=COP†
g2UL2-4	12324062		GTAGGGGCTACAAAGGAAGT	CCCAGGCTACACCATTATTA	43845450-43845697	WF=COP†
g2UL2-25	12324058		GGTGACTGGTGACAAAGATT	GTCAGCTTTTGAAGGAACAC	43911826-43912026	WF=COP†
g2UL2-5	12324063		TCTGGTTCAAAGAAGACCTG	AGGAGTTTTGGAGTGTCCTT	43934522-43934761	WF=COP†
g2UL2-6	12324064		AATCCCCTTCTCATTCTTTC	GCATCTTAAGGTCTGGAGTG	43935644-43935818	WF=COP†
g2UL2-7	12324065		TCTGTTCAGTGAGAGATCCA	AGTCAGGACAAAGATACACAAG	43951932-43951951	WF=COP†
g2UL2-26	12324059		AGCAGCTGGTATAGAAAAGGT	GCACACTTGAGAGTGAGTGA	43969019-43969339	WF=COP†
g2UL2-9	12324066		GGGAAGGAAAGACTGACTTC	AGTGCTTGTGATGTCCTCTC	44024519-44024723	WF=COP†
g2UL2-10	12324045		CGGATCTTTTGAGTTTGAAG	GTGGAATGGGGAAACTATTA	44033098-44033281	WF=COP†
g2UL2-18	12324052		TCGAGCTGTCTTCTGACTG	TTCTCCCTTCATCATGTCTC	44089085-44089302	WF=COP†
g2UL2-11	12324046		CGCTAAACCTGTCAACCTAC	AATAAAACCCACCACACAAC	44118131-44118361	WF=COP†
g2UL2-19	12324053		TTCCCACTTCTTCATTCCTA	GTCTGAACCCTACCATGAAA	44128126-44128291	WF=COP†
D2Rat298	12324043		TGCAACTATCTATGCCAGTTGG	CAGGACAGGCAGAAGATGCT	44149293-44149181	WF=COP†
g2UL2-12	12324047		GAGGAGTCATCGACTGGTTA	CCTTGCTCTCTTTCAAATTC	44164273-44164510	WF=COP†
g2UL2-27	12324060		TAAATGTGGTTTCCTTTGCT	TCAACGTAGCTGAAATTGTG	44195286-44195382	WF>COP†
D2Got27	D2Got27		ACCGTATGCACTTGATTTACAGAT	CCATTAAGTTGTCCTCTGCCC	44205860-44205910	WF<COP†

BN, Brown Norway; COP, Copenhagen; WF, Wistar Furth

*Position is *Rattus norvegicus* genome build version 3.4

†Microsatellite DNA genotypes were determined using 3% high resolution agarose

Table 4:

SNP markers tested on WF and COP rats based on published datAof SNPs in the Mcs1b interval

Samuelson lab ID	PubliCID	Ensemble ID	Forward Primer Sequence	Reverse Primer Sequence	Position*	BN (reference)	COP	WF
RN02UL-A12t			AAAATGCAGACCCATGTGC	TGATGCCCTTTCATTGTCC	42142131	C	C	T
RN02UL-A12u			GCAAAGGTTTCATCTGTTTTC	GCTACAAGCAGCACGTGAA	42170174	C	C	A
RN02UL-A12v+	rs63769957	ENSRNOSNP2740854	GAGCTGTCAGTGTGGCAAA	TGGAGCAAATGGCTTCAGA	42364155	A	A	C
RN02UL-A12w	rs64618233		AGTCCCTCACAAATGGTAGAA	GGCCAGGAATAATGAGCAA	42421698	A	G	G
RN02UL-A12x	rs64418578	ENSRNOSNP2785536	AATGATTGGCCCCAGTGT	AATGAACTCATCGGCCTGT	42514359	A	G	G
RN02UL-A12y	rs105808796	ENSRNOSNP1382723	GCCTTGGTGTTTACATGAGC	CAGCAATTACCGAGATGCAC	42633931	G	A	A
RN02UL-A12z	rs66162549	ENSRNOSNP2785538	TTCATTGCTACCTATGCAGTTG	AGCTATCCCACTGAATTCTCC	42679006	G	A	A
RN02UL-A12aa	rs105703830	ENSRNOSNP1382993	CTGTACCATTTCTGCCATGC	CATCATTGAATTTTGAGAAGCA	42782209	A	T	T
RN02UL-A12bb	rs105836290	ENSRNOSNP1383048	CCGAGCTTGCTTGTTTGTT	TGGAGACCCACGCTCTTT	42822772	T	C	C
RN02UL-A12cc	rs63938035	ENSRNOSNP2785541	GCCTATGATGTAATGCCCAGT	CACATGTGTGATCGCATAGC	42865664	T	C	C
RN02UL-A12dd	rs107042005	ENSRNOSNP2785542	GGCCATCTTGGACAGTGAG	TTGGATGGGGCAAAGACTA	43135289	A	G	G
RN02UL-A12ee	rs106693272	ENSRNOSNP2785543	TTTCTGGGATTGAGGACCA	CAAGCCGATCTCAGCAGTT	43252585	C	T	T
RN02UL-A12ff	rs64339116	ENSRNOSNP1383915	CTGCCAGAAGCATCTAAAAACT	AAGTGGGTGCTCTGTCTTCA	43523570	T	C	C
RN02UL-A12gg	rs107369228	ENSRNOSNP2785545	CTAATTTGTTTCTGCTTC	AATCCTATCCAAACTCTT	43544727	C	C	C
RN02UL-A12hh	rs66091881	ENSRNOSNP1384088	TGCAAACTGTTACAAAG	CCATGATTACTCCTTTT	43660982	T	C	C
RN02UL-A12ii	rs65824840	ENSRNOSNP2718470	TTAGACCAGAAACCTTC	CTGAACTTAGTTTTCCTG	43699296	A	G	G
RN02UL-A12jj	rs8143555		CCAGGAAATTAACACTT	GCTCCATGATACATCC	43836007	C	T	T
RN02UL-A12kk	rs8171129	ENSRNOSNP2785549	ACAAAGTCTTCTCAAGC	GTCCCATGTTTCTAGTTA	43838288	C	T	T
RN02UL-A12ll	rs107117602	ENSRNOSNP2785550	CTTAGCAACATACCTTCT	CTACGATGAGTTGTCTTT	43918079	G	A	A
RN02UL-A12mm	rs65844585	ENSRNOSNP1384598	CATCATTAGTCACTGCTT	CCAACAGTTTTAGTTTCT	43987419	G	A	A
RN02UL-A12nn	rs66189322	ENSRNOSNP2731297	GATGCATGGTATATAGTGT	GATGTTAAAACCATTAGG	44180666	C	C	C
RN02UL-A12oo+	rs107066908	ENSRNOSNP2785553	GCATTGACACTTGTTTA	AGCTATGTGTATTGGTTC	44210175	C	C	G
RN02UL-A12pp	rs64021638	ENSRNOSNP1384967	CGTGGTTGTATTCTAACT	CAGTAGTCACTCCACATT	44242105	C	T	T

Table 4 continuted

Samuelson lab ID	PubliCID	Ensemble ID	Forward Primer Sequence	Reverse Primer Sequence	Position*	BN (reference)	COP	WF
A48-ENSRNOSNP-1383684	rs105214022	ENSRNOSNP1383684	AGCCTGCACTGCCAACGTA	TGGTGGAATGCACCGCTAC	43294777	C	T	T
A48-ENSRNOSNP-1383696	rs106710684	ENSRNOSNP1383696	CTCTCCGGCAGTCACAGGTT	TGGAGACTGGGAATGCAAATG	43308266	T	C	C
A48-ENSRNOSNP-1383896	rs64839173	ENSRNOSNP1383896	TCAGGGCGAACTGAGAAAGG	GCGTCACAGCAAACATCTTCC	43507047	G	A	A
A48-ENSRNOSNP-1384351	rs64812820	ENSRNOSNP1384351	GAACCCCTTCCTCGCTTGAT	TGTGGCTGTGGGTCACCA	43843704	A	G	G
A48-ENSRNOSNP-1384874	rs65209285	ENSRNOSNP1384874	CTGTGCCAGCAGGCTAAGGT	AGCCTAGGTGGCATGTGATGA	44159855	C	T	T
A48-ENSRNOSNP-1384880	rs66201784	ENSRNOSNP1384880	CATGTCGTGCTTCCTGTCCTT	TCAGGAAGTCATGGAGGTGAGAC	44164060	A	G	G
A48-ENSRNOSNP-2645369	rs107488981	ENSRNOSNP2645369	TTCAAAATGGCGGAGTGCTT	TCTGGGAGTTCAGTGCTGTCTG	42712802	T	C	C
A48-ENSRNOSNP-2654615	rs64619323	ENSRNOSNP2654615	AATACAGGCGGTGGATAATGAAGA	AGCCTGGTCCAGTGTCTCAATC	43669623	A	G	G
A48-ENSRNOSNP-2657379	rs65363226	ENSRNOSNP2657379	TTCGGTGTCTACGTGGTTAGCA	CCTGCAAATATCTGGCATGTACTTC	43815443	A	G	G
A48-ENSRNOSNP-2685221	rs66162549	ENSRNOSNP2685221	CATTGCTACCTATGCAGTTGGA	GACGTGAGCGCATACAAATC	42679006	G	A	A
A48-ENSRNOSNP-2686459	rs63770151	ENSRNOSNP2686459	CAGTGTCTCCACCGACTGACC	CTTTGCCCCGATACATCCAC	43529896	C	T	T
A48-ENSRNOSNP-2705085	rs65463498	ENSRNOSNP2705085	AGTGGGCGCTACCTTCAACA	GGCCATTGAGTCGGTGGTTA	43418586	C	T	T
A48-ENSRNOSNP-2706005	rs66268295	ENSRNOSNP2706005	GGAGCTAAGCGGTGTGAGGA	AGCCTTCAAGGATGGGGAAG	43517914	G	A	A
A48-ENSRNOSNP-2706295	rs66174360	ENSRNOSNP2706295	AATGCCCAACATCCCTTCCT	CATGCCACACCCAAACATTG	43377935	A	A	A
A48-ENSRNOSNP-2708093	rs65540697	ENSRNOSNP2708093	TAGGCACCGAGAAGCCACAT	AGGGACTCCTGGAAGGGATG	43418057	G	C	C
A48-ENSRNOSNP-2744110	rs64996390	ENSRNOSNP2744110	AGCATGCCGTGACTCTGAAT	CAAGAGGTTCTTGGTTCTTTCA	43006013	C	T	T
A48-ENSRNOSNP-2769457	rs63803615	ENSRNOSNP2769457	AAGAGTCCAGGTAGGGTGGAG	GCACTAAGCGATACCCATGA	43376687	C	C	C
A48-ENSRNOSNP-2753529	rs64232116	ENSRNOSNP2753529	TCTTGGGTACACTGCCCACA	ACGGAGGGTTTTGCCTGTTT	43753769	A	T	T
A48-ENSRNOSNP-2778924	rs106934580	ENSRNOSNP2778924	ACTGGCTTGGCTGTTGGTGT	TGGTGGGCTAAAAGGGAGGT	43191996	T	C	C

Abbreviations: BN, Brown Norway (reference) COP, Copenhagen; WF, Wistar Furth

*Position is based on *Rattus norvegicus* Chromosome 2, genome build version 3.4

†Informative markers used in narrowinGMcs1b locus

Sources: RN02UL markers, Nijman eTal. BMCGenomics 2008; A48 markers, STAR Constortium database

Up to this point, all mapping had been performed using microsatellite markers. SNPs between WF and COP have not been tested and new markers were needed in the region to effectively fine-map the *Mcs1b* locus. I resequenced targeted regions in the newly delineated Mcs1b interval from *D2Rat142* to *g2Uwm65-18*. Cuppen and colleagues published a panel of SNP markers for the rat genome consisting of 820 different SNP assays tested in 34 different rat strains (Nijman et al., 2008). SNPs assayed in the study were chosen based on being polymorphic between BN, Wistar and Dahl/Salt Sensitive (SS) rats. Of the 820 SNPs tested, 22 lie within the *Mcs1b* locus, making this a good tool to identify polymorphisms between WF and COP alleles (Table 4). Although the Cuppen panel was tested on 34 different rat lines, the COP strain was not included. Further, WF rats were tested but they were not the Harlan (WF/Hsd) strain used in my study. It was therefore, necessary to resequence WF and COP gDNA at these 22 sites to determine if they were polymorphic.

Additionally, a rat SNP map has been developed by the Specific Targeted Research Project (STAR) Consortium (Saar et al., 2008). The consortium sequenced >100,000 SNPs in 60 different rat strains, including COP/Hsd; however, WF/Hsd was not included. To identify potential polymorphisms between WF and COP sequences I identified 20 non-redundant SNPs from the STAR Consortium database lying within the rat *Mcs1b* locus (Table 4). These were tested in WF and COP gDNA samples by Sanger sequencing.

Of the 22 Cuppen SNPs tested, 4 were identified to be polymorphic between WF and COP alleles. These were termed *A12t*, *A12u*, *A12v*, and *A12oo*. Markers *A12t*, *A12u*, and *A12v* lie relatively close to each other at positions *chr2:42142131*, *chr2:42170174*

and *chr2:42364155* (Figure 5). Marker *A12oo,* however, lies much further downstream near the proximal end of line I4 at position *chr2:44210175* (Figure 5). Sequence results of SNPs from the STAR Consortium database revealed no polymorphisms between WF and COP at any of these markers. All newly-identified markers will be useful in future mapping studies to narrow down the *Mcs1b* locus.

I also attempted to identify new microsatellite markers by amplifying regions of sequence containing di- and tri-nucleotide repeats, which are typical of microsatellite polymorphisms. Results of these studies uncovered no polymorphisms between WF and COP sequences. In addition, large regions of gDNA were resequenced in these areas across the *Mcs1b* interval to identify potential SNPs. The results of this study showed no polymorphisms between the two rat strains (Table 5).

The new SNP markers were tested on rats from congenic lines U2, W2, N3 and I4. Lines U2 and W2 tested homozygous for the COP allele at *A12v* and homozygous for the WF allele at the more distal SNPs. Line I4 tested homozygous for the COP allele at *A12oo* and homozygous for the WF allele at the proximal SNPs. Line N3 tested homozygous for the WF allele at *a12oo*. Taken together, these markers slightly shorten the *Mcs1b* interval to a region from *A12v* to *A12oo* (Chr2: 42364155 – 44210175).

Backcrossing Line T Generates a Novel WF.COP Congenic Line: 1b-11

There are 10 annotated transcripts that lie in the newly-defined *Mcs1b* locus (Figure 5). Also, there is a region of potential recombination between markers *g2UL1-5* and *A12oo*. Further mapping could eliminate some of these transcripts and reduce the number of potential gene candidates that need to be tested. To do this, I attempted to

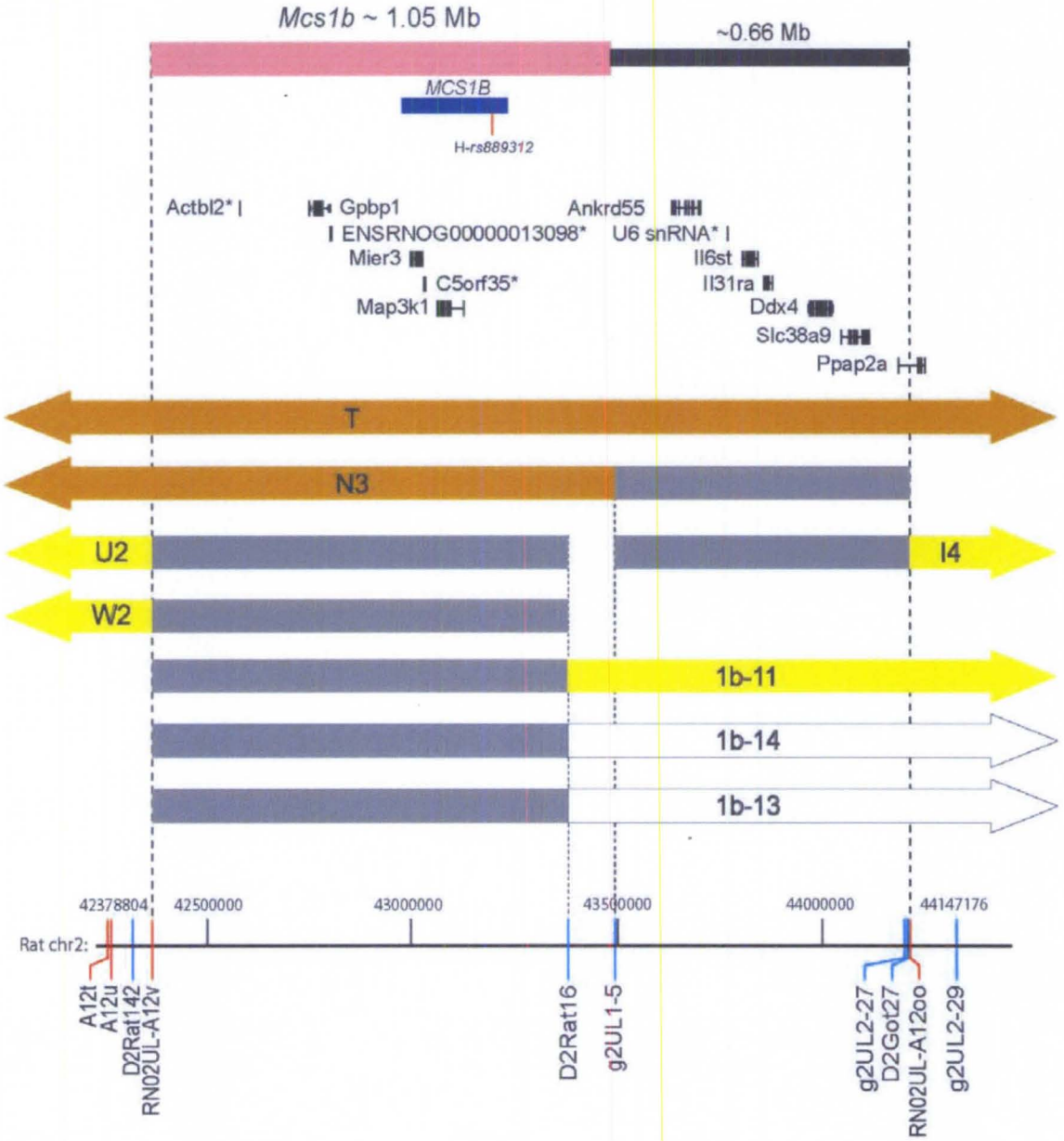

Mcs1b ~ 1.05 Mb
~0.66 Mb
MCS1B
H-rs889312
Actbl2*
Gpbp1
ENSRNOG00000013098*
Mier3
C5orf35*
Map3k1
Ankrd55
U6 snRNA*
Il6st
Il31ra
Ddx4
Slc38a9
Ppap2a
T
N3
U2
I4
W2
1b-11
1b-14
1b-13
42378804
42500000
43000000
43500000
44000000
44147176
Rat chr2:
A12t
A12u
D2Rat142
RN02UL-A12v
D2Rat16
g2UL1-5
g2UL2-27
D2Got27
RN02UL-A12oo
g2UL2-29

Figure 5: Genetic landscape of fine-mapped rat *Mcs1b* locus. ■ and ■ represent regions with a COP genotype with resistant and susceptible Mcs phenotypes, respectively. □ represents regions with the COP genotype where the Mcs phenotype is not known. ■ denotes regions of potential recombination where no markers are currently known. Red and blue lines demarcate the location of informative SNP and STS markers used to genotype the locus. The blue bar depicts the rat interval orthologous to the human breast cancer-associated locus; SNP *rs889312* and SNPs in LD are shown. Annotated transcripts obtained from UCSC Genome database shown in black.

Table 5:

Regions sequenced between Rat Chr*2*:43125940 and Chr*2*:43632545 to identify potential polymorphisms between WF and COP in the *Mcs1b* locus

Total conserved regions attempted to sequence	194
Regions Successfully Sequenced	142
Regions successfully sequenced with results that match reference gDNA files	103
% Regions sequenced resulting in both WF and COP sequences matching reference	24%
% Regions sequenced resulting in either WF or COP sequences matching reference	54%
Total potential polymorphic regions identified	86
Potential Microsatellites	4
Potential Insertion/Deletions	27
Potential SNPs	55
Potential Polymorphisms validated	0

generate a new congenic line containing a COP allele from the end of line N3 at microsatellite marker *g2UL1-5* and spanning marker *A12oo* and beyond. I backcrossed line T heterozygous rats to homozygous WF rats to generate novel unique recombinants containing this region. 268 progeny were tested from these crosses and generated 3 new recombinants. These recombinant animals contained COP alleles spanning markers *g2UL1-5* to *g2UL2-30* and were termed lines 1b-11, 1b-13, and 1b-14 (Figure 5). These recombinants were backcrossed to expand the population and progeny were genotyped. Pups containing the recombinant COP allele were inbred to establish a homozygous fixed congenic line. Due to complications with attaining recombinant pups and insufficient litter sizes, only 1b-11 was established as a new congenic line. This line was subsequently phenotyped as described.

Mammary Carcinoma Susceptibility Phenotype of WF.COP Congenic Line 1b-11 Is Not Different Than Homozygous WF Controls

Female rats homozygous for the 1b-11 COP allele and homozygous WF controls were administered DMBA at 50-55 days of age and euthanized at 15 weeks post-DMBA administration. Female 1b-11 rats developed 5.9 ± 2.5 (N=12) tumors per rat compared to 6.4 ± 1.9 (N=8) tumors for WF controls (Table 2, Figure 6). The difference in tumor multiplicity between the two lines is not statistically significant (P=0.4179). These data suggest that there is no difference in the mammary carcinoma susceptibility phenotype between 1b-11 congenic rats and WF rats.

WF.BN Congenic Rat Strain/Line Is Resistant To Developing Mammary Tumors

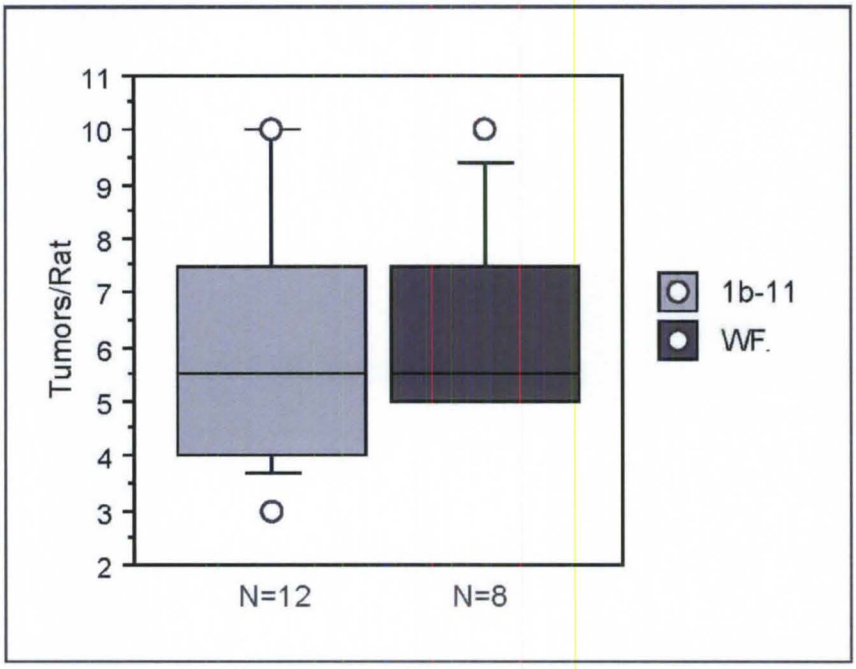


Figure 6: Endogenous mammary tumors formed in congenic line 1b-11 and homozygous WF animals. Tumors presented as box plots. Circles represent outlier datapoints. N value represents number of rats tested.

BN rats are almost completely resistant to chemical, radiation and oncogene-induced mammary carcinogenesis. The Mcs1b locus was identified between WF and COP rat strains and it is not known whether this locus is functioning in the BN rat to reduce mammary carcinoma susceptibility. If it is, the BN rat may be another useful strain to study the *Mcs1b* QTL.

To address this, I used a WF.BN congenic line. WF.BN congenic rats were developed by introgressing a BN allele at the *Mcs1b* locus onto a WF background using the method described previously. Female WF.BN N4F1 homozygous rats were phenotyped along with homozygous WF controls. WF.BN N4F1 homozygous rats developed 2.3 ± 1.5 (N=11) tumors per rat while WF controls developed 5.4 ± 2.6 (N=28) tumors per rat (Figure 7A). The difference between these means is significant (P=0.0012) and it suggests that the mammary carcinoma resistance exhibited by the BN rat strain may be partly conferred by heritable elements lying within the *Mcs1b* locus. To be certain there was no difference between WF.BN N4F1 rats carrying a WF allele at *Mcs1b* and inbred homozygous WF rats, these groups were also compared (Figure 7B). There is no statistical difference (P=0.2039) between congenic WF.BN N4F1 rats homozygous for the WF allele at Mcs1b and inbred WF rats; therefore, these could be pooled as susceptible controls.

With the knowledge that there are two other *Mcs* loci on rat chromosome *2*, it was necessary to define the full interval of the BN allele. To determine the length of the *Mcs1* locus within the WF.BN locus, the N4F1 rats were genotyped at markers along the original *Mcs1* locus (Figure 2, line B; Table 1). WF.BN N4F1 rats had a BN allele spanning markers *D2Uwm17* to *D2Rat210* (*chr2:*32051319-82193231) delineating a ~50

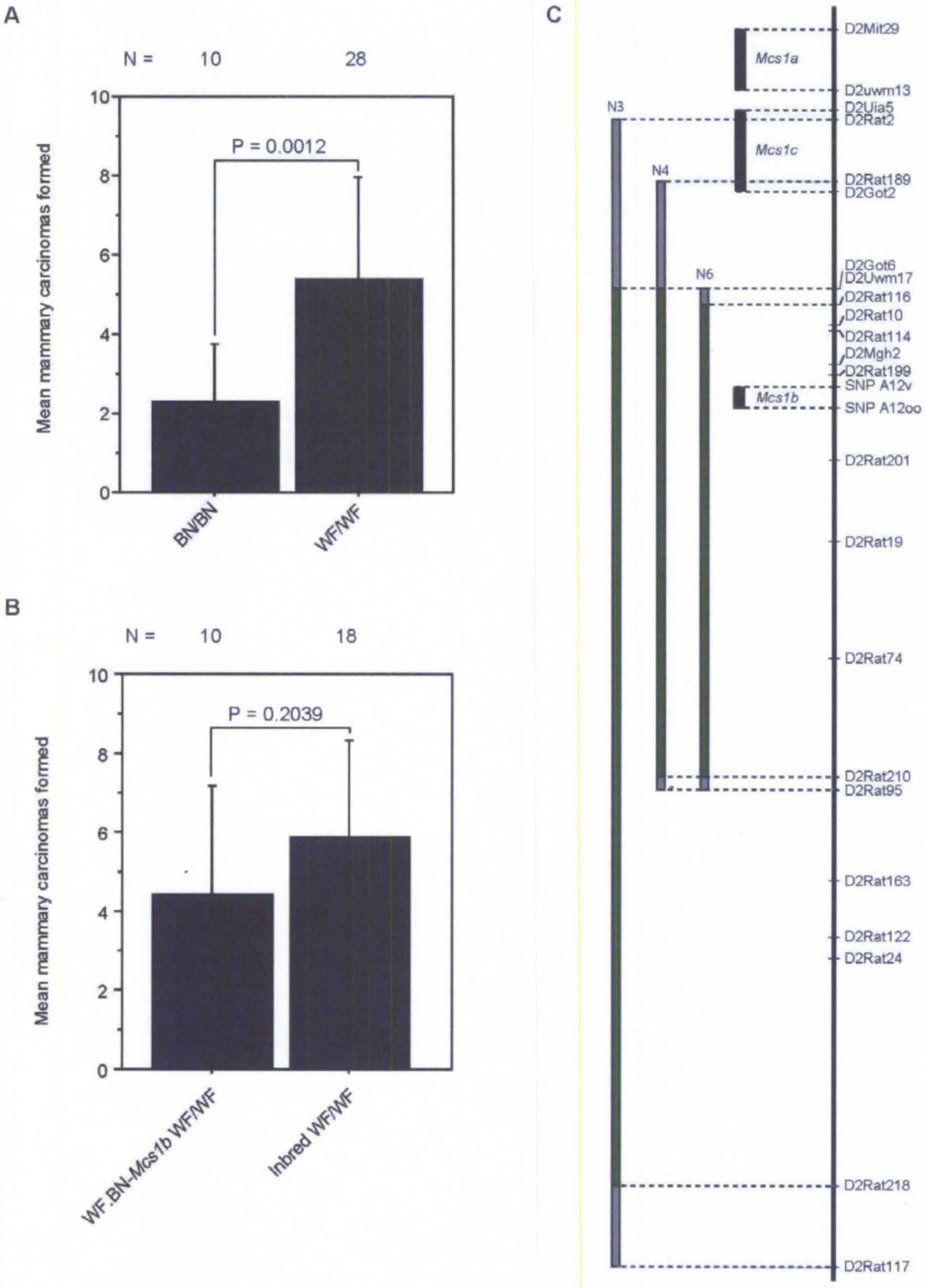

A
N = 10 28
Mean mammary carcinomas formed
P = 0.0012
BN/BN
WF/WF
B
N = 10 18
Mean mammary carcinomas formed
P = 0.2039
WF.BN-Mcs1b WF/WF
Inbred WF/WF
C
D2Mit29
Mcs1a
D2uwm13
N3
D2Uia5
D2Rat2
Mcs1c
N4
D2Rat189
D2Got2
D2Got6
N6
D2Uwm17
D2Rat116
D2Rat10
D2Rat114
D2Mgh2
D2Rat199
SNP A12v
Mcs1b
SNP A12oo
D2Rat201
D2Rat19
D2Rat74
D2Rat210
D2Rat95
D2Rat163
D2Rat122
D2Rat24
D2Rat218
D2Rat117

Figure 7: Tumor multiplicity in WF.BN congenic lines. A; Average mammary carcinomas formed in congenic WF.BN N4F1 and homozygous WF control female rats after being administered DMBA. Error bars denote standard deviation. Data was analyzed using Mann Whitney test. B; Average mammary carcinomas formed in congenic WF.BN female rats compared to inbred WF female rats. C; Map of WF.BN congenic lines at N4 and N6 generations. ■ denotes a BN genotype with a resistant Mcs phenotype. ■ denotes regions of potential recombination. Black bars designate Mcs loci 1a, 1b, and 1c previously identified by Haag *et al.*

Mb region (Figure 7C). There is a region of potential recombination at the proximal end of the interval that overlaps with the *Mcs1c* locus; therefore, the *Mcs1c* locus may be contributing to mammary carcinoma resistance in this line. Thus, I again attempted to reduce the interval by backcrossing the WF.BN congenic rats to WF rats up to the N6 generation and these animals were again genotyped to identify the ends of BN allele. The N6 generation contained the BN allele spanning markers *D2Rat116* to *D2Rat210* (*chr2:*33845498-82193231) shortening the interval to ~48 Mb and eliminating overlap with the *Mcs1c* locus. I attempted to further breed these animals out to the N10 generation to eliminate the long stretches of BN genomic sequence flanking the *Mcs1b* locus; however, due to complications, the WF.BN congenic line was lost. It was decided that it was not feasible to pursue reestablishing the WF.BN congenic line.

Discussion

The *Mcs1b* locus was previously delineated to a ~15 Mb interval bounded by markers *D2Uwm17* and *g2UL2-30* on rat chromosome *2* (Haag et al., 2003)(Figure). More than 50 transcripts lie in this region. To shorten this list and reduce subsequent comparative genetics work I sought to fine-map this locus using a combination of WF.COP and WF.BN congenic rats.

Five congenic lines were developed using a similar approach to what was used previously to identify the *Mcs1* and its subloci: *Mcs1a*, *b* and *c*. Lines F3, U2, W2, I4 and N3 all contain COP alleles spanning various portions of the original line T background genome (Figure 4). Testing these lines for mammary carcinoma susceptibility enabled me to narrow the *Mcs1b* locus. The tumor multiplicity data

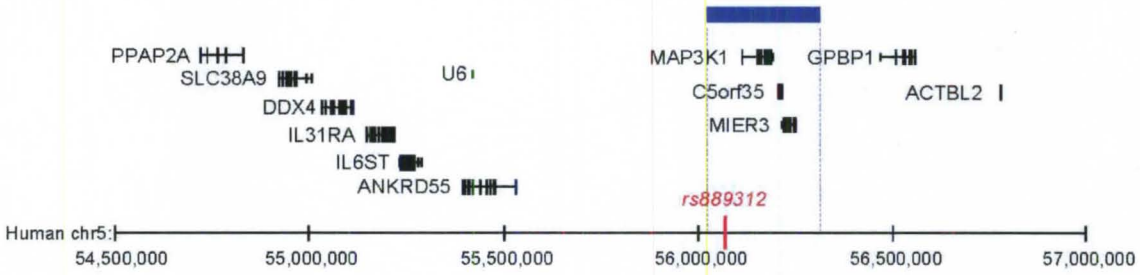


Figure 8: Map of human *Mcs1b* orthologous locus on human chromosome *5*. Base positions are labeled at the bottom moving proximal to distal from left to right. The blue bar denotes 280 kb haplotype block containing SNP *rs889312*. The genomic position of SNP *rs889312* is shown in red. Transcripts are shown with exons designated by vertical bars.

obtained from lines U2, W2, I4 and N3 delineate the *Mcs1b* locus to a ~1.8 Mb interval on chromosome *2* between markers *A12v* and *A12oo*. This markedly reduces the number of transcripts to be functionally tested to thirteen.

As anticipated, the rat ortholog to the human breast cancer-associated 280 kb haplotype block lies within this 1.8 Mb region. It is likely that a heritable element(s) lying in this region is controlling rat mammary carcinoma susceptibility in a manner potentially analogous to the human orthologous locus. Therefore, the rat may be used to identify the mechanism of action and translate this to human disease susceptibility. The human block contains 2 known transcripts, *MIER3* and *MAP3K1*, as well as a predicted gene transcript, *C5ORF35* (Figure 3). Nearby, but lying outside of the human haplotype block associated with *rs889312*, are approximately 7 other annotated transcripts (Figure 8). The rat contains orthologs to these transcripts within the *Mcs1b* locus although the orientation of the rat *Mcs1b* locus is reverse of the human locus (Figure 5). I anticipate that one or a combination of the three transcripts lying inside the human haplotype ortholog is causative for the mammary carcinoma susceptibility phenotype. However, all of the transcripts within the *Mcs1b* locus, including those lying outside the human haplotype block orthologous interval, must be tested as they cannot be ruled out as causal.

The *Mcs1b* locus was mapped with the use of polymorphic microsatellite markers published in the RGD public database (Dwinell et al., 2009). However, many of these were not informative between WF and COP genotypes. Targeted resequencing of regions in the narrowed *Mcs1b* locus uncovered 4 SNPs polymorphic between the WF and COP strains. Using these new SNP markers, the *Mcs1b* locus was mapped to a

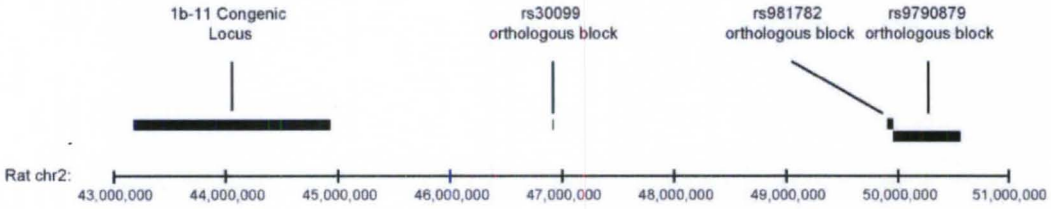
shorter interval between markers *A12v* and *A12oo*. Also, short genomic regions rich in di- and tri-nucelotide repeats within the *Mcs1b* interval were sequenced; however, no polymorphisms were identified between WF and COP alleles. Taken together these data suggest that WF and COP share similar haplotypes. However, since these newly identified SNPs are positioned at the ends of the *Mcs1b* interval, it is unlikely that they are causative; therefore, more polymorphisms likely exist in this region.

Breeding of line T led to the development of another congenic line, 1b-11. Line 1b-11 has a COP genotype from marker *D2Rat16* to marker *g2UL2-30*. This interval crosses the 0.66 Mb region from *g2UL1-5* to *A12oo* making it useful in identifying if this area of potential recombination is contributing to the resistance phenotype (Figure 5). Tumor multiplicity between homozygous 1b-11 female rats and homozygous WF littermates was not statistically different (Table 2). These data imply that this 0.66 region from D2Rat16 to A12oo does not contain the element responsible for the *Mcs1b* phenotype. It stands to reason, then, that the element conferring resistance to mammary carcinoma development resides in the interval from *A12v* to *D2Rat16*, thereby narrowing the *Mcs1b* locus to ~1.01 Mb.

These data signify that a more complex mechanism may be at work. Tumor multiplicity data for line I4 and WF littermate control rats was 9.3 ± 3.0 and 7.9 ± 3.7, respectively. Although the difference is not significant, it suggests a trend that may indicate that the mechanism controlling susceptibility in the *Mcs1b* locus is much more complex than originally thought. A resistant phenotype could be compensated for by the action of another risk allele lying outside of the *Mcs1b* locus on the distal side. Indeed, the GWAS by Easton *et al.* reported two additional alleles identified upstream of SNP

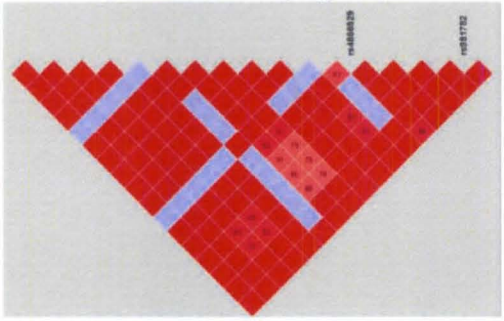
rs889312 and an additional breast cancer-associated allele was identified on human chromosome 5 in a subsequent breast cancer-GWAS (Easton et al., 2007; Turnbull et al., 2010). SNPs *rs981782* and *rs30099* reported by Easton *et al.* are positioned on human chromosome 5 at bases *45321475* and *52454339*, respectively. SNP *rs981782* has a minor allele frequency of 0.47 in the SEARCH population studied and associates with a reduction in breast cancer incidence with an OR of 0.92 (95% CI=0.87-0.97) when homozygous for the minor allele (Easton et al., 2007). SNP *rs30099* has a minor allele frequency of 0.08 in the SEARCH population and associates with an increase in breast cancer incidence with an OR of 1.09 (95% CI=0.96-1.24) when homozygous for the minor allele (Easton et al., 2007). The magnitude of the affect associated with these loci is much smaller than what is seen for SNP *rs889312*. Regardless, the results from the 1b-11 phenotype suggest that an additional risk allele may be present in the 1b-11 locus. To address this, I used a bioinformatics approach to identify rat orthologs of human SNPs *rs981782* and *rs30099*. I determined that LD blocks associated with these SNPs have orthologs on rat chromosome *2* (Figure 9A). SNP *rs981782* lies in a 56kb haplotype block on human chromosome 5 and is in LD with one other SNP, *rs4866929* (Figure 9B). This haplotype block has a rat ortholog that spans from base pairs (bp) 49907723 to 49954324 on chromosome 2 (Figure 9A). SNP *rs30099* is also in LD with only one other SNP, *rs30727* and lies in a 3kb haplotype block on human chromosome *2* (Figure 9C). The haplotype block has a rat ortholog that lies on chromosome *2* from bases 46916997 to 46918680 (Figure 9A). SNP *rs9790879* reported by Turnbull *et al.* lies within a 488 kb haplotype block on the *p* arm of human chromosome 5 and is in LD with approximately 100 other SNPs (Figure 9D). There is a rat ortholog to this block

A



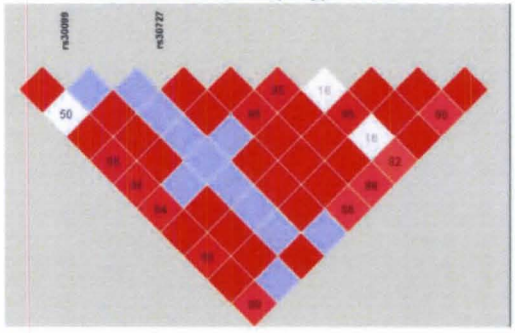
B

rs981782 Human Haplotype Block



C

rs30099 Human Haplotype Block



D

rs9790879 Human Haplotype Block

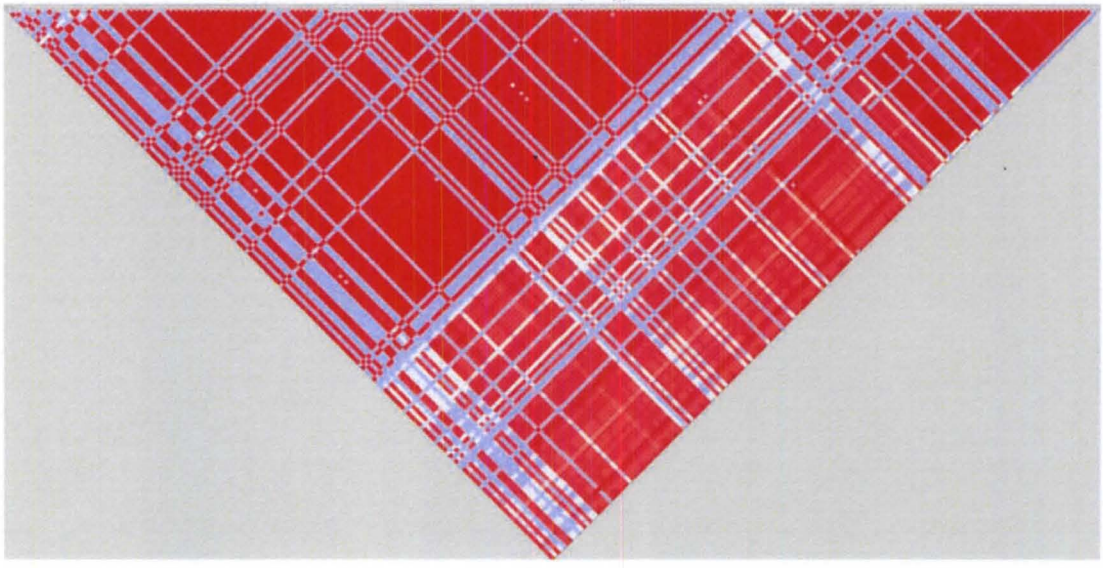


Figure 9: Additional breast cancer-associated loci identified in GWAS with rat chromosome *2* orthologs. A; Map of line 1b-11 interval on rat chromosome *2* showing positions of blocks orthologous to human haplotype blocks in LD with other breast cancer-associated SNPs. B – D; Haploview images of human haplotype blocks associated with individual breast cancer associated SNPs. Lower panel shows pairwise r^2 values for LD; light blue represents lower disequilibrium, white is intermediate, and red represents strong disequilibrium. An r^2 threshold of 0.80 was used to determine SNPs in LD. SNPS in LD are shown for *rs981782* and *rs30099*. There are too many SNPs in LD with *rs9790879* to maintain resolution; therefore, these are not shown.

delimited by bp 49961197 and 50564367 (Figure 9A). When mapped to the rat genome, none of these rat orthologs lie within the 1b-11 congenic region (Figure 9A). This negates the notion that there is a second allele that is offsetting the affect of the *Mcs1b* locus. However, this is based on the assumption that there must be a human ortholog. It is possible that a second risk allele is present in this region but is specific to the rat *or* that an ortholog to a human risk allele lies within the 1b-11 locus yet unidentified in human studies. Testing this idea would require the development of new congenic lines containing varying portions of the distal end of line 1b-11 to define the specific region for this phenotype. However, this would be meaningless to human breast cancer susceptibility if the effect is rat-specific.

Another potential explanation for the susceptible phenotype seen in the 1b-11 line is that *Mcs1b* is a compound locus, i.e., resulting from the action of two or more variants working in concert to give rise to the phenotype. Line N3 shows a ~56% reduction in tumor multiplicity compared to WF controls (Table 1). This indicates that the N3 allele contains the necessary variants required to give rise to the resistant phenotype. The 1b-11 interval overlaps the distal end of the N3 interval. Therefore, the 1b-11 allele could harbor a variant that is insufficient alone and requires an additional variant lying upstream. This variant would be present in N3 but lost in 1b-11, thereby eliminating the resistance phenotype in 1b-11 congenic rats. Work here is not conclusive, but other complex QTLs were identified that operate this way (Samuelson et al., 2007). To further study this potential mechanism would require modifying the current approach by incorporating cross-breeding of different congenic lines to determine whether the resistant phenotype can be restored, i.e., breeding of U2 or W2 rats to 1b-11 rats and

phenotyping offspring containing both U2/W2 and 1b-11 alleles. If *Mcs1b* is a complex locus, these offspring should exhibit a resistant phenotype.

Gould and colleagues show that line T heterozygous female rats average 7.6 ± 0.8 (N=18) mammary carcinomas per rat (Haag et al., 2003). This result was not statistically different from WF-homozygous littermates (8.3 ± 0.8, N=18) but was significantly increased (P<0.0001) compared to COP-homozygous line T rats (3.5 ± 0.5, N=21). These data suggest that the WF allele at the line T locus exerts a dominant phenotype in a Mendelian sense. On the other hand, Line N3 heterozygous and COP-homozygous littermates develop 5.4 ± 3.6 (N=15) and 3.4 ± 2.0 (N=25) carcinomas per rat, respectively (Table 1). Both of these values are statistically reduced compared to WF-homozygous line N3 littermate female rats (7.8 ± 3.1, N=25). Tumor multiplicity in line N3 heterozygous rats is approximately half compared to COP-homozygous littermates signifying neither the WF nor the COP allele is acting dominantly. Taken together, these figures show that there is a decided difference between the line T and N3 alleles. This substantiates the notion that there is a more complex mechanism occurring at this locus. The simplest explanation is that there is an epistatic allele placed in either the proximal or distal regions flanking the N3 allele; however, other mechanisms may be at play. A compound locus could exist as described above. Moreover, it may be that there is a cis-interaction requiring a second element that is present in the line T interval but is absent in line N3. My data are insufficient to favor one scenario over another but suggest that the mechanism of action at this locus is much more complex than originally anticipated.

In addition to the COP rat, the BN rat strain is also almost completely resistant to developing spontaneous, chemically-induced, and oncogene-induced mammary

carcinomas. As previously mentioned, the BN sequence has been completely sequenced. It is currently not known whether the resistant phenotype exhibited by the BN rat is associated with the *Mcs1b* locus. The BN rat was tested to determine if the mammary carcinoma resistant phenotype is controlled in part by the *Mcs1b* locus. If it is, the BN could be an additional tool to help discover SNPs that may be involved in the COP *Mcs1b* phenotype. A WF.BN *Mcs1b* congenic line had been initiated and I attempted to establish this line in order to determine if the BN *Mcs1b* region confers resistance to developing mammary carcinomas in similar manner to the COP strain.

Results of WF.BN phenotype studies were to be applied to further characterize the *Mcs1b* locus in the COP line as well as determine whether the *Mcs1b* locus is involved in controlling mammary carcinoma susceptibility in the BN line. By examining the phenotype of the BN allele a 3-way haplotype analysis would rule in/out specific polymorphisms in the *Mcs1b* locus for their contribution to the mammary carcinoma resistance phenotype. The BN haplotype carries a set of SNPs that is shared with the COP allele (BN=COP at *A12v* and *A12oo*) and can, therefore, be used to develop the *Mcs1b* locus with a BN haplotype. Markers between WF and COP are infrequent in the *Mcs1b* interval; however, the BN strain has been sequenced and many more markers have been identified. Therefore, these may be used to identify markers that may be used to narrow the *Mcs1b* locus.

Unfortunately, however, the line was lost before the N10 generation and the *Mcs1b* region narrowed. Regardless, data from initial studies with the WF.BN congenic line at the N4F1 generation imply that the mammary carcinoma resistance exhibited by BN rats is, at least in part, controlled by the *Mcs1b* locus. This cannot be stated with

complete confidence as the proximal end of the WF.BN rats contains a region of potential recombination that overlaps with the *Mcs1c* locus. Therefore, the *Mcs1c* locus could be involved. Additionally, these WF.BN congenics have large intervals of BN sequence that flank the distal side of the *Mcs1b* locus. These rats potentially harbor other risk alleles that are controlling mammary carcinoma susceptibility, i.e. the rat orthologs of the human haplotype blocks associated with SNPs *rs981782*, *rs30099* or *rs9790879*. With the loss of the WF.BN congenic line further investigation of the BN phenotype is not practical at this time.

Overall, the data delineate the *Mcs1b* locus to a 1.01 Mb region flanked by SNP markers *A12v* and *D2Rat16*. The variant(s) contributing to the *Mcs1b*-conferred resistance phenotype likely lies inside this interval unless the mechanism is much more complex, as discussed. To determine the specific variant(s) contributing to mammary carcinoma susceptibility deep sequencing of this region is needed to identify new variants and test them accordingly. If no variant is identified lying inside the region delineated by *A12v* and *D2rat16*, this strongly implicates that *Mcs1b* requires additional cooperative loci *or* that the 1b-11 line contains an epistatic allele downstream.

One of the goals for this project was to shorten the interval containing the *Mcs1b* locus to minimize the list of potential gene candidates that would need to be functionally tested. This was achieved by mapping the *Mcs1b* locus to 1.01 Mb. Within the 1.01 Mb *Mcs1b* locus there are five potential gene candidates: *Actbl2*, *Gpbp1*, *Mier3*, *C5orf35* and *Map3k1*. *Mier3*, *C5orf35* and *Map3k1* all lie inside the region orthologous to the *rs889312* LD block. Although these three transcripts are the most likely candidates, any one or combination of the genes contained within *Mcs1b* may be involved in controlling

mammary carcinoma susceptibility and they must be tested further to determine their role.

CHAPTER III

RAT *MCS1B* ACTS IN A MAMMARY GLAND CELL-AUTONOMOUS MANNER TO CONTROL MAMMARY CARCINOMA SUSCEPTIBILITY

Introduction

The etiology of mammary carcinogenesis is complex, involving events that occur within the mammary parenchyma as well as events that are external to the mammary gland. In both humans and rats, mammary tumors arise primarily in the cells that line the mammary duct, the ductal epithelial cells. One would expect, then, that genetic differences that give rise to sustained proliferation and carcinogenesis would originate in these cells. Likewise, it would be likely that anti-oncogenic properties, such as the phenotype conferred by the *Mcs1b* COP allele, also lie within these cells. Indeed, there are both oncogenes and tumor suppressor genes expressed in the ductal epithelia that affect tumor development; however, as mentioned previously, different groups have shown that other cell types external to the mammary gland can affect pathogenesis (Pollard, 2004; Sica et al., 2006; Smits et al., 2011b; Trimboli et al., 2009).

Consonant with data illustrating mammary tumors originating in the mammary epithelial cell (MEC) population, it has been shown previously that the effects of the *Mcs1* COP locus are generally intrinsic to the mammary gland (Zhang et al., 1990). However, in these studies a slight mammary gland-external effect was observed, likely due to mixed effects of multiple QTLs: *Mcs1a*, *b* and *c*. It is necessary, then, to test the

Mcs1b allele independently of the other COP *Mcs* alleles. This is critical to functionally analyze potential candidate genes. For instance, results of these mammary gland autonomy studies will guide us as to what tissues or cell types to examine when assessing transcript levels or protein expression. Moreover, knowing the site of function of the *Mcs1b* allele may lend some insight into a viable gene candidate or a plausible role that a gene may be performing to control mammary carcinoma susceptibility. There are five potential gene candidates lying in the narrowed *Mcs1b* locus; *Actbl2*, *Gpbp1*, *Mier3*, *C5orf35* and *Map3k1*. Of these, only *Map3k1* has been fully characterized. Therefore, identifying the tissue or cell type that the Mcs1b locus is conferring resistance to mammary carcinoma development may provide leads as to potential functions of these transcripts in the context of mammary carcinogenesis. Further, these mammary gland cell-autonomy studies are critical in that they provide the basis for all functional studies of the *Mcs1b* locus and will potentially provide needed insight into the etiology of mammary carcinogenesis.

Design and Methods

Whole Mammary Gland Transplant Assays

Donor nulliparous female homozygous WF and line N3 WF.COP congenic rats (N=~25 animals per group) were euthanized at 30-35 days of age and their abdominal and inguinal mammary glands excised. The glands were scissor-minced and, using an IACUC-approved protocol, transplanted into the interscapular white fat pad of age-matched nulliparous female rats in a reciprocal manner; i.e. WF into WF.COP and WF.COP into WF. Also, WF into WF and WF.COP into WF.COP transplants were

performed as controls. At 50-55 days of age, recipient animals were administered DMBA dissolved in sesame oil (20 mg/ml) by oral gavage (65 mg/kg body mass) to induce mammary carcinogenesis. At 15 weeks post-DMBA administration, recipient rats were necropsied and interscapular fat pads were examined for tumor development (Figure 10). Fat pads were whole mounted onto glass slides and stored in 70% ethanol for two weeks to allow for clearing of lipids from the pad. After two weeks, the whole mounts were fixed with 1 part glacial acetic acid and 3 parts 100% ethanol for 1 hour, hydrated by serially-decreasing ethanol washes (70%, 50%, 40%, distilled water) for 15 minutes each, and stained in aluminum carmine for four days. Aluminum carmine was prepared by dissolving 2.5 g alum potassium sulfate (Spectrum Chemical, New Brunswick, NJ) and 1 g Carmine stain (TCI America, Portland, OR) in 500 ml distilled water, boiling for 20 minutes and filtering out precipitate. Once stained, whole mounts were dehydrated again with serially-increasing ethanol washes (50%, 70%, 100%) and finally a xylene wash. Whole mounts were stored long term in mineral oil. To verify properly mammary gland development in the transplant site whole mounts were examined for proper mammary gland duct elongation and TEB formation. Whole mounted fat pads were examined microscopically for frank carcinoma, ductal carcinoma *in situ* (DCIS) and hyperplasia development. Tumor outcome was analyzed using logarithmic regression analysis. DCIS and hyperplasia multiplicity data were analyzed by Mann-Whitney rank test.

Mammary Gland Cell Preparation and Injection into Interscapular Fatpad

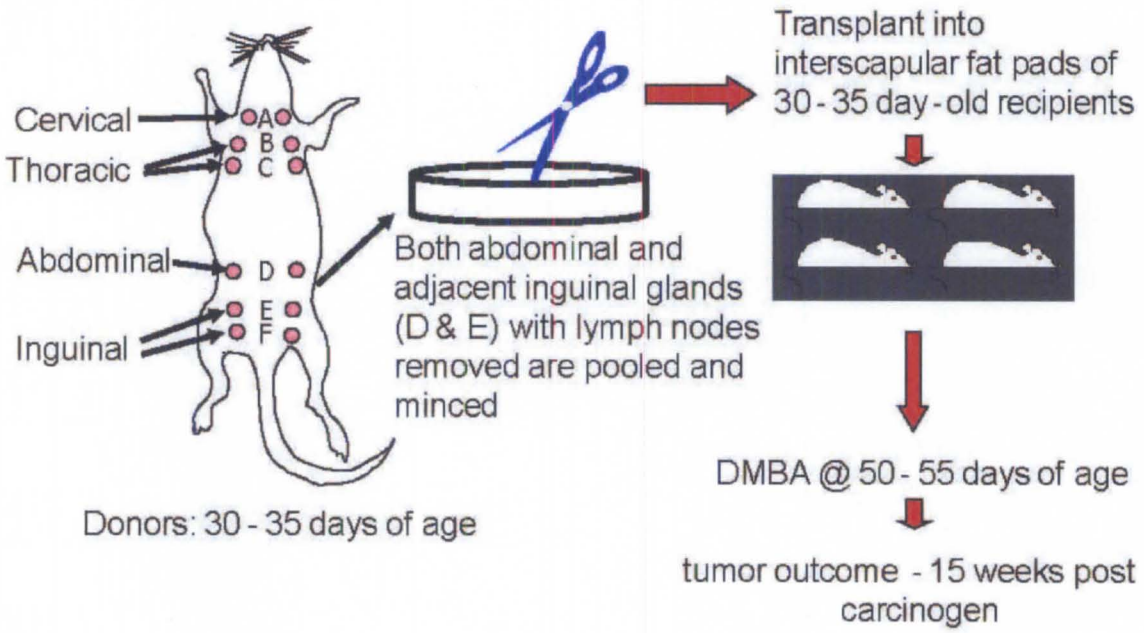


Figure 10: Flow diagram of mammary gland grafting assay.

Donor nulliparous female homozygous WF and line N3 WF.COP congenic rats (N=~25 animals per group) were euthanized at 30-35 days of age and their abdominal and inguinal mammary glands excised. The glands were scissor minced and epithelial cells were enriched by enzymatic digestion in collagenase III (Worthington, Lakewood, NJ) for 3-4 hours at 37° C. Details of this method have been previously published (Gould and Clifton, 1985). The mammary epithelial cell-enriched suspensions (hereafter referred to as MEC preparations) were injected into the interscapular white fat pads of 30-35 day old nulliparious WF female recipients (N=16 animals per group) at 1 x 10^6 cells per animal in 100 µl of DMEM/F12 (Life Technologies,) media (1 x 10^7 cells/ml). Whole mammary gland transplants from WF donors were performed as positive controls.

Tumor development in the ectopic mammary glands was examined 15 weeks following DMBA administration to determine the mammary carcinoma susceptibility of the ectopic mammary gland. Transplants that did not develop tumors were examined by whole mounting the fat pad and staining with aluminum carmine to confirm that the transplanted gland developed properly in the fat pad, as previously described. Fat pads that failed to develop mammary glands greater than 3 x 3 mm^2 were excluded from further analysis. Previous studies utilizing this method have reported that a very small percentage of recipient rats develop multiple tumors in the transplant site. Therefore, tumor outcome was assessed as a binary response and analyzed by logistic regression.

Mammary Gland Cell Injections supplemented with Adherent Cell Fraction

Mammary gland cell preparations were made from WF donors using the mammary gland cell preparation protocol described above. Adherent cells left in the

flask were removed with Trypsin +EDTA (Life Technologies) for 5 minutes and placed into 15 ml centrifuge tubes. Approximately 250 µl FBS was gently added to the bottom of each tube to form a distinct layer for the cells to be centrifuged into. The cell solution was centrifuged briefly to collect cells and the supernatant discarded. The cell pellet was resuspended in media and counted and herein referred to as adherent cell fraction (ACF) preparations. ACF preparations were added to MEC preparations at 1:10 ratio resulting in a final MEC concentration of 1×10^7 cells/ml plus a final ACF concentration of 1×10^6 cells/ml. MEC/ACF preparations were injected into the interscapular fat pads of nulliparious WF female rats. MEC preps alone were injected into the fat pads of nulliparous WF female rats as a control. At 12 weeks of age all recipient rats were administered DMBA. Recipient rats were necropsied 15 weeks post-DMBA and examined for tumor development in the ectopic site. Fat pads were mounted and processed and stained with aluminum carmine. Stained fat pads were evaluated for proper mammary gland development and hyperplasia and DCIS development was assessed. Tumor, hyperplasia and DCIS data were analyzed using a Mann-Whitney rank test.

Results

Whole Mammary Gland Transplants Suggest Rat Mcs1b Is Mammary Cell Autonomous

To determine if the *Mcs1b* COP allele reduces mammary carcinoma susceptibility in a manner that is intrinsic to the mammary gland, rats were transplanted with whole mammary gland tissue into their interscapular fat pads and exposed to DMBA (Figure

10). All of the donor and recipient female rats used had a WF background and only differed at the *Mcs1b* locus having either a resistant COP or susceptible WF allele. We expected that there would be no rejection of the graft as both genotypes have the same major histocompatibility complex (MHC) type and, thus, have compatible immune systems. To verify this, we determined that recipients did not reject mammary tissue grafts from donors of different genotypes by microscopic assessment of whole mounted fat pads. There was no statistically significant association (P=0.1869) between the outcome of ectopic mammary gland development and donor or recipient genotype (Table 6).

Data from ectopic mammary gland-positive fat pad whole mounts were analyzed for associations of donor and recipient genotypes with ectopic carcinoma development. Ectopic mammary glands formed from mammary gland tissue from *Mcs1b* resistant donors, when grafted into either resistant or susceptible recipients, resulted in fewer ectopic tumors forming compared to ectopic mammary glands from susceptible donors when grafted into animals of either genotype (Figure 11). The genotype of the donor was significantly associated (P=0.0019) with ectopic mammary carcinoma outcome while there was no effect seen from the recipient genotype (Table 7). These data signify that the resistant phenotype conferred by the *Mcs1b* COP allele is acting in a mammary gland-autonomous manner. Based on this, future studies to functionally characterize the *Mcs1b* locus should be focused on the mammary gland.

Enzymatically-Isolated Mammary Gland Cells Injected into the Interscapular Fat Pad Reconstitute a Normal Mammary Gland in the Ectopic Site

Table 6:

Graft site MG development outcome (dependent) and *Mcs1b* donor and recipient genotypes (independent)

	Coefficient	P value	Odds Ratio (95% CI)
Donor Effect	0.99	0.1869	2.69 (0.62 – 11.68)
Recipient Effect	-1.31	0.116	0.27 (0.05 – 1.38)
Intercept	2.7	0.0004	

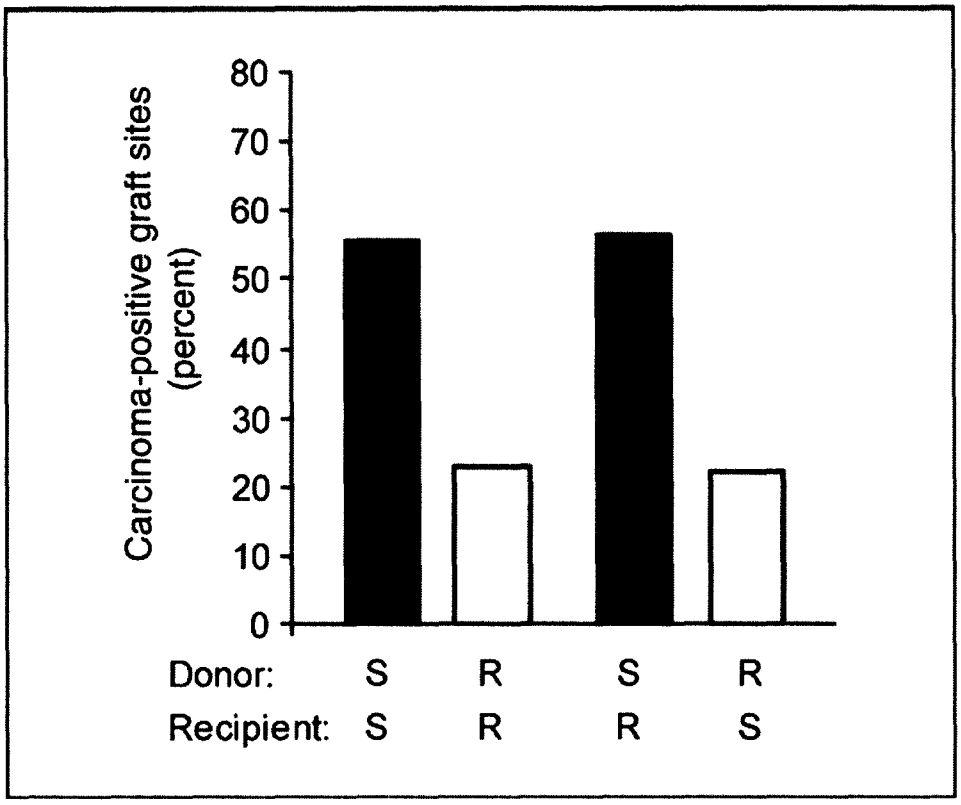


Figure 11: Rat *Mcs1b* is mammary gland autonomous. Percentage of mammary gland-graft–positive recipients that developed ectopic mammary gland carcinomas are shown for each susceptible (S) and *Mcs1b*-resistant (R) donor:recipient group. Groups with S donors are shown as filled bars, and groups with R donors are shown as unfilled bars. The total number of mammary gland-graft–positive recipients that were evaluated for tumor outcome in each group were, respectively, 27, 22, 23, and 18 for S:S, R:R, S:R, and R:S.

Table 7:

MG graft site tumor outcome (dependent) and *Mcs1b* donor and recipient genotypes (independent)

	Coefficient	P value	Odds Ratio (95% CI)
Donor Effect	1.48	0.0019*	4.4 (1.73 - 11.18)
Recipient Effect	-0.04	0.9381	0.96 (0.39 – 2.36)
Intercept	-1.22	0.0045	

The *Mcs1b* COP resistance effect is intrinsic to a mammary gland cell type. As mentioned, the majority of rat mammary tumors arise in the epithelial cells lining the duct. In addition to these ductal epithelial cells there are myoepithelial cells that surround the duct and TEB and multipotent cap cells that can differentiate into myoepithelial cells or move inward and differentiate into luminal cells (Williams and Daniel, 1983). Based on the fact that mammary tumors originate in the cells lining the mammary gland duct we hypothesized that one or more of these epithelial cell types are responsible for *Mcs1b* COP-conferred resistance. However, there is evidence that other cell types in the mammary mesenchyme affect tumor development (Pollard, 2004; Sica et al., 2006; Smits et al., 2011b). To test which cell type the *Mcs1b* COP allele is acting on to control mammary carcinoma development I used a MEC-enriched clonogen transplant approach that had been previously developed (Gould and Clifton, 1985). This method is used to enrich for MECs and non-adherent leukocytes and lymphocytes by removing mammary gland adipocytes, fibroblasts, and other strongly adherent cells by negative selection. These cells are injected into the interscapular fat pad of recipients to induce development of an ectopic mammary gland similar to the whole mammary gland transplantation experiments.

Enzymatic preparation of mammary glands to enrich for MECs is the standard protocol. However, this method is time-consuming and I wished to decrease the time required to dissociate mammary glands. To do this, I compared cells dissociated mechanically using a Medicon tissue dissociator to cells from enzymatic preparations to determine whether they recapitulated a morphologically normal mammary gland in the transplant site. Rats receiving mechanically-dissociated MECs did not develop

mammary glands in the ectopic site while 75% of the rats injected with enzymatically-isolated MECs developed ectopic mammary glands when a 1:1 donor:recipient ratio was used (Table 8). Mammary glands grown in the ectopic site looked morphologically normal (Figure 12). Based on these results, subsequent MEC preparations were performed using the enzymatic method.

To estimate the number of rats to use for these studies we used standard deviations (SD) data from previous mammary cell autonomy experiments resulting in SD of approximately 0.6 ectopic mammary tumors. For a 95% confidence interval and using a tolerable error of ± 0.15 of the mean ectopic mammary tumors with a SD of 0.6, approximately 60 animals must be tested. Given we achieved 75% transplant efficiency in the pilot study, to be successful 80 animals per group are required.

Mammary Gland Cell Preparations Are Not Sufficient to Allow for Efficient Mammary Carcinoma Development at Ectopic Graft Sites

Since injection of MECs into the interscapular fat pad resulted in mammary glands formed in the ectopic transplant site, I intended to use this method to determine if the *Mcs1b* locus was controlling mammary carcinoma susceptibility in MECs. MECs isolated from susceptible WF (N=16) or resistant WF.COP (N=14) rats and injected into the fat pads of age-matched susceptible WF female rats to determine the effect of the genotype on tumor outcome in the transplant site. Upon inspection of stained whole mounts, all recipients (100%) exhibited mammary gland development in the ectopic site. However, tumor development was not significantly different (P=0.9221), as each group developed only one tumor each (Figure 13A). Tumor outcome in MEC-derived

Table 8:

Mammary gland outcome in enzymatically-dissociated versus mechanically-dissociated mammary gland cell preparations.

MEC isolation method	# Rats Administered	# Ectopic Mammary Glands formed	% Positive
Medicon	3	0	0
Collagenase Digestion	4	3	75

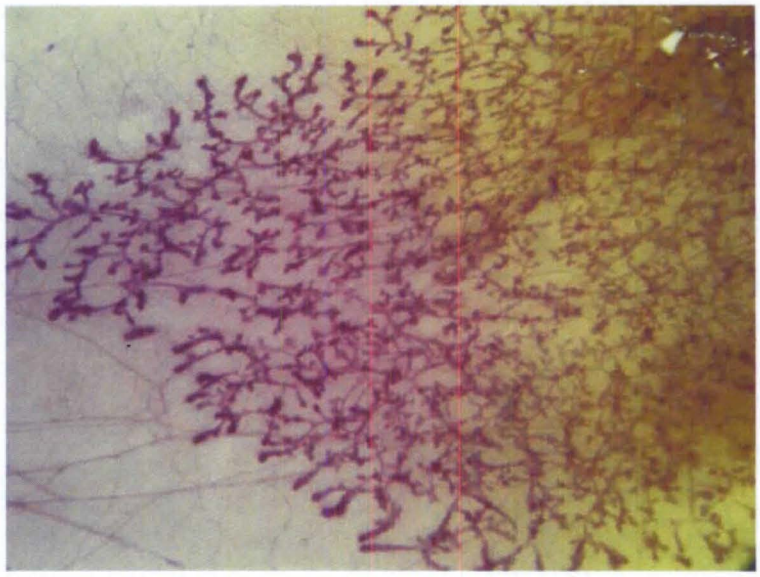


Figure 12: Representative whole mounted interscapular fat pad from MEC pilot study. 10X magnified image of aluminum carmine stained mammary gland formed in the fat pad following injection of MECs isolated via collagenase digestion.

mammary glands was determined to be significantly reduced when analyzed against whole mammary gland transplant controls (P=0.0007, Figure 13A).

Since tumor development was considerably reduced in animals receiving mammary gland cell preparations compared to whole mammary gland transplants, I examined the fat pads microscopically to determine if there were any gross morphological changes compared to ectopic mammary glands developed from whole mammary gland transplants. There did not appear to be gross morphological differences in ductal branching or TEB formation at the microscopic level; however, hyperplasia and DCIS formation was detected in the majority of whole mounts (Figure 12B). When quantified, hyperplasia were not significantly different (P=0.4884, Kruskal-Wallis test) between ectopic mammary glands from recipients of WF, WF.COP, or whole mammary gland transplants (Figure 12C). DCIS development was not significantly different (P=0.3545, Mann-Whitney test) between ectopic mammary glands formed from MECs from either genotype (Figure 13C). However, DCIS formation in whole mammary gland transplants was significantly greater (P=0.0003, Mann-Whitney test) than for ectopic mammary glands formed from MECs of both genotypes.

Injection of Mammary Gland Cell Preparations Along With Adherent Cell Fractions Does Not Allow for Mammary Carcinoma Development at Ectopic Graft Sites

Since there was no distinction in tumor outcome between ectopic mammary glands from susceptible and resistant MEC preparations and the ability to form frank carcinomas was nearly ablated, I sought to determine what was missing in cell preparations that was necessary for tumorigenesis. The MEC isolation method I

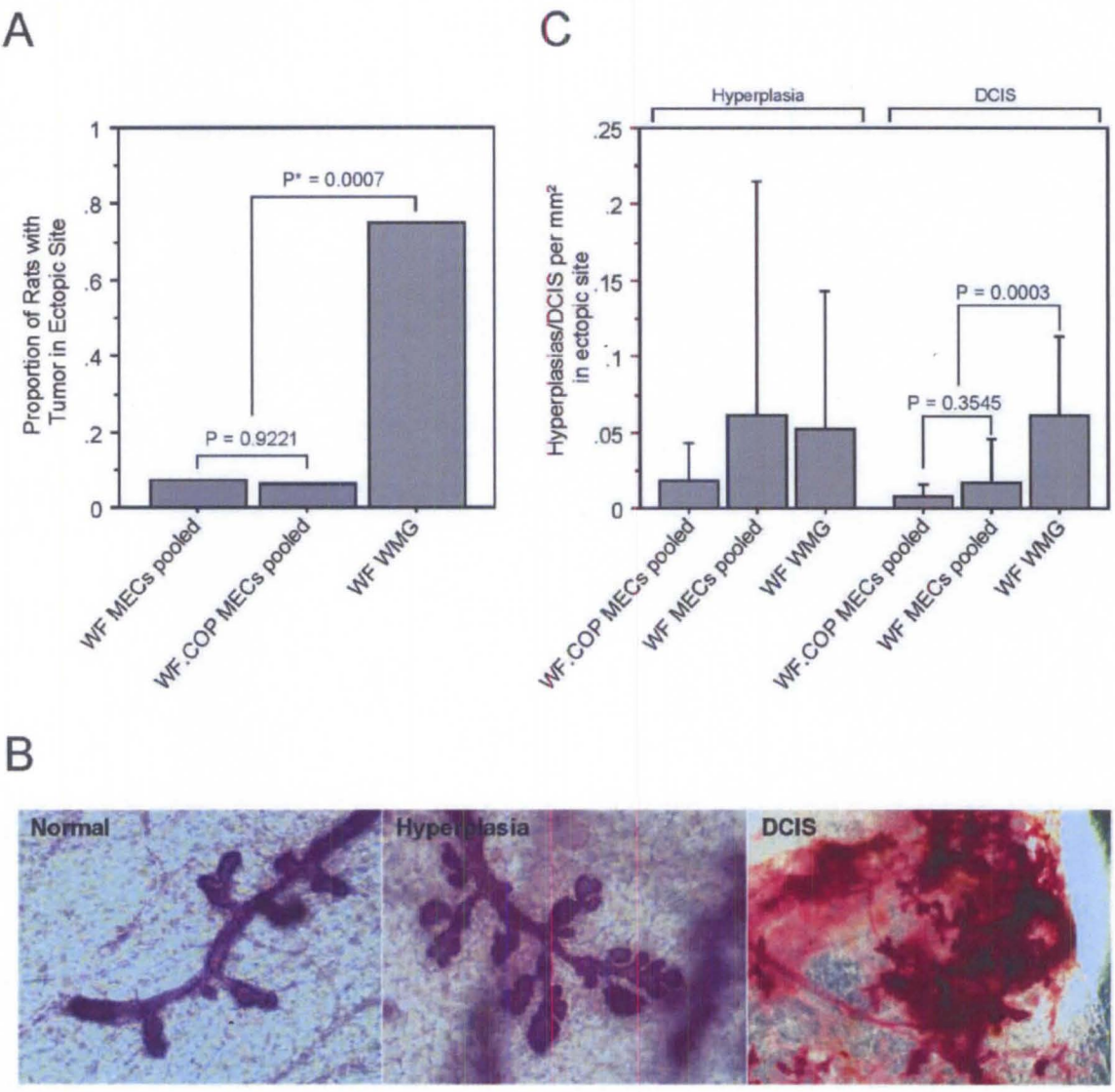


Figure 13: Tumor outcome in ectopic mammary glands derived from MECs. A; Proportion of interscapular fat pads with ectopic mammary glands exhibiting tumor development. P values based on logistic regression analysis. B; Microscopic images of aluminum carmine-stained normal TEB, hyperplastic TEB, and precancerous lesion stained from interscapular fat pad whole mounts. C; Quantification of hyperplasia and DCIS formed per mm^2 in ectopic mammary glands. P values based on nonparametric Mann-Whitney test.

originally employed enriches MECs by negative selection removing most adherent cells, i.e., fibroblasts, macrophages, mast cells, as well as, adipocytes, which remain on the top layer when centrifuged. These cell types play roles in mammary gland function; therefore, we hypothesized that the absence of one or more of these cell types resulted in retarding tumor development. To test this I adapted the original MEC isolation method to include the ACF that was discarded in the original protocol. These ACFs were retained and injected along with MECs. MEC-only injections and whole mammary gland transplants were performed as controls. I was attempting to restore the DMBA-induced mammary tumorigenesis susceptibility phenotype seen previously with WF whole mammary gland transplants; therefore, resistant cells were not considered and only WF donor cells were used.

I anticipated that co-injecting the adherent cells lost during dissociation along with MECs would restore mammary tumor development in the transplant site. Surprisingly, there was no difference (P=0.6664) in ectopic tumor outcome in WF rats receiving WF MECs (N=9) and WF MECs co-injected with the AHC (N=9) but, combined, these were significantly different from those receiving whole mammary gland transplants demonstrating that inclusion of the ACF was not sufficient to restore tumorigenesis (Figure 14A). Additionally, I quantified hyperplasia and DCIS formation microscopically by counting hyperplastic TEBs or DCIS foci. I noted no significant difference (P=0.8345, Kruskal-Wallis test) in hyperplasia formation between ectopic glands formed from MECs, MEC plus ACF cells or whole mammary gland transplants. DCIS development did not statistically differ between rats receiving WF MECs and those getting WF MECs plus the adherent cell fraction. However, DCIS formation was

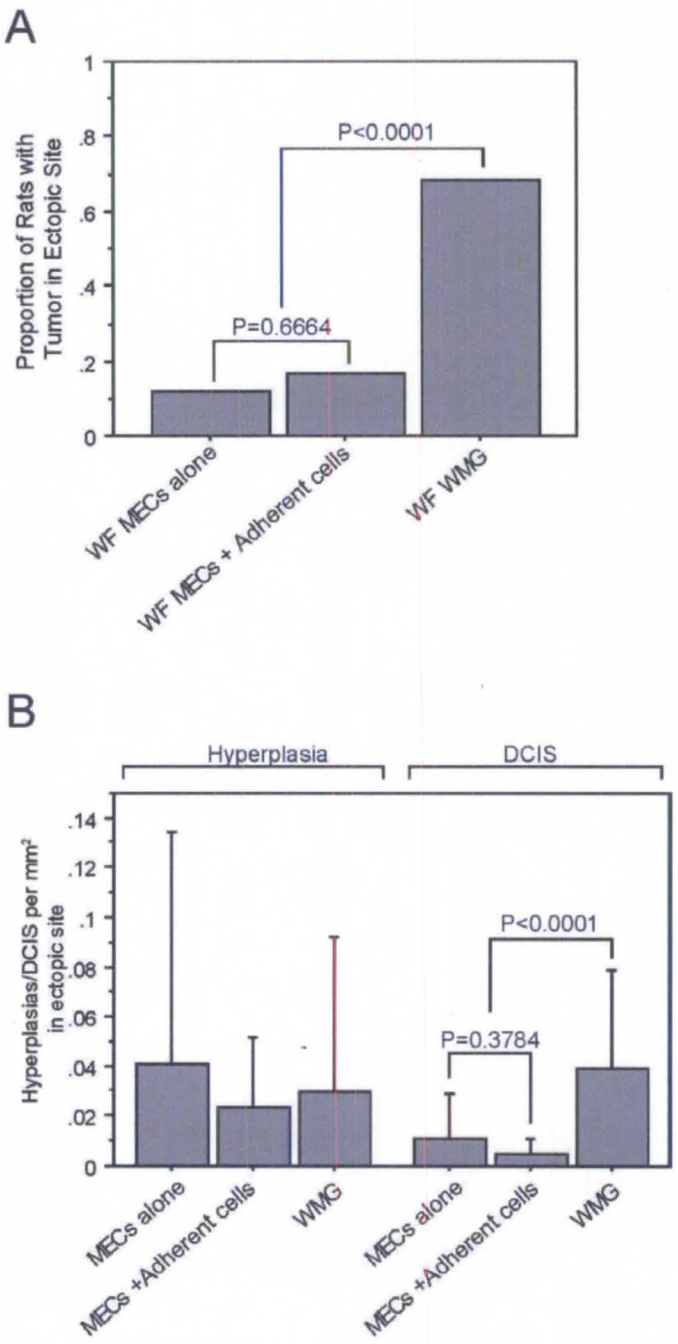


Figure 14: Tumor outcome in ectopic mammary glands derived from MECs or MECs supplemented with adherent cells. A; Proportion of interscapular fat pads with ectopic mammary glands exhibiting tumor development. P values based on logistic regression analysis. B; Quantification of hyperplasia and DCIS formed per mm^2 in ectopic mammary glands. P values based on nonparametric Mann-Whitney test.

significantly increased ($P<0.0001$) in whole mammary gland transplants compared to either cell injection (Figure 14B).

DMBA Susceptibility Is Delayed In Ectopic Mammary Glands Formed from Injection of Mammary Gland Cell Preparations

Injection of cellular components of mammary glands into the interscapular white fat pad failed to form tumors in the resulting ectopically-formed mammary gland. Co-injecting the AHC along with MECs again showed no difference in formation of frank carcinomas, DCIS or hyperplasia suggesting that there was a more complex process taking place. Within the process of dissociation of the mammary gland for MEC enrichment adipocytes are removed. Adipocytes are critical in mammary gland development and could be the missing link in efficient tumor formation that is lost in the mammary cell preparation transplants. However, there is an adequate concentration of host adipocytes within the interscapular fat pad cellular milieu making this less likely. Additionally, extracellular matrix (ECM) proteins are removed during MEC enrichment by collagenase digestion. The ECM provides an essential substrate required for normal mammary gland development; therefore, ECM may be required for tumor growth in these ectopic mammary glands. However, reintegrating MEC preparations with the adherent cells would restore fibroblasts which are responsible for ECM deposition, making this scenario less likely as well. DMBA-induced mammary carcinogenesis in the rat has been established to be dependent on a distinct mammary gland developmental window of 50-55 days of age; however, this window is based on fully developed mammary glands (Russo and Russo, 1994). We hypothesized that a different time span is required for

dissociated mammary gland cells to reconstitute a functional mammary gland than for whole gland transplants, i.e., the DMBA-susceptibility window is shifted.

To test this idea I used the original MEC isolation protocol, including susceptible and resistant MEC preparations only and administered DMBA at 40-45, 50-55, 60-65, and 70-75 days of age days. By including susceptible and resistant donor MECs, I would be able to identify whether the *Mcs1b* COP resistant phenotype is acting in the MEC population.

Interestingly, only 32% (8 of 25) of female rats receiving DMBA at the 40-45 day susceptibility window developed a mammary gland in the interscapular fat pad while those receiving DMBA at 50-55, 60-65, or 70-75 days of age developed 81%, 74%, and 68%, respectively (Figure 15A). This may suggest that administering DMBA too early after MEC injection may affect the efficiency of the graft to develop into a proper mammary gland.

Tumor outcome in the ectopic site was higher at the 40-45, 60-65, and 70-75 day susceptibility windows compared to the 50-55 day window (Figure 15B). These results are without regard to genotype of the MEC transplant as no genotype effect was detected (Figure 15B, bottom). Results for the 40-45 and 60-65 day windows were not statistically different from that of the original 50-55 day window (P=0.2352 and 0.0659, respectively). However, fat pad whole mounts from rats that received DMBA at 70-75 days exhibited a significantly higher tumor outcome when compared to the original 50-55 day DMBA window (P=0.0041). Furthermore, tumor outcome for the 70-75 day window was not significantly different than what was observed in whole mammary gland transplants (P=0.2025). These results suggest a shift in the susceptibility window of

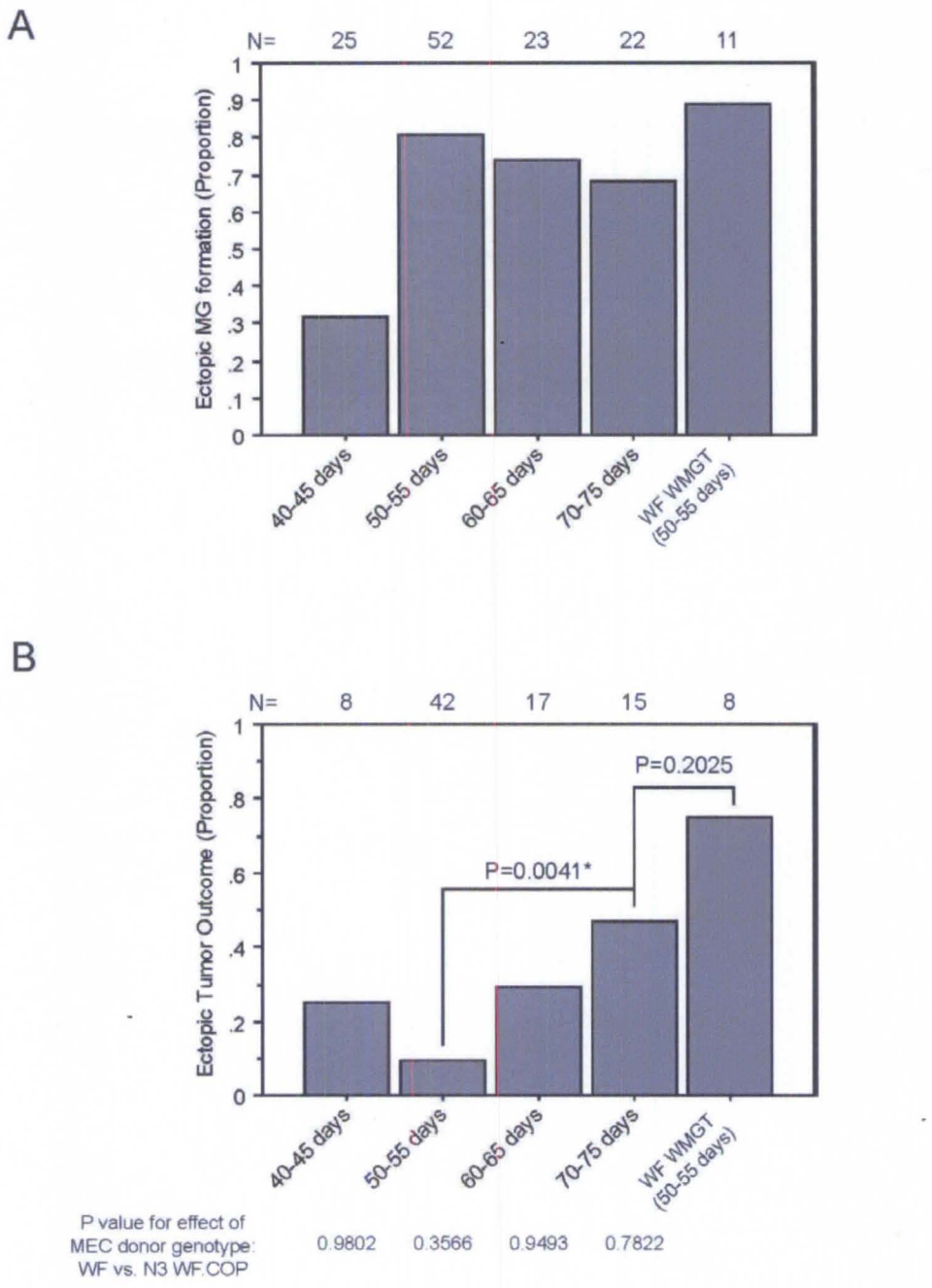


Figure 15: Results of DMBA susceptibility window studies with MEC injections. A; Mammary gland development outcome in the interscapular fat pad. Bars represent the proportion of rats injected that developed a mammary gland in the injection site. B; Tumor outcome in the ectopic mammary gland at varying windows of DMBA-administration. P values are based on logistic regression analysis.

ectopic mammary glands derived from MECs compared to those from whole mammary glands. Also, since no genotype effect was observed it is likely that the action of the *Mcs1b* locus to control mammary carcinogenesis is not intrinsic to a mammary epithelial cell type.

Discussion

Carcinogenesis in the mammary gland is complex involving multiple cell types and signaling mechanisms. On these lines, we set out to determine if the resistance to mammary carcinoma development conferred by the *Mcs1b* locus was acting in a mammary gland-intrinsic or extrinsic manner. The majority of mammary carcinomas arising in the rat originate in the ductal epithelium; thus, we hypothesized that the *Mcs1b* locus was acting within a mammary gland cell type to confer resistance.

Transplanting immature mammary glands into the interscapular fat pad results in the development of a properly branched and formed mammary gland (Gould and Clifton, 1985). We used this approach and induced mammary carcinogenesis with a single dose of DMBA. The results of these experiments supported our hypothesis that the *Mcs1b*-conferred phenotype is mammary gland intrinsic.

Transcripts within the *Mcs1b* locus remain to be functionally tested and these results indicate that the mammary gland should be the focus of these assays, e.g., gene expression profiles and splice variant analysis. Moreover, these data focus on a cell type of interest. Cells from the immune system can have an effect on mammary carcinogenesis in the tumor microenvironment (Pollard, 2004; Sica et al., 2006; Smits et al., 2011b; Trimboli et al., 2009). Herein, we have shown that this is not likely the case

and focused on the mammary gland cellular milieu. The mammary gland is composed of a variety of cell types all of which have roles in normal gland development as well as progression to a neoplastic and malignant state (Lanigan et al., 2007). Therefore, identifying a candidate cell type will afford a greater understanding of the etiology of mammary cancer.

Given that rat mammary tumors arise predominantly in the epithelial cells lining the ducts, I hypothesized that MECs were responsible for the *Mcs1b*-conferred phenotype. To test this I modified the original transplant protocol by injecting MEC-enriched cell preparations and tracking the phenotype. I anticipated that approximately twice as many WF MEC-derived ectopic mammary glands would exhibit tumor development than glands developed from WF.COP MEC donors, similar to what was seen in whole mammary gland transplant assays. However, the ability to form tumors in the transplant site was lost, suggesting that there is a component missing necessary for tumor development.

I estimated that 80 animals of each genotype would have to be tested to estimate a mean tumor outcome to within ± 0.15; however, I only tested 16 homozygous WF and 14 resistant congenic WF.COP rats in these studies. Although this is below the target number, the ability to form tumors in the fat pad was nearly completely ablated in all animals. Moreover, the lack of tumor development was not an artifact of poor transplant efficiency since the majority of recipients (97%) developed mammary glands in the interscapular fat pad. These ectopic glands appeared to have developed normally and, upon microscopic inspection, exhibited hyperplasia and DCIS development suggesting that DMBA was sufficiently initiating carcinogenesis within the ectopic site. Hyperplasia

and DCIS were quantified to determine if there was a difference between the MEC and the whole mammary gland transplants. No significant difference in hyperplasia was identified between ectopic mammary glands formed by MECs from the two genotypes, nor were these different than hyperplasia in whole mammary gland transplants. Similarly, DCIS formation was not statistically different in ectopic glands formed from congenic WF.COP MECs compared to ectopic glands from homozygous WF MECs. However, DCIS formation in ectopic mammary glands from either genotype was much lower than what was observed in whole mammary gland transplants suggesting carcinogenesis was being initiated in the MEC-derived glands but something involved in the MEC-isolation process removed a component required for progression to a frank carcinoma.

My hypothesis was that a cellular component was removed during MEC-isolation that was required for complete tumor-induction by DMBA. To address this hypothesis, I first attempted to restore the ability to form tumors by reclaiming the cells removed during MEC enrichment and re-introducing them along with MECs into the fat pad and following the phenotype. However, including these adherent cells with MECs had no effect on tumor outcome compared to MECs alone, suggesting that the mechanism is more complex. The loss of tumorigenisis could be explained by the absence of adipocytes since these were discarded during MEC-enrichment. However, this seemed unlikely since the recipient fat pad contains adequate adipocyte content and allowed for mammary gland development in the majority of recipients. Therefore, it appeared that this effect could not be explained simply by a single missing cellular component.

Whole WF mammary gland transplants were performed as controls in these

experiments and these exhibited a normal ectopic tumor outcome. Also, hyperplasia and DCIS development was quantified in ectopic glands from MEC-, MEC plus adherent cell-, and whole mammary gland-induced grafts. All grafts exhibited hyperplasia development, but there was no statistical effect associated with MECs, MECs plus adherent cells, or whole mammary gland transplants. Moreover, DCIS formation was not statistically different between ectopic mammary glands formed from MEC or MEC plus adherent cell grafts; however, again the level of DCIS formation in either cell-derived ectopic gland was significantly reduced compared to whole mammary gland grafts.

Taken together, these data suggest that DMBA is driving neoplastic events to form hyperplasia and precancerous DCIS lesions but is insufficient for progression to development of frank carcinomas. We hypothesized that there was a developmental gap in these ectopic mammary glands affecting their susceptibility to DMBA-induced carcinogenesis. I sought to determine if there was a difference in susceptibility to DMBA in WF MEC-grafted ectopic mammary glands at different developmental windows. DMBA was administered in previous experiments at 50-55 days of age of the recipient rat; approximately 20 days after cell injections were performed. Rather than adjust the age at which cell injections were performed, we chose to keep this constant at 30-35 days and, instead, administer DMBA at 4 different time points: 40-45 days, 50-55 days, 60-65 days, and 70-75 days. If there was an effect from the developmental stage of the ectopic mammary gland, it was anticipated that one of these windows would return the mammary carcinogenesis phenotype towards a level similar to that seen in the whole mammary gland transplant studies. One may predict that an earlier DMBA window would likely result in an increase in tumor formation as these cells would be expected to be more

mitotic and, therefore, more vulnerable to DNA damaging agents. However, outcomes of the susceptibility window tests showed that female rats receiving DMBA at the 70-75 day window developed significantly more tumors than rats receiving DMBA at the 40-45, 50-55, and 60-65 day windows. Tumor outcome in the 70-75 day group was not statistically different than that observed in whole mammary gland transplants, suggesting a developmental dependency for mammary carcinogenesis induction by DMBA. This may hold true for other PHCs as well.

In addition, efficiency of proper mammary gland development in the interscapular fat pad was retarded at the 40-45 day DMBA window. This implies that DMBA may have an effect on mammary gland development. At this time point, DMBA was administered only ten days following cell injections. DMBA causes acute immunotoxicity immediately after it is given (Gao et al., 2005). It may be that immune cells necessary for appropriate ductal branching and elongation are not present, resulting in atrophy of the mammary gland in the ectopic site. However, the high dose of DMBA administered to rats is not physiologically relevant; therefore, this effect is not applicable to human health but has implications on future rat studies. More studies will need to assess the mechanism by which DMBA affects mammary gland growth in these studies, e.g. measure immunological cells in the plasma and in the ectopic mammary gland.

The shift in the DMBA-susceptibility window suggests that at earlier stages of mammary gland development, DNA-damaging agents, such as PHCs, have less effect on mammary carcinogenesis. It is not clear what the implications of this are. These data may suggest a reduced role for mammary "stem" cells in progression to full carcinoma development. Cancer stem cells have been proposed to be principal mediators in primary

tumor development and relapse following therapy. My data suggest, however, at earlier stages the stem cell niche may not be capable of driving full carcinogenesis and, rather, a more differentiated cell type is requisite for tumor development. Still, it could be argued that these undifferentiated cells may require more than a single dose of DMBA to develop frank carcinomas. Indeed, DMBA-induced carcinogenesis relies on steroidal hormone activation, such as estrogens, as a "second hit" to drive carcinogenesis and potentially these stem niche cells lack ample ER expression. We currently have no information regarding the cell populations present or how they act within the 60-65 or 70-75 day window. Therefore, there is no evidence for or against a role for a particular mammary gland cell type in DMBA-induced mammary carcinogenesis.

Previous studies have demonstrated that the clonogen transplantation method is sufficient to induce mammary gland formation in an ectopic site; yet, to date, no one has shown a capacity to stimulate mammary cancer development in these ectopic glands. These experiments are important because they provide insight into mammary carcinogenesis in the context of environmental exposures. Herein, I show that the mammary carcinogenic effects of DMBA are dependent on a specific stage of mammary gland development. Although the data imply that the *Mcs1b*-conferred phenotype is not intrinsic to a mammary epithelial cell population, work to identify a specific cell type responsible for *Mcs1b*-induced resistance, regrettably, remains inconclusive. Regardless, the initial goal of this aim was to identify a tissue type involved in *Mcs1b*-conferred resistance to mammary carcinoma development. My data show that the action of the *Mcs1b* COP allele is specific to the mammary gland. Future efforts to functionally test gene candidates within the *Mcs1b* locus will, consequently, be focused on the mammary

gland as a target tissue. Although these studies do not provide definitive answers as to what the specific factors are affecting tumor development, they afford a means by which to study these effects further. On the whole, my data identify that the *Mcs1b* COP allele is acting in mammary gland-autonomous manner and indicate that the effectiveness of DMBA to induce mammary carcinogenesis is dependent on the developmental stage of the mammary gland.

CHAPTER IV

MIER3 IS A CANDIDATE BREAST CANCER-ASSOCIATED GENE

Introduction

There are thirteen transcripts annotated within the 1.8 Mb Mcs1b locus delineated by lines N3, W2, U2, and I4 and six of these transcripts reside in the 1 Mb region delineated by addition of line 1b-11 (Figure 5). All of these genes are also annotated to the human breast cancer-associated locus marked by SNP *rs889312*, and any combination of these transcripts could be the responsible gene(s) involved in *Mcs1b* COP-conferred mammary carcinoma resistance (Figure 8). The most plausible breast cancer candidate gene within this region is *Map3k1*. However, no direct evidence is available to support *Map3k1* as the causal gene. It is relevant, then, to test the transcripts annotated to this region. We hypothesize that a genotypic variant between WF and COP in the *Mcs1b* locus lies within the coding region of one or more of these transcripts and results in an amino acid change that gives rise to a differential mammary carcinoma susceptibility phenotype. This approach is important in that SNP haplotype maps in the rat are incomplete; thus, identifying variants will increase current knowledge of rat genetics and potentially identify causative SNPs. Many cancers are caused by detrimental mutations or variants that alter protein function and this may be the basis the *Mcs1b* phenotype.

All of the human SNPs tagged along with *rs889312* lie outside of annotated transcribed genes (Figure 3). Consequently, it is possible that the inherited element(s)

controlling the risk phenotype are non-coding regulatory elements. Indeed, many low penetrance alleles associated with disease phenotypes are now being found to involve non-coding variations. There may be additional SNPs not yet identified that lie in transcribed regions that may be causative; thus, it is important to first test for changes in the coding region.

To not bias the approach towards any specific transcript(s), I first set out to resequence the open reading frames (ORFs) of each transcript lying in the 1.8 Mb *Mcs1b* locus as it was delineated from marker *A12v* (ch*2*:42364155) to marker *a12oo* (ch*2*:44210175). The mechanism by which the *Mcs1b* locus is controlling mammary carcinogenesis susceptibility is not clear; therefore, including all of the transcripts in this interval reduces the probability of missing the causative factor.

In the past decade there has been increasing interest in regulatory microRNAs (miRNAs) in cancer research. These miRNAs are short pieces of untranslated RNA transcribed by the cell that can recognize and bind specific target sequence sites on translated mRNAs and mark them for degradation. They regulate normal processes within the cell and have been shown to be dysregulated in some cancers. Differential effects can result either through sequence variation in the miRNA target sequence or through alterations in the sequence or the expression of the miRNA itself. These miRNA target sequences are principally located in the 3' untranslated regions (UTRs) of the mRNA to be regulated. Therefore, in addition to resequencing the ORFs, I also resequenced the 3' UTRs of these transcripts.

Sequence differences between WF and COP sequence in any of the transcripts will be analyzed *in silico* to determine if the variation(s) disrupts the normal structure of

the protein, e.g., results in an amino acid change or introduces a premature stop codon. It may be that variation between the two rat lines results in differences in stability of a transcript. Efforts will be focused on determining transcript expression levels between susceptible WF and resistant WF.COP line T and N3 rats. Conversely, if no variation is identified in the coding region of any of these transcripts, it will be tacit that the mechanism of the *Mcs1b*-conferred phenotype is regulatory in nature. In this case, gene expression of all the transcripts will be tested. In each case, gene expression levels will be first assessed in mammary gland tissue as we found *Mcs1b* COP-conferred mammary carcinoma resistance to be autonomous to the mammary gland. Other tissues will be tested to establish whether the effect is mammary gland-specific.

Design and Methods

Resequencing Mcs1b Transcripts

Spleen or thymus tissue from WF/Hsd and WF.COP lines N3 and T rats was excised and total RNA was extracted using TriReagent (Molecular Research Center) and standard chloroform/isopropanol precipitation. RNA samples were treated with TURBO DNase (Life Technologies) to reduce DNA contamination and cDNA was made using Superscript III reverse transcriptase (Life Technologies). Sequences were not attainable from cDNA in some instances; therefore, genomic DNA, extracted from frozen spleen or liver tissues using standard phenol-chloroform/isopropanol precipitation, was used. Samples were PCR amplified using Accuprime HiFi Taq polymerase (Life Technologies) and, subsequently, cleaned with the QIAquick PCR Purification Kit (Qiagen) to remove unincorporated nucleotides. Amplified samples were sequenced using the BigDye

Terminator v.3.1 Cycle Sequencing Kit (Life Technologies and purified with Agencourt CleanSeq magnetic beads (Beckman Coulter). These were submitted for analysis to the University of Louisville Center for Genetics & Molecular Medicine DNA Core using an ABI PRISM 3130XL Sequence Detection System (Life Technologies). Primer sequences for amplifying and sequencing *Mcs1b* ORFs and 3'-UTRs are in Table 9. Nucleic acid sequences were submitted to NCBI/GenBank and assigned accession numbers JQ013728 through JQ013739.

Gene Expression Assays

Tissues were excised from WF/Hsd and WF.COP lines N3 and T rats that had been treated with DMBA or not and flash-frozen in liquid nitrogen. Total RNA was extracted with TRI-Reagent (Molecular Research Center) followed by standard chloroform/ethanol precipitation. To reduce possible solvent and DNA contamination RNA samples were further processed by a 1/10 v/v 3M sodium acetate and 2.5x v/v 100% ethanol wash on ice for 10 minutes followed by 80% ethanol wash followed by Turbo DNase (Life Technologies) treatment. Total RNA quantity and quality were measured with a Nanodrop 1000 (Fisher Scientific) and a Bioanalyzer with RNA 6000 NanoChips (Agilent). cDNA was made by reverse transcription reactions using (20 µl f.v.) 1µg total RNA, 0.5x RNAsecure, 5µM random hexamers, 25ng/µL oligo(dT_{18}), and 0.5 mM dNTPs were incubated 5 minutes at 65° C prior to adding 1× first strand buffer, 100mM DTT, and 1µL Superscript III (Life Technologies). Reactions were incubated 5 m at 25° C, 1 h at 50° C, and 15 m at 70° C. *Taq*Man QPCR primers and MGB probes (Applied Biosystems) were designed by using Primer Express v 2.0 (Applied

Table 9:

Primers for resequencing ORFs and 3' untranslated regions of transcripts in 1.8 Mb Mes1b region

Gene (GenBank Accession #)	GenBank accession WF and COP ref sequence	Amplification Primer Lab ID	Amplification Primer Sequence	Nested Sequencing Primer Lab ID	Nested Primer Sequence
Map3k1	JQ013733	Map3k1-F14	CGGTCGAGAGATGGAGAATAA	Map3k1-Sq1	AGCTCAAGGCGACCTGTATG
(NM_053842)		Map3k1-R1	AGGCCAGCAGACTATCAACG	Map3k1-Sq2	CTGACTCCTTGAAAGCTCCA
				Map3k1-Sq3	CCCCATGCTATGGAGAAAAA
				Map3k1-Sq4	CGACTTCGGACTTGAAAGTGA
				Map3k1-Sq5	GCTCGTTGGCTGCTTATTCT
				Map3k1-Sq6	GGGTCTGCCTTTTGTTTGAA
Mier3	JQ013732	Mier3-F1	ATGGCCACACCCCAGGTCGCAGTTG	Mier3-Sq1	CTGATCAGGAATGGGCTCAG
(NM_001168000)		Mier3-R1	TGAGAACGGTTCAAACTTCC	Mier3-Sq2	TCACCACTGTACTCGCCAAG
				Mier3-Sq3	ACCCAGCAAAAGAACAAAGG
				Mier3-Sq4	GAACATGCATGGATTTGTGC
Gpbp1	JQ013734	Gpbp1-F1	AACACCTTCCTGTGCAGTCC	Gpbp1-Sq1	CAGCATGGCTTTCATCTTCA
(NM_001106410)		Gpbp1-R1	CATCTTCTAATGGGGTAGGTTCA	Gpbp1-Sq2	GTGTGTGTCTGGGCGTGTAG
Ankrd55	JQ013731	ANKRD55-F1-1	GAGGCACAAGATGCAAATAGCAG	ANKRD55-F2-1	GTTTCGGGACGTCAAGTGGA
(XM_342195)		ANKRD55-R7-1	AGCAGCAAACATGGCACAAA	ANKRD55-F3-1	GAGGCTTCTCATCGTCCTGCT
				ANKRD55-F4-1	AAGGTCCCCGAGTGTAACCTG
				ANKRD55-F5-1	TGACATCATCGCCACCTTTG
				ANKRD55-F6-1	TCAGAAAAGCCGAAGCGAAC
				ANKRD55-F7-1	TGGAGAACCCCTGAGAACGA
				ANKRD55-F8-1	TGTGCCATGTTTGCTGCTTT
				ANKRD55-R1-1	AAACTGGTGCGACCGTATGC
				ANKRD55-R2-1	TGTGAAAAGCTGCCCAATGA
				ANKRD55-R3-1	CTGGACGCATTCTGCCTTTC
				ANKRD55-R4-1	CCAGGCTGATCTTGGCAGTT
				ANKRD55-R5-1	CCAAGCCTTGCAGGAGAATG
				ANKRD55-R6-1	CTTGGAATCGCTGGAAGCAC
Il6st	JQ013730	Il6st_mRNA_#1_amplify_start_F	AGGAAGGGCGGACAACTGTTA	Il6st-F2-1	ACGTCCAGCTCACCTGCAAT
(NM_001008725)		Il6st_mRNA_#1_amplify_end_R	GGTGGAAGGGATTCCTTTGG	Il6st-F3-1	AACATCGAGGTCTGGGTGGA
				Il6st-F4-1	ATGGGAAGGGCTACTGGAGTG
				Il6st-F5-1	TGGACACCGCCATCTAAACC
				Il6st-F6-1	GGTCCACATGGCAGCATACA
				Il6st-F7-1	CCCCCAAGGCACAATTTTAAC
				Il.6st-R2-1	GGTGGGCTGGGTTTCACTTT
				Il.6st-R3-1	GAAACTTGGCGCTTTGGATG
				Il.6st-R4-1	TCTGGGATGCAGGGTGAGTT
				Il.6st-R5-1	AAAGTGAATTCCGGCCCATC
				Il.6st-R6-1	CCAAGGATTTCAGGTCATCTGG
				Il.6st-R7-1	CCTGGCTTGACCTTTCGAAGT
				Il6st mRNA #1 amplify end F	TTCTAGGCCCAGCATCTCCA
				Il6st mRNA #1 amplify start R	CAAACCTGAATTGACCCATGC
Il31ra	JQ013736	Il31ra-mouseortho2-F1	TTCTCTAGCCTCCCTTCCAG	IL31RA-F2-1	CAGCTACACCGTGAGGATGACTT
(XM_226759)		IL31RA-R7-1	TCATTTCAAGGGGCGTCTGT	IL31RA-F3-1	CGAACGGGTCTTCAACCTCA
				IL31RA-F4-1	ACTGCTTCTGTTGGGCCAGA
				IL31RA-F5-1	AGCCATATTCCATCCAAGCTTATG
				IL31RA-F6-1	CAGGAAGCCAAACCGATTGA
				IL31RA-F7-1	GGGAATGGCGGAAGAGACATA
				IL31RA-R2-1	TCTTTGCTCCACTTGCTCCAG
				IL31RA-R3-1	AACTCTGCCAGTTCACCACCA

Table 9 continued

Gene (GenBank Accession #)	GenBank accession WF and COP ref sequence	Amplification Primer Lab ID	Amplification Primer Sequence	Nested Sequencing Primer Lab ID	Nested Primer Sequence
				IL31RA-R4-1	TGACCACGGCTGTCTTCAAA
				IL31RA-R5-1	TTGAAATCATCTCCAAACCAGGTAG
				IL31RA-R6-1	TCTGAGCTCGCTCTTCACAGG
Ddx4	JQ013737	A49-Ddx4-New-F1	GAGAAGTGGATTTTCCTCTGGA	A49-Ddx4-New-F4	TTGTGGAACCTGAAGCTATC
(NM_001077647)		Ddx4-R8-1	AAGGGTTGGGAGTAAGAACAGAAGA	A49-Ddx4-New-R1	CCCACTTCTGCTCTGGAAAG
				A49-Ddx4-New-R2	TCCAGAGGAAAATCCACTTCTC
				A49-Ddx4-New-R3	CATTCCCATCTGGATAGCCG
				Ddx4-F3-1	AGGTGACAGCTCTGGTTTCTGG
				Ddx4-F4-1	TATACCCCCTCCTCCACCAGA
				Ddx4-F6-1	AGGTTGGCTGGCGAGTTTTT
				Ddx4-F7-1	TGGCTGCCAGAGGACTTGAT
				Ddx4-F8-1	GCACAAGAGGGGCTGTGTTT
				Ddx4-R3-1	TGGTGCATCATGTCCAGATACTTC
				Ddx4-R4-1	AGTGCCCAAACTGGGTTCCT
				Ddx4-R5-1	GGCCAACTTGAAGAATGGACTG
				Ddx4-R6-1	GCCCTGCCAGTATTTCCACA
				Ddx4-R7-1	ACTGGTTGGCAGCTTGTGC
Ppap2a	JQ013728	Ppap2a-F1-1	TCCTGGGAATTGGGTATCTGG	Ppap2a-R1-1	TGTAGCCATCGCTGCAGTTG
(NM_022538)		Ppap2a-R4-1	TGTTCGCAAGCATTTTTATTATA	Ppap2a-F2-1	AGCCGTTGGAGCCTTTTTGT
				Ppap2a-R2-1	CAGGGCTCATGATTCCTTGC
				Ppap2a-F3-1	TCAAGGAGCTGTTGTGGC-AAT
				Ppap2a-R3-1	TGTACAGGTGGGGCAGTGTTT
Slc38a9	JQ013729	Slc38a9-F1-1	CCAACAGCAAATCCCCATTC	Slc38a9-F2-1	AGCGACCATGCTGCTGCTA
(NM_001035251)		Slc38a9-R6-1	CAACATGCCTTGACCCAACA	Slc38a9-F3-1	TGGTGAAGTCCCGGAGTATGA
				Slc38a9-F4-1	GAGTGGCCATCCTGACAACAG
				Slc38a9-F5-1	TCCTGGTTTTTGCCTCATTCC
				Slc38a9-F6-1	CAAGAGGAGCCGTCTGACATGG
				Slc38a9-R1-1	AGTGCCTTGTCTGCAGGAGTG
				Slc38a9-R2-1	TTGCTCCAATGAGAGACACCAA
				Slc38a9-R3-1	AAGTTGAGCAGCGGCAAGAG
				Slc38a9-R4-1	TGCTGGGGAAGTTGTCCAAG
				Slc38a9-R5-1	TGCCAAGGTTCAGAACACCA
Map3k1-3'UTR		Map3k1-3'UTR-F1	ATGTGGGTGCCATGCAAAG		
		Map3k1-3'UTR-R1	ACCGAGACGGAACTGCAGAA		
		Map3k1-3'UTR-R2	GGGCCAACAGTAAAATGCACA		
Mier3-3'UTR		Mier3-3'UTR-F1	ACGCAGCCCTTCATTCAGAG		
		Mier3-3'UTR-R1	CGTCTCGTCATTTGCACACAG		
		Mier3-3'UTR-F2-2	AGGTACCAACCCCGATCC		
		Mier3-3'UTR-R2-2	AAATGTAATTAAGACGGTGCAAA		
		Mier3-3'UTR-F2-3	AGTGGAATCACAATGTTTACAGG		
		Mier3-3'UTR-R2-3	AAGACGGTGCAAAAAGTAATTG		
		Mier3-3'UTR-F3	GCACAGTGCAGAGCCTTGAA		
		Mier3-3'UTR-R3	GAGCCCAGGCTACCCATTATTT		
Gpbp1-3'UTR		Gpbp1-3'UTR-F1	CCTCTGCCTCCTTCAGCAAA		
		Gpbp1-3'UTR-F2	TCTCATGCAGTTTGGGGTGA		
		Gpbp1-3'UTR-F3	TTGGCTAAAACTGGAGCAAAGTG		
		Gpbp1-3'UTR-F4	TGGCCTGGAATTCTATGTAGC		
		Gpbp1-3'UTR-F5	AGGTACAAGGCAGCTTCCAG		
		Gpbp1-3'UTR-F6	GAGGTGCTTGTCCAGAGCTT		
		Gpbp1-3'UTR-R1	CAAGAGAAATGCTGCAGTGATGA		
		Gpbp1-3'UTR-R2	AGCTCTGGACAAGCACCTCCT		
		Gpbp1-3'UTR-R3	TACCCTAGCGTTGGCAGGAA		
		Gpbp1-3'UTR-R4-1	TCCTTTTGCTGTTGCTTTG		
		Gpbp1-3'UTR-R4-2	TGCTCCAGTTTTAGCCAACAT		
		Gpbp1-3'UTR-R5	TCCAATCACCTCTGAAAATCC		

Biosystems). Primer and probe sequences are provided in Table 10. One µl of the dilution (≈12.5 ng of RNA-equivalent cDNA) was used in a 16-µl *Taq*Man QPCR. The reaction components were 1× *Taq*Man Buffer A (Applied Biosystems); 5.5 mM $MgCl_2$; dATP, dCTP, dGTP, and dTTP at 400 µM each; experimental primers at 300 nM each; 200 nM *Taq*Man experimental probe (Applied Biosystems); *Rplp2* primers at 100 nM each, 200 nM rodent *Rplp2* probe; and 0.4 units of *Taq* Gold DNA Pol (Applied Biosystems). Real-time QPCR was run on an ABI PRISM 7900HT real-time PCR machine. Real-time QPCR cycling conditions were 50°C for 2 min and 95°C for 10 min, followed by 40 cycles of 95°C for 15 sec and 60°C for 1 min. FAM (*Mcs1b* region target gene probe) and VIC (Rodent *Rplp2* probe; Applied Biosystems) fluorescence values were measured by using Applied Biosystems SDS v 2.3 software; quantities of transcripts were measured by comparison of cycle threshold values with a standard curve calculated from serial dilutions. Sample measurements are an average of four replicates per sample and were standardized by dividing the quantity of rodent *Rplp2*. Data were analyzed by Mann–Whitney tests.

Comparative Genomics

Human and mouse sequences for genes annotated to the *Mcs1b* locus were obtained using the UCSC Genome Browser using the *Homo sapiens* version GRCh37/hg19, *Rattus norvegicus* version 3.4/rn4, and *Mus musculus* version NCBI37/mm9 genome assemblies. Sequences were aligned using the DNAStar SeqMan sequence analysis program (DNAStar) to identify orthologous regions. Primers were

Table 10 :

Rat Mcs1b Target and Rplp2 Primers and Probe Sequences Used for QPCR

Assay Name	Forward Primer Sequence (5' to 3')	Reverse Primer Sequence (5' to 3')	Probe Sequence (5' to 3')	NCBI Gene Reference ID
Rplp2	TGAACGACTCAACAAGGTCATCA	CAACACCCTGAGCGATGACA	VIC-CTGAATGGAAAGAATATTGAG	NM_001030021.1
Map3k1	TCCTCATATTGTCAGTACCGATGTC	TTGCAAGGCAAAGGCTAAGAG	6FAM-CCTGTTGAAATCAGGTATAA	NM_053887
Mier3	CGAAAGGTACTGCTGTAATGGAAA	GCACTCCTCTTCAGTCCAAGCT	6FAM-CGTCTCAAGAAGGAATG	NM_001168000.1
Gpbp1	GAGTAGAAGAGGAGCATGAAGATGAA	TGGTGAGTACTATTGCTGTTATGCAA	6FAM-CTCAGAGAAGGATGACGAC	NM_001106410.1
Il6st	GACAACGCTGCTGGGAGTCT	GGAGTTAAAATTGTGCCTTGGG	6FAM-AAACACATCTGGCCGAAT	NM_001008725.3
Ankrd55	ATGGTGCCAAGCACAACATC	GATGAGAAGCCTCACATCAGGTT	6FAM-CAGATAAAAATGGCCGCCTG	XM_342195.4
Il31ra	GCCAAAGAAAAAGCTCCAACAT	ACGGCTGTCTTCAAACCAATG	6FAM-TCCCAATACCACGGTGAA	XM_226759.4
Ddx4	CTGGGAAGACGGCAGCTTT	GCAGTTATTCCATCCCTCATCATA	6FAM-CTCTTGCCTATTTTGGCTC	NM_001077647.1
Slc38g	GGTTTGCGCCAACAGAATTT	CCCTGTCAGCTGTGGAAACTG	6FAM-TTGTACCAGAGATAAGAGC	NM_001035251.1
Ppap2a	TGTCAAGGGAATGAACAGAAGGT	AACAGCATGCAGTACATAGAGAATGA	6FAM-AAGGCAGGTTGTCCTT	NM_022538.2

designed against the rat orthologous sequence using Primer3 (Rozen and Skaletsky, 2000) and used to test for alternate start sites for *Mcs1b* transcripts in rat cDNA samples.

Genomics and Statistical Analysis

Mammary carcinoma multiplicity phenotypes were compared by nonparametric Mann-Whitney tests. Results from mammary gland grafting experiments were analyzed using logistic regression. Donor and recipient genotypes were incorporated as dependent variables. In independent models, graft site tumor outcome and grafting ability were used as independent variables. Quantitative PCR (QPCR) data were analyzed using ANOVAs with $\log_2$ (*Target* quantity/*Rplp2* quantity) as the dependent variable. Independent variables for comparing mammary gland transcript levels were *Mcs1b* genotype and DMBA exposure. *Mcs1b* genotype and tissue source were independent variables for mammary carcinoma and non-diseased mammary tissue QPCRs. Fisher's PLSD tests were used to compare groups following a significant F-test ($\alpha \leq 0.05$). Statview software (SAS Institute) was used.

Cloning of Mier3 Splice Variants

Splice variants for rat *Mier3* were amplified from susceptible and *Mcs1b* resistant rat mammary gland cDNAs by standard PCR and cloned into a pCR® 2.1-TOPO® vector (Life Technologies) according to manufacturer specifications. Clones were transformed into chemically-competent DH5α E. coli cells, plated on Miller's LB Agar (Amresco) plates containing 50 μg/ml Kanamycin and coated with 40 μl of 40 mg/ml X-gal, and incubated at 37° C overnight. White colonies were picked from each plate and inoculated

into individual wells of a 96 well plate containing 40 µl Miller's LB broth (Amresco) and incubated for 2 hours while shaking. 0.2 µl of inoculated media was amplified by PCR using M13 reverse and T7 promoter primers specific for flanking regions of the multiple cloning site within the pCR® 2.1-TOPO vector. PCR products were resolved on 1% agarose TBE gels and stained with SYBR Gold. Large and small bands corresponding to full-length and spliced *Mier3*, respectively, were counted for each sample. Analysis was performed on the proportion of full-length bands in the total number counted. Proportions were arcsine-transformed and analyzed by a student's t-test.

Results

Mcs1b Potential Candidate Open Reading Frame Sequences Yields Are Not Different Between WF and COP

As shown in Figure 5, rat *Mcs1b* was found to contain thirteen potential candidate gene transcripts as well as sequence orthologous to human *5q11.2*, a GWAS-identified breast cancer risk associated allele marked by SNP *rs889312* (Easton et al., 2007). To prioritize potential candidates, I resequenced conserved protein coding ORFs that were within the 1.8 Mb interval that delimited *Mcs1b,* and based on RT-PCR gel electrophoresis, were expressed in mammary glands of susceptible WF and *Mcs1b* resistant females (lines N3 and T). Transcripts from *Gpbp1, Map3k1, Mier3, Ankrd55, Il6st, Il31ra, Ddx4, Slc38a9,* and *Ppap2a* genes were detected in mammary gland total RNA pools from each genotype by RT-PCR. No genetic variants were identified between susceptible WF and *Mcs1b* resistant genotype ORFs or 3' UTRs for these transcripts. Nucleotide sequences were submitted to NLM-NCBI and the GenBank

accession numbers are provided in Table 9.

Four of the *Mcs1b* candidate genes are predicted transcripts based on sequence containing gene-coding properties, e.g. intron-exon boundaries and polyadenylation signals, or sequence similarity to known expressed transcripts: *Actbl2*, *ENSRNOG00000013098*, *C5orf35* and *U6 snRNA* (labeled with asterisks in Figure 5). Rat *Actbl2* was identified as a pseudogene and is located outside rat genomic sequence orthologous to the human *5q11.2* haplotype block that associates with breast cancer risk. Predicted transcript *ENSRNOG00000013098* was listed on the Ensembl genome browser (Flicek et al., 2011). I found no evidence by RT-PCR of a transcript from *Actbl2* or *ENSRNOG00000013098* in total RNA samples from multiple susceptible and *Mcs1b* resistant mammary glands or in rat mixed tissue total-RNA samples that included embryo, brain, testes, ovary, thymus, spleen, and liver. Since cDNA was not attainable and *Actbl2* was predicted to be a single-exon transcript, we sequenced genomic DNA spanning this predicted pseudogene and found no sequence differences between WF and COP alleles.

Rat *C5orf35* is an ortholog of human *C5ORF35*. This gene was recently identified to have a SET domain and renamed *SETD9* (SET domain containing 9). SET family members are histone methyltransferases involved in regulating chromatin structure by methylation of lysine residues on tails of histones. Dysregulation of SET proteins plays roles in the progression of many cancer types. For instance, the SET member *EZH2* has been implicated in the development of breast and prostate cancer (Kleer et al., 2003; Yu et al., 2007). Although the human *C5ORF35* protein has not been fully characterized it could be a potential candidate based on its sequence similarity to

SET proteins. However, rat *C5orf35* was not present in any total RNA samples tested from various rat tissues. Additionally, I successfully amplified *C5ORF35* from human thymus, spleen, and ovary, but not human breast tissue cDNA (Figure 16A) suggesting that human *C5ORF35* is not highly expressed in the breast. However, in an Oncomine (Rhodes et al., 2004) database search I found that other groups have reported detection of *C5ORF35* in human breast carcinoma and non-diseased breast tissue. Using bioinformatic analysis I discovered that the annotated 5'- and 3'- UTRs of human *C5ORF35* are poorly conserved between humans and rodents (Figure 16B); therefore, we concluded that *C5ORF35* is a human, but not a rat transcript.

A predicted small nuclear RNA (snRNA) at rat position Chr*2*:43765811-43765918 named *U6* or *ENSRNOG00000034909* is estimated to be 108 bp on the forward strand. I noted that *ENSRNOG00000034909* sequence aligned to approximately 100 distinct regions of the rat genome using both NCBI/BLAST and UCSC/BLAT (den Dekker et al., 2012; van Boxtel et al., 2011) (Table 11). Because of the highly repetitive nature of the sequence, I was unable to design specific probes to determine if this predicted single exon gene was transcribed from rat *Mcs1b.*

Mcs1b Potential Candidate Gene Expression Levels Uncover Mier3 as a Potential Gene Candidate

Rat *Mcs1b* did not contain any protein coding genetic variation between *Mcs1b* susceptible and resistant alleles; therefore, rat *Mcs1b* may contain variation in one or more non-protein-coding regulatory elements that differentially control gene expression between mammary cancer-susceptible and resistant genotypes. To test this hypothesis, I

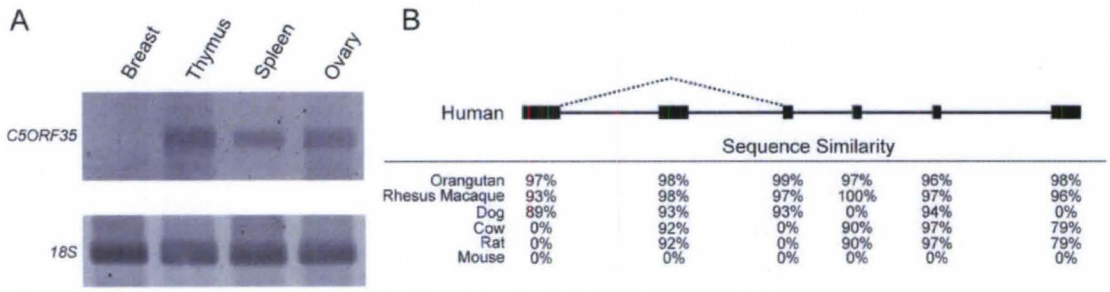


Figure 16: Predicted human *C5ORF35* is not conserved in rats. A; Predicted *C5ORF35* transcript was detected in multiple human tissues by RT-PCR. B; Splice variants of *C5ORF35* detected in human tissue and sequence similarity between human *C5ORF35* exons and other species. A *C5ORF35* splice variant that did not contain exon 2 was detected in human thymus and spleen samples. *C5ORF35* is not conserved between primates and rodents based on degeneration of 5' and 3'UTR sequences in rodents.

Table 11:

BLAT alignment of ENSRNOG00000034909 sequence against UCSC Genome Browser Rat DNA build (Baylor 3.4/rn4 assembly)

Transcript ID	Transcript Description	Chromosome	Start Position	End Position	Orientation	Score	%ID	Span (bps)
No rat transcripts annotated in browser		*Chr:2*	43765811	43765918	+	108	100.00	108
RefSeq ID: NM_153738	Rat Prlpn intronic region	*Chr:17*	43789233	43789333	-	82	88.20	101
No rat transcripts annotated in browser		*Chr:6*	136766481	136766580	-	79	88.70	100
RefSeq ID: NM_001108397.1	Rat Sfrs14 intronic region	*Chr:16*	19667327	19667415	-	78	93.10	89
RefSeq ID: NM_001191653.1	Rat Tanc2 intronic region	*Chr:10*	95198700	95198796	-	78	89.40	97
No rat transcripts annotated in browser		*Chr:7*	98699371	98699471	-	77	91.10	101
RefSeq ID: NM_173101.1	Rat Myo1e intronic region	*Chr:8*	74882368	74882468	-	75	89.90	101
No rat transcripts annotated in browser		*Chr:4*	110618884	110618978	-	75	92.80	95
No rat transcripts annotated in browser		*Chr:11*	63198076	63198175	-	75	89.70	100
No rat transcripts annotated in browser		*Chr:1*	103440155	103440263	+	75	90.50	109
No rat transcripts annotated in browser		*Chr:1*	75525663	75525763	+	75	89.90	101
No rat transcripts annotated in browser		*Chr:3*	69993031	69993128	-	74	90.70	98
No rat transcripts annotated in browser		*Chr:2*	19495997	19496085	-	74	90.70	89
No rat transcripts annotated in browser		*Chr:1*	245228475	245228564	-	74	91.20	90
RefSeq ID: NM_031337	Rat Siat9 intronic region	*Chr:4*	105395193	105395722	-	73	95.20	530
No rat transcripts annotated in browser		*Chr:7*	79649136	79649236	+	73	88.80	101
No rat transcripts annotated in browser		*Chr:20*	29978369	29978469	+	73	88.80	101
No rat transcripts annotated in browser		*Chr:17*	25409944	25410031	+	73	90.60	88
RefSeq ID: NM_001011915	Rat Plekhc1 intronic region	*Chr:15*	21384589	21384682	+	73	88.80	94
RefSeq ID: NM_001011915.1	Rat Fermt2 intronic region	" "	" "	" "	""	" "	" "	" "
No rat transcripts annotated in browser		*Chr:2*	173740365	173740453	+	72	89.60	89
No rat transcripts annotated in browser		*Chr:9*	112012955	112013153	+	71	94.00	199
No rat transcripts annotated in browser		*Chr:2*	124284322	124284416	+	71	90.40	95
No rat transcripts annotated in browser		*Chr:11*	70013852	70013937	+	71	90.40	86
RefSeq ID: NM_001135761.1		*Chr:3*	17548151	17548246	-	70	93.90	96
No rat transcripts annotated in browser		*Chr:7*	135345735	135345826	+	70	88.10	92
No rat transcripts annotated in browser		*Chr:18*	27228793	27228894	+	70	86.70	102
No rat transcripts annotated in browser		*Chr:8*	31027110	31027202	-	69	90.20	93
No rat transcripts annotated in browser		*Chr:5*	15608653	15608741	-	69	92.30	89
No rat transcripts annotated in browser		*Chr:2*	240633692	240633780	-	69	92.30	89
Rat mRNA from GenBank, ID: BC166504	Rat Fam189a2 intronic region	*Chr:1*	227408075	227408163	-	69	92.30	89
No rat transcripts annotated in browser		*Chr:1*	18291369	18291554	+	69	92.70	186
RefSeq ID: NM_001170548.1	Rat Thumpd3 intronic region	*Chr:4*	148933048	148933143	+	68	85.60	96
Rat mRNA from GenBank, ID: FQ230829	No identified homolog	*Chr:19*	41214051	41214136	+	68	84.40	86
No rat transcripts annotated in browser		*Chr:14*	18935943	18936031	+	68	87.30	89
No rat transcripts annotated in browser		*Chr:9*	47499507	47499595	-	67	91.00	89

Table 11 continued

Transcript ID	Transcript Description	Chromosome	Start Position	End Position	Orientation	Score	%ID	Span (bps)
RefSeq ID: NM_001197907.1	Rat Oxr1 intronic region	*Chr:7*	77197714	77197802	-	67	91.00	89
RefSeq ID: NM_001171177.2	Rat Tmtc2 intronic region	*Chr:7*	43891984	43892076	-	67	84.50	93
No rat transcripts annotated in browser		*Chr:7*	11235427	11235515	-	67	91.00	89
RefSeq ID: NM_031057	Rat Aldh6a1 intronic region	*Chr:6*	108710588	108710676	-	67	91.00	89
No rat transcripts annotated in browser		*Chr:6*	93546629	93546717	-	67	91.00	89
No rat transcripts annotated in browser		*Chr:6*	72766763	72766851	-	67	91.00	89
RefSeq ID: NM_001134463.1	Rat Camkmt intronic region	*Chr:6*	8704054	8704142	-	67	91.00	89
No rat transcripts annotated in browser		*Chr:5*	151444756	151444844	-	67	91.00	89
RefSeq ID: NM_001108005.1	Rat Mast2 intronic region	*Chr:5*	136630633	136630721	-	67	91.00	89
No rat transcripts annotated in browser		*Chr:5*	60215398	60215486	-	67	91.00	89
No rat transcripts annotated in browser		*Chr:5*	56854404	56854492	-	67	91.00	89
No rat transcripts annotated in browser		*Chr:3*	36733997	36734085	-	67	91.00	89
No rat transcripts annotated in browser		*Chr:3*	35650885	35650973	-	67	91.00	89
No rat transcripts annotated in browser		*Chr:2*	233090377	233090465	-	67	91.00	89
No rat transcripts annotated in browser		*Chr:2*	168539521	168539609	-	67	91.00	89
RefSeq ID: NM_001034961	Rat Sohlh2 intronic region	*Chr:2*	144422171	144422259	-	67	91.00	89
No rat transcripts annotated in browser		*Chr:2*	96996138	96996226	-	67	91.00	89
No rat transcripts annotated in browser		*Chr:2*	35329196	35329284	-	67	91.00	89
No rat transcripts annotated in browser		*Chr:18*	13043912	13044000	-	67	88.90	89
RefSeq ID: NM_001004020.1	Rat Tmprss11b intronic region	*Chr:14*	23014788	23014876	-	67	91.00	89
No rat transcripts annotated in browser		*Chr:14*	9590393	9590481	-	67	91.00	89
No rat transcripts annotated in browser		*Chr:10*	60526129	60526217	-	67	91.00	89
RefSeq ID: NM_001191634.1	Rat Gbf1 intronic region	*Chr:1*	251394207	251394295	-	67	91.00	89
Rat mRNA from GenBank, ID: FQ232940	No identified homolog	*Chr:1*	23671516	23671604	-	67	91.00	89
No rat transcripts annotated in browser		*Chr:1*	18892950	18893050	-	67	85.40	101
RefSeq ID: NM_012774.1	Rat Gpc3 intronic region	*Chr:X*	139315241	139315329	+	67	91.00	89
No rat transcripts annotated in browser		*Chr:7*	127637150	127637238	+	67	91.00	89
No rat transcripts annotated in browser		*Chr:6*	72766085	72766173	+	67	91.00	89
RefSeq ID: NM_001134628.1	Rat RGD1564943 (Predicted) intronic region	*Chr:5*	148336835	148336923	+	67	91.00	89
No rat transcripts annotated in browser		*Chr:4*	102436898	102436986	+	67	91.00	89
RefSeq ID: NM_022217.1	Rat Amph intronic region	*Chr:17*	53682453	53682541	+	67	91.00	89
No rat transcripts annotated in browser		*Chr:17*	48877826	48877914	+	67	91.00	89
No rat transcripts annotated in browser		*Chr:14*	111622041	111622131	+	67	89.90	91
No rat transcripts annotated in browser		*Chr:10*	89929380	89929468	+	67	91.00	89
No rat transcripts annotated in browser		*Chr:X*	132443461	132443547	+	66	91.00	87
RefSeq ID: NM_001106706.1	Rat Ttc27 intronic region	*Chr:6*	20562368	20562456	-	65	89.70	89
No rat transcripts annotated in browser		*Chr:4*	26126909	26126997	-	65	91.80	89
No rat transcripts annotated in browser		*Chr:2*	216457973	216458061	-	65	89.70	89
No rat transcripts annotated in browser		*Chr:18*	33879457	33879556	-	65	92.40	100

Table 11 continued

Transcript ID	Transcript Description	Chromosome	Start Position	End Position	Orientation	Score	%ID	Span (bps)
RefSeq ID: NM_001007145	Rat Catna1 intronic region	*Chr:18*	27639818	27639906	-	65	89.70	89
No rat transcripts annotated in browser		*Chr:18*	265348	265436	-	65	89.70	89
No rat transcripts annotated in browser		*Chr:14*	40179238	40179338	-	65	84.30	101
No rat transcripts annotated in browser		*Chr:11*	32577174	32577262	-	65	89.70	89
No rat transcripts annotated in browser		*Chr:7*	113696949	113697045	+	65	92.30	89
No rat transcripts annotated in browser		*Chr:2*	190279006	190279094	+	65	89.70	89
No rat transcripts annotated in browser		*Chr:18*	16341735	16341823	+	65	89.70	89
No rat transcripts annotated in browser		*Chr:1*	231873185	231873266	+	65	86.60	82
Rat mRNA from GenBank, ID: FQ227174	No identified homolog	*Chr:17*	15279892	15279979	-	64	89.50	88
No rat transcripts annotated in browser		*Chr:14*	77213446	77213534	-	64	89.50	89
Rat mRNA from GenBank, ID: DQ100481	LINE-1 retrotransposon nucleic acid binding protein mRNA	*Chr:1*	25793445	25793533	-	64	84.10	89
No rat transcripts annotated in browser		*Chr:6*	116937744	116937821	+	64	91.10	78
No rat transcripts annotated in browser		*Chr:14*	32819567	32819654	-	63	89.10	88
RefSeq ID: NM_001107047.1	Rat Ikzf3 intronic region	*Chr:10*	87362236	87362320	-	63	90.50	85
RefSeq ID: NM_001007630.1	Rat Zfand6 intronic region	*Chr:1*	140906354	140906442	-	63	88.40	89
No rat transcripts annotated in browser		*Chr:1*	56584008	56584096	-	63	88.40	89
No rat transcripts annotated in browser		*Chr:9*	103107997	103108085	+	63	88.40	89
No rat transcripts annotated in browser		*Chr:3*	104572110	104572198	+	63	88.40	89
No rat transcripts annotated in browser		*Chr:1*	194118835	194118923	+	63	88.40	89
RefSeq ID: NM_001135718.1	Rat Prex1 intronic region	*Chr:3*	157849079	157849155	+	62	89.20	77
No rat transcripts annotated in browser		*Chr:16*	36604669	36604747	-	61	91.40	79
No rat transcripts annotated in browser		*Chr:14*	54341337	54341425	-	60	90.00	89
No rat transcripts annotated in browser		*Chr:1*	232461672	232461742	-	60	91.20	71
RefSeq ID: NM_001191669.1	Rat Thsd7b intronic region	*Chr:13*	42749088	42749194	+	60	97.00	107

(Hits with score ≥ 60)

measured mammary gland transcript levels of genes located at *Mcs1b* in 12-week old virgin female rats that were exposed to DMBA at 50-55 days and age matched controls without DMBA. These studies focused on mammary gland transcript levels due to the mammary gland autonomous nature of *Mcs1b* previously discussed. Twelve-week old animals were used because this is the age after the acute phase for DMBA-toxicity and before induced frank mammary carcinomas are detectable in susceptible strains.

Differences in expression between genotypes were analyzed by non-parametric Mann-Whitney tests and the effects of *Mcs1b* genotype and DMBA exposure on candidate gene transcript levels were analyzed by two-way ANOVA (Table 12). Transcript levels of *Ankrd55, Il31ra, Ddx4, Slc38a9,* or *Ppap2a* were not significantly different between *Mcs1b* resistant and susceptible mammary glands exposed to DMBA or not. However, the effect of *Mcs1b* genotype was statistically significant ($P < 0.05$) for *Gpbp1*, *Mier3*, and *Map3k1*. Furthermore, there was a significant effect of DMBA exposure on *Map3k1* transcript levels (P=0.0003) and the interaction between *Mcs1b* genotype and DMBA exposure approached statistical significance for *Map3k1* (P=0.0588). Additionally, there was an effect of *Mcs1b* genotype on *Il6st* expression by ANOVA; however, there was no difference in expression between genotypes in mammary glands not exposed to DMBA (P=0.1137) and expression only approached significance in glands exposed to DMBA (P=0.0734). When *Mcs1b* genotypes were compared by DMBA exposure, mammary gland transcript levels were significantly different for *Gpbp1, Mier3,* and *Map3k1* between *Mcs1b* resistant and susceptible mammary glands that were not exposed to DMBA. However, significant expression differences between *Mcs1b* resistant and susceptible genotypes were sustained only for

Table 12:

Analysis and statistics of *Mcs1b* potential candidate gene mammary gland transcript levels in *Mcs1b* -resistant and susceptible genotypes at 12 weeks of age

	Two-way ANOVA *F* test *P* values				Log_2 Target/Rplp2 mean ± SD (*n*)		
Target	***Mcs1b* Genotype**	**Exposure**	**G X E**	**Exposure**	**Susceptible**	***Mcs1b* Resistant**	**P-value[a]**
Gpbp1	0.0101	0.209	0.6422	Control	0.586 ± 0.600 (34)	0.044 ± 0.734 (29)	0.002
				DMBA	0.281 ± 1.309 (45)	−0.097 ± 1.246 (42)	0.1716
MIER3	0.0023	0.7911	0.6682	Control	0.115 ± 0.594 (34)	−0.522 ± 1.278 (34)	0.0104
				DMBA	0.154 ± 1.557 (45)	0.688 ± 1.943 (48)	0.024
Map3k1	0.0002	0.0003	0.0588	Control	−0.092 ± 0.818 (34)	−0.725 ± 0.767 (32)	0.0019
				DMBA	0.105 ± 0.564 (47)	−0.104 ± 0.651 (45)	0.1036
Ankrd55	0.4694	0.2025	0.9019	Control	−0.691 ± 0.678 (24)	−0.826 ± 1.108 (22)	0.618
				DMBA	−0.377 ± 1.296 (17)	−0.567 ± 1.006 (22)	0.609
Il6st	0.0199	0.1744	0.8435	Control	−0.066 ± 0.755 (36)	−0.418 ± 1.054 (33)	0.1137
				DMBA	0.189 ± 1.006 (44)	−0.227 ± 1.181 (48)	0.0734
Il31ra	0.2869	0.8674	0.9928	Control	−0.331 ± 1.072 (24)	−0.559 ± 0.761 (23)	0.4072
				DMBA	−0.368 ± 0.942 (20)	−0.592 ± 1.159 (23)	0.4949
Ddx4	0.0555	0.5442	0.4045	Control	−0.107 ± 0.983 (36)	−0.359 ± 0.911 (33)	0.2748
				DMBA	−0.055 ± 1.122 (18)	−0.690 ± 1.575 (17)	0.1769
Slc38a9	0.1008	0.3929	0.973	Control	−0.285 ± 0.600 (24)	−0.575 ± 0.681 (23)	0.1284
				DMBA	−0.144 ± 0.970 (20)	−0.422 ± 0.954 (23)	0.3499
Ppap2a	0.3918	0.8314	0.5788	Control	−0.385 ± 0.632 (24)	−0.447 ± 0.765 (23)	0.7629
				DMBA	−0.315 ± 1.357 (20)	−0.605 ± 1.029 (23)	0.4315

[a]Fisher PLSD test *P*- values from comparing susceptible and *Mcs1b* -resistant genotypes by exposure.

Mier3 when females given DMBA were compared between genotypes (P=0.024). Transcript levels of *Gpbp1* and *Map3k1* were not different between genotypes when DMBA-exposed females were evaluated.

Mammary gland transcript levels were lower in *Mcs1b* resistant genotype females for all genes with a significant difference between genotypes: *Gpbp1*, *Mier3*, and *Map3k1*. *Mier3* mean transcript levels were approximately 4.5-fold lower in *Mcs1b* resistant compared to susceptible genotype mammary glands whether animals were exposed to DMBA or not (Table 12). Thus, exposure to DMBA had no appreciable effect on *Mier3* differences between susceptible and *Mcs1b* resistant genotype females. No significant differences in *Mier3* transcript levels were detected between *Mcs1b* resistant and susceptible genotypes in spleen, thymus, ovary, or brain tissues (Figure 17). This suggests that *Mier3* transcript level differences between *Mcs1b* alleles may be specific to mammary gland tissue.

Mier3 Is Expressed as Three Different Variants in the Rat Mammary Gland

Gene expression studies identified *Mier3* as a potential rat mammary carcinoma susceptibility-related gene. Interestingly, I also noted that *Mier3* migrated as 2 bands on electrophoretic gels suggesting that different variants of *Mier3* may be being expressed; therefore, I sequenced these variants. Moreover, I also sought to check if these *Mier3* variants were expressed in the mammary gland and whether they were expressed at the same level between WF and COP tissue.

Mier3 was cloned into a TOPO vector and many clones were sequenced. Sequencing showed that both *Mier3* variants were expressed in the mammary gland

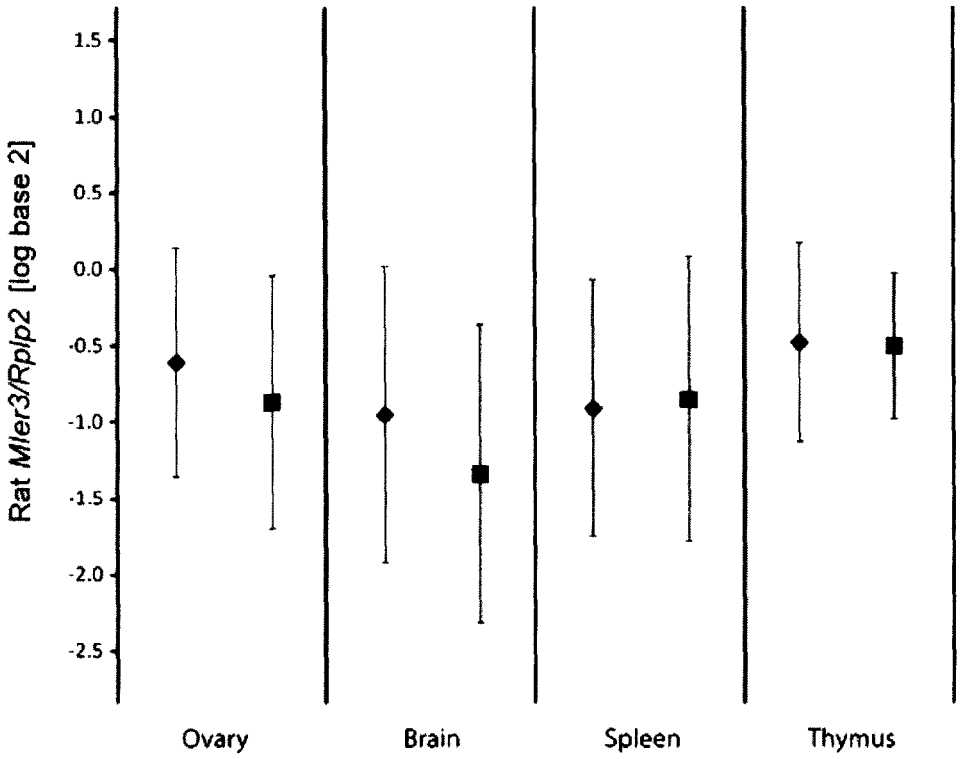
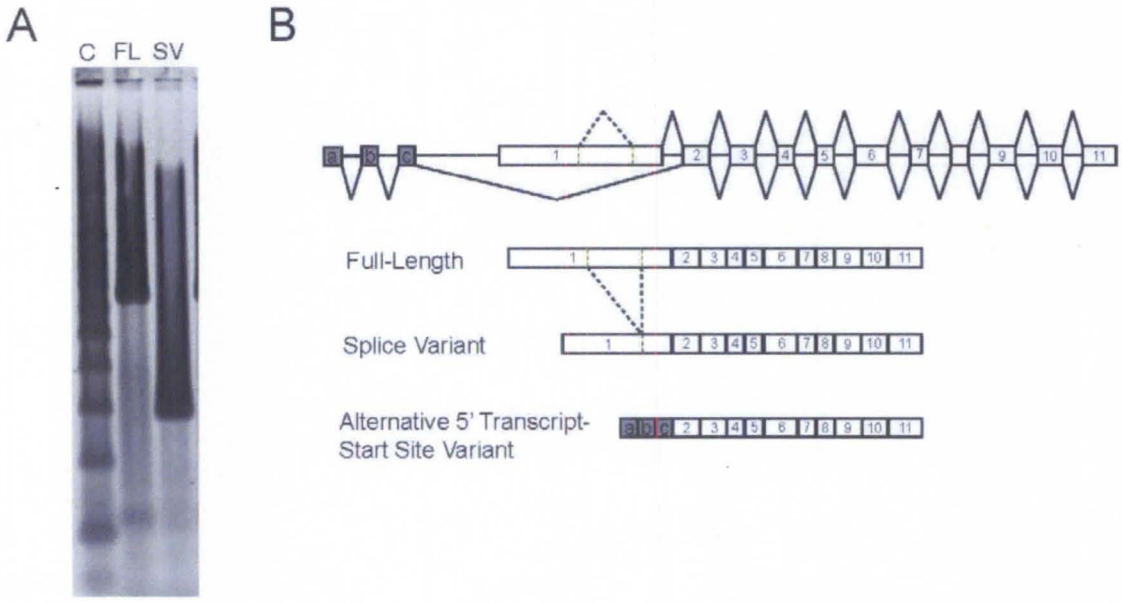


Figure 17: *Mier3* is not differentially expressed between WF susceptible and *Mcs1b* resistant WF.COP congenic females in rat ovary, brain, spleen, or thymus tissues. Expression (QPCR) levels are represented as log base 2 mean quantities of rat *Mier3* relative to rat *Rplp2* expression +/- SD for each tissue listed on the X axis. Comparisons between WF susceptible (◆) and *Mcs1b* resistant (■) females for each tissue were not statistically significant in one-way ANOVA with post hoc Fischer's PLSD test (Ovary: P=0.4755, N=8 WF, N=13 *Mcs1b* resistant line N3; Brain: P=0.3788, N=11 WF, N=10 *Mcs1b* resistant line N3; Spleen: P=0.8854, N=8 WF, N=13 *Mcs1b* resistant line N3; Thymus: P=0.9277, N=8 WF, N=12 *Mcs1b* resistant line N3).

(Figure 18A). One sequence identified the full length *Mier3* transcript (Figure 18B). The other showed intra-exonal splicing of exon 1 from bp 587 to bb 950 resulting in a 3611 bp splice variant. This variant disrupts the normal ATG start site for translation and results in a new start site at position 1013 in the transcript. This new start site is in frame and would result in a 23 amino acid truncation at the N terminus of the protein (Figure 18C). It is not clear what the consequences of this truncation are on the function of the Mier3 protein. Additionally, it is not known whether this splice variant is conserved between species.

I attempted to design Taqman gene expression assays to test whether there is a difference in expression of these splice variants between WF and COP mammary gland tissue samples; however, due to the constraints of the Taqman assay design this was not possible. Instead, I resorted to cloning *Mier3* amplicons and counting clones (Table 13). *Mier3* amplicons were cloned into a pCR2.1-TOPO vector, clones were picked and amplified and run on agarose gels. 96 clones were counted per sample and analyzed as the proportion of full length *Mier3* clones out of the total number of clones (containing full length *Mier3* plus spliced *Mier3*) counted. Clone assay results demonstrated there was not a significant difference in *Mier3* splice variant expression versus the full length variant expression between WF and resistant WF.COP mammary glands (Figure 18D). Overall, the full length *Mier3* transcript expression was much higher than that of the splice variant (76% and 79% for resistant N3 and susceptible WF samples, respectively).

Further bioinformatic analysis showed that *Mier3* sequence in the mouse and human have an alternatively used exon upstream of exon 1 that results in an alternative start site. The alternate exon usage transcribes three short exons and excludes



C

```
Mier3 Full Length                  ----MTLCFIPVGSLSSEDHDFDPTAEMLVHDYDDERTLEEEELMGDGKNFSSEIEDLEK MW=60.86 kD
Mier3 Splice Variant               --------MITIGSLSSEDHDFDPTAEMLVHDYDDERTLEEEELMGDGKNFSSEIEDLEK MW=60.41 kD
Mier3 Alternate 5' Start Variant   MAEASFGSSSPVGSLSSEDHDFDPTAEMLVHDYDDERTLEEEELMGDGKNFSSEIEDLEK MW=61.11 kD
                                     .:**************************************************
```

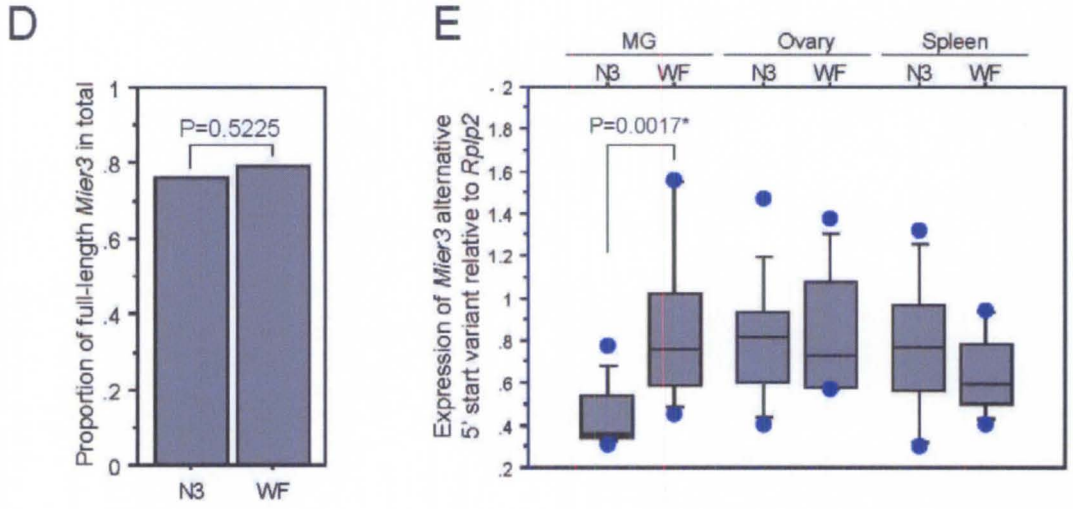


Figure 18: Three different *Mier3* variants are expressed in the rat mammary gland. A; Image of electrophoretic gel of full-length *Mier3* and *Mier3* splice variant. B; Diagram of gene structure of full-length, spliced, and alternative 5'-transcript start variant of *Mier3*. C; Amino acid sequence comparison of the three different variants at the N-terminus. Molecular weights of each protein are predicted based on sequence. D; Expression levels of full-length *Mier3* compared to total *Mier3* expressed. Levels depicted as proportion of full-length *Mier3* in total. P values based on t-test of Arcsine-transformed proportion values. E; Expression levels of alternative 5'-transcription start *Mier3* variant. Expression is relative to Rplp2 expression. P values based on results of ANOVA and Fisher PLSD post-hoc test. Abbreviations: C, 100 base pair DNA ladder standard; FL, full length rat *Mier3*; SV, rat *Mier3* splice variant.

Table 13:

Primer and Probe Sequences Used for Analysis of Rat Mier3 Variants

Assay Name	Primer/Probe Name	Primer/Probe Sequence
Mier3 Splice Variant Cloning and Sequencing	Mier3-F3	AGCCCGTGTATTCCAGACAG
	Mier3-R3	TGCAACCGTTGGAATTGTAG
Mier3 Alternate 5' Transcription Start Site Sequencing	Mier3-alt-5'-end-F3	CTGGCGATTGGCTCAGG
	Mier3-alt-5'-end-R3	TGTCTAGGACTTTTTCATTTCCA
Mier3 Alternate 5' Transcription Start Site Taqman Expression	Mier3-alt5'-1-F	ATGGCGGAGGCTTCCTTT
	Mier3-alt5'-1-R	TCAAAATCATGATCCTCAGAAGACA
	Mier3-alt5'-1-Probe	AGCCCAGTTGGGTCT
	Mier3-alt5'-2-F	AGCCCAGTTGGGTCTTTGTCT
	Mier3-alt5'-2-R	TCATAGTCATGGACCAGCACTCA
	Mier3-alt5'-2-Probe	CTGAGGATCATGATTTTG

exon 1 (Figure 18B). I designed primers to the orthologous rat sequence and amplified and sequenced these in rat mammary glands (Table 13). To further determine whether there were expression differences in the alternative start form of *Mier3* compared to full length *Mier3*, I designed Taqman assays (Table 13). The results indicate a differential expression of the alternative start end variant between WF and COP untreated mammary glands (P=0.0017) (Figure 18E). Moreover, differential expression of the alternative 5'-start variant appears to be specific to the mammary gland as differential expression was not observed in ovary and spleen tissue (Figure 18E). The alternative start site variant results in a transcription start site beginning upstream of the normal start site; however, the alternate transcription start site stays in frame with the downstream coding of *Mier3*. Although, the amino acid sequence is slightly altered at the N-terminus of the protein, there is no evidence suggesting this change should disrupt translation (Figure 18C). It is not clear what the implication of this alternative sequence has for *Mier3* function, but expression differences may be implicated in the reduced expression of total *Mier3* seen in previous gene expression assays.

Mcs1b Genotype Exhibits Pleiotropic Effects on Body Weight

GWA studies have identified several breast cancer susceptibility loci; however, little is known about how the relative risks associated with these regions are affected by the established reproductive, behavioral, and anthropometric risk factors for breast cancer (often referred to collectively as environmental factors, although some, such as height, are in part genetically determined). Travis *et al.* detected a significant association between human breast cancer risk associated SNP *rs889312* and stature in women (Travis

et al., 2010). To determine if rat *Mcs1b* might also exhibit pleiotropy, we analyzed rat body weight, which is information we routinely collect and relevant because body weight is genetically correlated to stature in humans (Czerwinski et al., 2007). Significant effects of *Mcs1b* genotype (P<0.0001) and DMBA exposure (P=0.0014) on body weight at 12 weeks of age were detected (Figure 19). The interaction between *Mcs1b* genotype and DMBA exposure was also significant (P=0.0004). Females with the *Mcs1b* resistant genotype had mean ± SD body weights of 200 ± 11 grams with DMBA (N=47) and 201 ± 7.7 grams without (N=33), which were not significantly different (P=0.7880). Comparatively, mammary cancer susceptible females had higher (P<0.0001) mean ± SD body weight at 192 ± 11 grams with DMBA (N=45) than unexposed susceptible females (N=34) who had a mean ± SD body weight of 180 ± 12 grams.

Rat Mammary Carcinomas Express Higher Mier3 Transcript Levels Compared to Normal Rat Mammary Gland Tissue

There was a significant difference in expression of *Mier3, Gpbp1, and Map3k1* between untreated N3 and WF mammary glands. However, only *Mier3* exhibited a significant effect after DMBA induction. Therefore, I next sought to determine whether there was an effect of *Mcs1b* genotype on levels of any of these transcripts in mammary carcinoma tissue. Further, *Il6st* was also included because it had been reported to be higher in rat mammary carcinomas compared to normal mammary gland tissues although in our studies it narrowly missed statistical significance (Qiu et al., 2003b).

Mier3 mRNA expression was measured in DMBA-induced mammary carcinomas from *Mcs1b* resistant genotype (N=25) and susceptible (N=28) mammary glands by

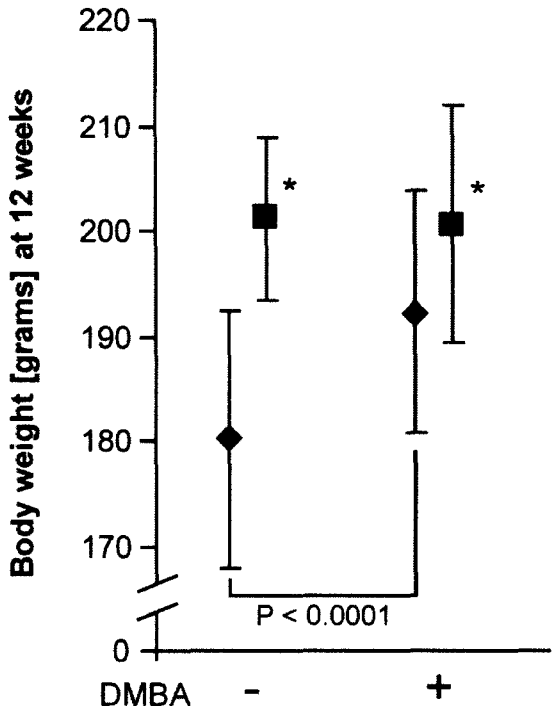


Figure 19: Rat *Mcs1b*-resistant genotype is associated with higher body weight. Lower body weight at 12 weeks of age was observed in mammary carcinoma susceptible (◆) compared with *Mcs1b*-resistant females (■) with DMBA and without ($P <$ 0.0001 and P=0.0007, respectively). Body weight was significantly higher in susceptible females that received DMBA compared to females not receiving DMBA ($P < 0.0001$).

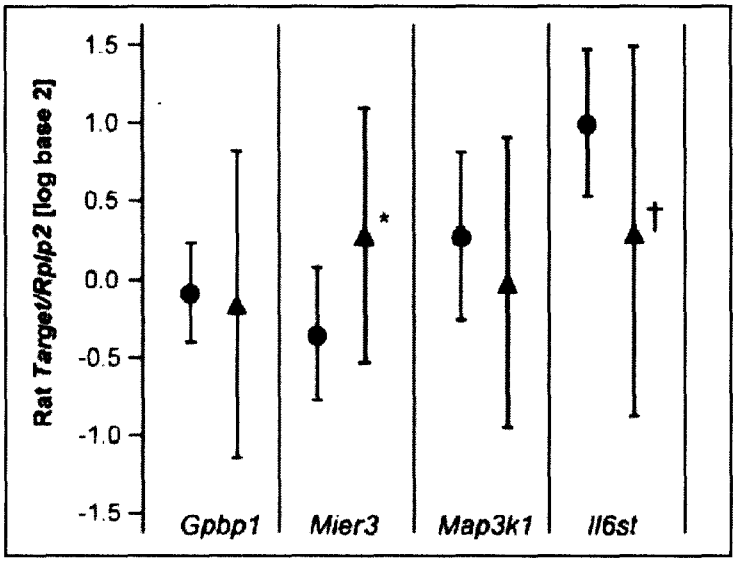
QPCR and compared to adjacent "non-diseased" mammary glands. We collected total RNA from DMBA-induced mammary carcinomas (N=1 or 2 *per* rat) and adjacent non-diseased mammary gland tissue from 21-23 week old females (N=6 *per* genotype). There were no statistically significant differences in mammary carcinoma transcript levels between *Mcs1b* genotypes for any of the four genes tested (Figure 20A). However, *Mier3* transcript levels were significantly higher (1.8-fold) in mammary carcinomas compared to non-diseased mammary tissue. We also observed that *Il6st* was potentially different between mammary carcinomas and non-diseased mammary glands but did not meet statistical significance (Figure 20A).

To determine whether *Mier3* expression differences had been recognized in human samples, I queried The Cancer Genome Atlas (cancergenome.nih.gov) gene expression database by using the Oncomine (Rhodes et al., 2004) data-mining platform. Results show that levels of human *MIER3* were, respectively, 1.33 and 1.20 fold higher in invasive ductal (N=392) and invasive lobular (N=36) breast carcinoma samples compared to pathologically normal breast tissues (N=61)($P=2.8x10^{-13}$, ductal; $P=6.3x10^{-4}$, lobular; t-tests, Figure 19B). Thus, both human/rat *MIER3/Mier3* levels are higher in breast/mammary carcinoma compared to non-diseased breast/mammary tissues.

Discussion

Within the interval delineated by SNP markers A12v and A12oo lie thirteen transcripts on rat chromosome *2* (Figure 5). To determine the mechanism by which the *Mcs1b* locus confers resistance to developing mammary carcinomas when induced with DMBA I sought to, first, determine if any SNPs in the coding regions of these transcripts

A


Rat Target/Rplp2 [log base 2]
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
Gpbp1
Mier3
Map3k1
Il6st
*
†

B

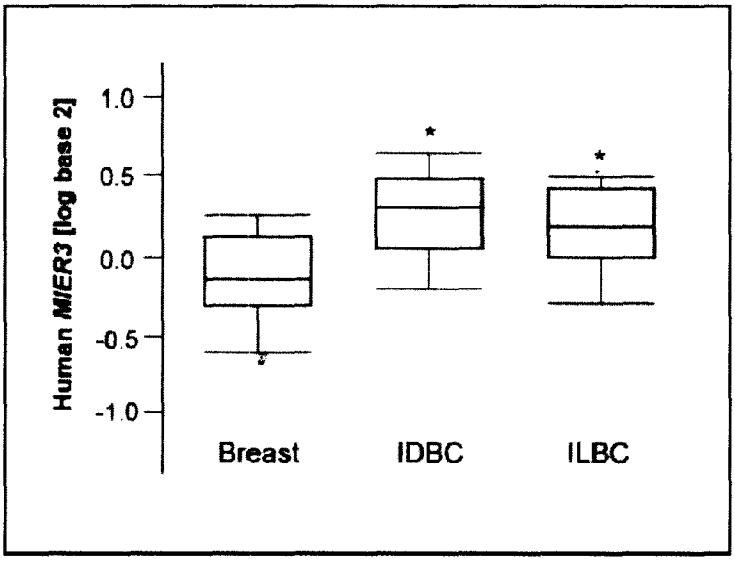

Human MIER3 [log base 2]
1.0
0.5
0.0
-0.5
-1.0
Breast
IDBC
ILBC
*
*

Figure 20: Rat *Mier3* transcript levels were significantly higher in DMBA-induced mammary carcinomas compared to non-diseased mammary gland tissue and human *MIER3* was significantly higher in breast carcinomas compared with pathologically normal breast tissues. A; Mean ± SD are graphed for each variable. Expression of *Mier3* mammary carcinoma (▲) is significantly different than in adjacent non-diseased mammary gland tissue (●; *, P=0.0120; †, P=0.0569). B, Oncomine (www.oncomine.org) was used to query The Cancer Genome Atlas (cancergenome.nih.gov) gene expression database. Box plots of $\log_2$ median centered *MIER3* transcript levels are shown for invasive ductal breast carcinomas (IDBC, N=392) and invasive lobular breast carcinomas (ILBC, N=36) compared with pathologically normal breast tissues (Breast, N=61). *MIER3* transcript levels are significantly elevated in both tumor types compared to normal tissue (*, $P < 0.05$).

existed between susceptible and resistant lines. Of the thirteen, 9 were amplified and sequenced in rat spleen and thymus tissue: *Gpbp1, Map3k1, Mier3, Ankrd55, Il6st, Il31ra, Ddx4, Slc38a9,* and *Ppap2a.* The other four were predicted transcripts and had not been validated. These were identified to be pseudogenes upon further investigation using bioinformatic tools. Sequencing results showed that there was no variation between susceptible WF and resistant WF.COP cDNAs. This suggests that the basis for *Mcs1b* COP-conferred resistance is regulatory in nature; therefore, further studies would be needed to identify potential gene candidates.

There are many mechanisms that could be responsible for the *Mcs1b*-conferred mammary carcinoma resistance phenotype, e.g., a variant in a promoter or enhancer disturbing transcription or a variant in a splice site disrupting mRNA processing. To investigate this mechanism the most suitable starting point was to examine expression levels of the transcripts. Gene expression of the nine validated transcripts was measured in mammary glands from animals treated with DMBA or not. Three transcripts were differentially expressed between susceptible and resistant untreated mammary gland tissue samples: *Gpbp1*, *Mier3*, and *Map3k1*. However, only *Mier3* was statistically different between susceptible and resistant mammary glands after DMBA was administered. Moreover, *Il6st* was previously reported to be upregulated in mammary carcinomas versus non-diseased mammary gland tissue (Qiu et al., 2003a) but missed statistical significance in my studies. The interaction between *Mcs1b* genotype and DMBA exposure approached statistical significance for *Map3k1*; however, *Map3k1* expression was not differentially expressed in DMBA-treated mammary glands. The loss in statistical significance between DMBA-exposed susceptible and *Mcs1b* resistant

females for *Map3k1* was due to an increase in mean level of *Map3k1* (P=0.0003) in the *Mcs1b* resistant genotype females with DMBA compared to age-matched controls of the same genotype without DMBA (Table 13). *Map3k1* levels were not different (P=0.2038) between susceptible WF mammary glands with or without DMBA. Regardless, *Map3k1* expression was not different between genotypes in DMBA-exposed mammary glands suggesting it was not a likely candidate. Hence, *Mier3* stood out as the most likely candidate. Indeed, further study of *Mier3* in rat tumor samples showed that *Mier3* expression was also increased in mammary carcinomas compared to adjacent non-diseased mammary gland tissue. This was reinforced by Oncomine microarray data confirming significantly elevated MIER3 expression in human breast cancer samples compared to normal breast tissue.

Mier3 (*M*esoderm *i*nduction *e*arly *r*esponse 1, family member *3*) has not been characterized. However, based on sequence similarity, human MIER3 has been identified as having two key domains: ELM2 and SANT (2012). The MIER3 ELM2 domain is from amino acid position 174 to 272. The ELM2 (Egl-27 and MTA1 homology 2) domain was initially identified in the protein MTA1 a component of the NuRD chromatin regulatory complex and is involved in recruiting HDAC leading to changes in chromatin structure and resulting in transcriptional repression (Ding et al., 2003; Solari et al., 1999). The SANT domain is located from amino acid position 277 to 329. SANT domains are present in nuclear receptor co-repressor proteins and in the subunits of many chromatin-remodeling complexes (Aasland et al., 1996). Moreover, SANT domains are characterized by tandem repeats of three alpha-helices arranged in a helix-turn-helix motif, each alpha helix containing a bulky aromatic residue making them

similar to the DNA-binding domain of Myb proteins involved in DNA binding and transcriptional repression (Grune et al., 2003; Vargova et al., 2011). Taken together, the presence of these domains suggests that *Mier3* may bind DNA or regulate chromatin structure.

Another *Mier* family member, *Mier1*, has also been implicated in breast cancer progression (McCarthy et al., 2008a). Human and rat MIER3/Mier3 (GenBank ref| NP_689835.3 and NP_001161472.1) gene products share 93% amino acid sequence identity, and human MIER3 and MIER1 (GenBank ref|NP_001071172.1) have 54% identical amino acids based on BLAST (van Boxtel et al., 2011). *Mier1* is a transcriptional regulator that was discovered during a screen for fibroblast growth factor response genes (Paterno et al., 1997; Paterno et al., 1998; Thorne et al., 2005). Notably, MIER1 physically interacts with ERα, Sp1, and Creb-binding protein (Blackmore et al., 2008; Ding et al., 2004b; McCarthy et al., 2008b). The impact of hormone receptors and co-regulators on breast cancer development has been noted. Of particular interest, MIER1 contains a C-terminal LxxLL motif referred to as LXD (McCarthy et al., 2008a). The LXD is a highly conserved sequence shown to interact with variety of hormone receptors (Heery et al., 1997). Moreover, studies have shown that different arrangements of the LXD confer varying specificities for different hormone receptors. For example, whereas a single LXD is sufficient for activation by ERα, different combinations of two, appropriately spaced, LXDs are required for actions of the thyroid hormone, retinoic acid, peroxisome proliferator-activated, or progesterone receptors (McInerney et al., 1998b). Strikingly, MIER1 contains a single C-terminal LXD while MIER3 contains 2 LXDs. Concordant with this, studies demonstrate that loss of *MIER1* may contribute to

breast cancer progression while our data indicate that reduction in *Mier3* expression is protective against mammary carcinoma development. Taken together, a potential functional difference between *MIER1* and *MIER3* may be that a difference in the number of LXD motifs between them results in physical interactions with different nuclear hormone receptors.

Amplification of *Mier3* in rat thymus cDNA for sequencing resulted in two distinct bands on agarose gel. I further examined these bands by, first, confirming their expression in mammary tissue and sequencing. Sequencing revealed a full-length transcript and a splice variant in which a 363 bp segment was internally spliced out of exon 1. Additionally, the human and mouse *MIER3/mier3* annotated sequences indicate that additional exons may exist upstream of exon 1. To test this, I successfully amplified and sequenced the rat orthologous region. Sequencing identified alternative usage of three short exons upstream of the original exon 1 being transcribed and skipping exon 1. Further, I measured gene expression of both variants. No differences were identified between the full-length and splice variant forms; however, the alternative start site variant was specifically down-regulated in resistant compared to susceptible mammary gland tissue. This was similar to the gene expression for all *Mier3* isoforms performed previously, suggesting that the differences in total *Mier3* expression may be defined by expression of the *Mier3* alternative transcription start site variant.

Both *Mier3* variants have different translation start sites compared to the original full length *Mier3* transcript. This would change the N-terminus of the Mier3 protein for each of these variants; however, neither rat *Mier3* nor its human ortholog has been characterized and it is not clear what the consequence of these variants may be on protein

function. Examining the N-terminus with The Eukaryotic Linear Motif resource (Dinkel et al., 2012) (http://elm.eu.org) shows that full length Mier3 contains a USP7-binding motif that is lost in the intra-exonal splice variant. Additionally, the alternative transcription start variant retains this USP7-binding domain and adds a Casein kinase 1 (CK1) phosphorylation site (Dinkel et al., 2012). USP7 is a deubiquitinating enzyme most commonly involved in regulation p53 regulation by deubiquitinating it and protecting it from MDM2-mediated degradation (Li et al., 2002). CK1 is a serine/threonine kinase involved in a variety of cell signaling pathways with a myriad of targets; however, it has been recognized to be involved in the activation of the Wnt signaling pathway (Davidson et al., 2005). Therefore, the alternative start site form of *Mier3* may be more mitogenic compared to the other isotypes. Taking everything together, *Mier3* emerges as a positive regulator of cell proliferation affecting carcinogenesis in the mammary gland.

Travis *et al.* reported pleiotropic effects on stature in women associated with the breast cancer-associated SNP *rs889312* (Travis et al., 2010). During the course of our experiments we routinely gather data on body mass of the rats. I used this to test whether a correlation existed between body mass and *Mcs1b* genotype since body mass can be used as a proxy for stature in humans. Lower body weight was observed in susceptible female rats compared to resistant congenic rats at 12 weeks of age. These animals differ essentially only at the *Mcs1b* locus and resistant female rats express less *Mier3* than female rats harboring a susceptible Mcs1b allele. Therefore, this seems counter-intuitive as it is believed that *Mier3* is mitogenic and increased expression would result in increased weight gain. However, as previously mentioned, PR is known to attenuate

effects of ERα (Hsueh et al., 1976; Tseng and Gurpide, 1975). What is more, loss of estrogen activity has been identified to associate with a reduction in catabolism and an increase in weight, femur length, and bone density in male mice (Ford et al., 2011). As discussed earlier in this chapter, human *MIER3* contains 2 LXDs which may confer affinity for PR (McInerney et al., 1998b). Therefore, a likely scenario would be that in resistant congenic females expressing lower amounts of *Mier3*, there is less PR bound to *Mier3* and more that is available to inhibit ERα activity. ERα inhibition in these animals results in higher body mass compared to susceptible animals expressing higher amounts of *Mier3*. However, this is based on an assumption that *Mier3* is binding to PR supported only by sequence similarity to *Mier1* and the presence of 2 LxxLL motifs. Therefore, this is entirely speculative and more empirical work is necessary to determine the mechanism by which *Mier3* is working.

CHAPTER V

CONCLUDING REMARKS

Rat mammary carcinoma susceptibility, like human breast cancer risk, is complex as both are controlled by multiple susceptibility alleles and environmental factors. We have mapped rat *Mcs1b* to a 1.1 Mb region of rat chromosome *2* using multiple congenic lines. We found that rat *Mcs1b* is highly relevant to human breast cancer susceptibility as it contains genomic sequence orthologous to a low-penetrance breast cancer risk allele at human chromosome *5q11.2.* This human susceptibility allele was first reported by Easton *et al.*(2007) in the first population-based breast cancer risk GWAS. Human *5q11.2* has been confirmed to strongly associate with breast cancer risk in multiple independent studies of European- and Asian-descent populations (Antoniou et al., 2008; Broeks et al., 2011; Campa et al., 2011; Garcia-Closas et al., 2008a; Han et al., 2011; Zheng et al., 2010). This is the first report of a rodent complex disease susceptibility QTL with a GWAS-identified concordant human ortholog that had a probability of association below a stringent significance level of $P \leq 10^{-7}$, which is widely deemed to be required for genome-wide studies.

An experimental organism with a segregating concordant susceptibility allele implies that functional genetic studies may translate directly to human biology and disease. For example, Gould and colleagues reported that rat *Mcs5a*, a WKY strain resistance QTL that is concordant to human *MCS5A*, acted in a non-mammary cell-

autonomous manner that involves immune cells (Smits et al., 2011a). Here, we used rat genetic lines to show that *Mcs1b* controls mammary cancer susceptibility by an undetermined mammary gland cell autonomous mechanism. While our result is in agreement with previous work that concluded a majority, but not all, of the COP rat strain resistance to mammary cancer is mammary gland autonomous (Zhang et al., 1990); it further highlights that the WKY and COP strains may achieve mammary carcinoma resistance through different genetically determined cellular and molecular mechanisms that are likely genetically determined in humans as well.

Further, Gould and colleagues developed a clonogenic transplant assay using rat MEC preparations to reconstitute mammary glands in the intrascapular fat pads of recipient rats (Gould and Clifton, 1985). To date no one has demonstrated an ability to induce carcinomas in these ectopic glands. The window for DMBA to induce carcinogenesis in endogenous rat mammary glands and whole mammary gland transplants was established to be at 50-55 days of age. Herein, I show that glands formed from the injection of MECs have a longer latency compared to whole mammary gland grafts since DMBA-induced carcinogenesis is delayed to 70-75 days of age. This is a significant result for future use of this protocol.

Most common genetic variation associated with human complex disease susceptibility appears to be located in non-protein-coding DNA. Since we found no genetic variation between susceptible and resistant allele *Mcs1b* ORFs, we conclude that *Mcs1b* is likely a noncoding gene regulatory element(s), such as a transcription factor binding site or noncoding RNA. This would be similar to the hypothesized identity of the human *5q11.2* breast cancer risk associated element. Human polymorphisms that are

contained in public databases and highly correlated with human *5q11.2* breast cancer risk associated SNP *rs889312* are in non-protein-coding DNA. There are no known noncoding RNAs in either the human or rat ortholog; therefore, another type of gene regulatory element is likely responsible for or associated to susceptibility differences.

Our studies suggest that *MIER3* is a strong candidate breast cancer susceptibility gene at human *5q11.2.* We identified *Mier3* as a strong *Mcs1b* candidate gene in this study based on different *Mier3* mammary gland transcript levels between susceptible and *Mcs1b* resistant genotypes. Lower levels of *Mier3* in *Mcs1b* resistant genotype females were genetically determined and not dependent on the induction of mammary carcinogenesis by DMBA. We also found *Mier3* levels significantly lower in non-diseased rat mammary tissue compared to mammary carcinoma. Further, we queried The Cancer Genome Atlas gene expression database and noted that human *MIER3* levels were higher in both ductal and lobular breast carcinomas compared to breast tissue.

MIER3 or *mesoderm induction early response 1, family member* 3 (GenBank ref|NM_152622) is an uncharacterized gene. We determined that MIER3 localized to the nucleus. Human and rat MIER3/Mier3 (GenBank ref| NP_689835.3 and NP_001161472.1) gene products share 93% amino acid sequence identity, and human MIER3 and MIER1 (GenBank ref|NP_001071172.1) have 54% identical amino acids based on BLAST (van Boxtel et al., 2011). MIER1 physically interacts with ERα, Sp1, and Creb-binding protein (Blackmore et al., 2008; Ding et al., 2004b; McCarthy et al., 2008b). MIER1 contains one, while MIER3 has two conserved LxxLL sequences, which is a motif that facilitates nuclear hormone receptor interactions (Heery et al., 1997). A potential functional difference between MIER1 and MIER3 may be that a difference in

the number of LxxLL motifs between them results in physical interactions with different nuclear hormone receptors (McInerney et al., 1998a). In addition, MIER1 and MIER3 harbor ELM2 and SANT domains. The ELM2 domain is involved in recruitment of HDAC activity, which leads to changes in chromatin structure and results in transcriptional repression (Ding et al., 2003). Likewise, the MIER1 SANT domain functions in gene repression by interacting with Sp1 and interfering with its ability to bind to its cognate site on responsive promoters (Ding et al., 2004a). The presence of two LxxLL motifs along with the ELM2 and SANT domains suggests that MIER3 is a potential transcriptional repressor whose activity is mediated by interaction with nuclear hormone receptors. To test this idea, more functional studies are necessary to determine the mechanism by which *Mier3* is controlling mammary carcinogenesis in the rat.

In addition to *MIER3, MAP3K1* and *C5ORF35* reside within the human *5q11.2* haplotype block that associates with breast cancer risk. Even though there are no published studies in support, *MAP3K1* is often considered the candidate breast cancer susceptibility gene at *5q11.2* due to its location within the breast cancer risk associated haplotype block and known function as a serine/threonine kinase. In our rat studies, *Map3k1* was differentially expressed between susceptible and *Mcs1b* resistant congenic rats that had not been induced to undergo mammary carcinogenesis; however, mammary glands that had been exposed to mammary carcinogen did not show a difference in *Map3k1* levels between *Mc1b* alleles. An interesting result in our study with respect to *Map3k1*, which may have important implications for human studies of potential genotype-environment interactions, is exposure to mammary carcinogen resulted in increased mammary gland *Map3k1* levels for the *Mcs1b* resistant, but not the susceptible

genotype. We found no evidence of a rat orthologous transcript to human *C5ORF35* in multiple rat tissues. Further, exonic elements of *C5ORF35* have not been conserved in the rat. Therefore, we conclude that *MAP3K1* and *C5ORF35* are not as likely as *MIER3* to be breast cancer susceptibility genes.

I also identified multiple variants of *Mier3* expressed in the rat mammary gland. One of these variants exhibits alternate exon usage compared to the annotated *Mier3* transcript (GenBank ref|NM_152622). Similar to what was observed in the original *Mier3* gene expression studies, expression of the alternate exon variant is reduced in resistant Mcs1b mammary gland tissue compared to susceptible controls. The sequence of the variant differs only at the 5' end of the transcript, but is predicted to introduce a casein kinase phosphorylation site at the N-terminus. However, the implications are not clear and more study is necessary to determine whether differential expression of this variant affects mammary carcinoma susceptibility. Regardless, differential expression of this variant adds to our evidence that *Mier3* is the most likely breast cancer gene.

We noted that both rat *Mcs1b* and human *5q11.2* exhibit pleiotropy. Travis *et al.* reported that carriers of the increased risk allele at human *5q11.2* were significantly shorter in height than non-carriers (Travis et al., 2010). In our study, high risk female rats had lower body weight than *Mcs1b* resistant females. There is a predicted rat body weight QTL named *Bw1* that overlaps *Mcs1b* and is associated with mesenteric body fat amount (Ogino et al., 2000). Both human and rat study results are counter intuitive as one might expect taller women and heavier rats to be at greater cancer risk. Thus, it is important to note that, as expected with low-penetrance alleles, the quantitative difference between the means for each human genotype were subtle with overlapping

distributions. Mean height difference was 7 mm between non-carriers and carriers of the increased risk allele. In our study, we analyzed only body weight, and not specific components of body weight, such as bone density or fat. Thus, better descriptive traits would likely be more informative. It is notable that the pleitropic effects of these alleles opens the possibility that other experimental organisms, approaches, and study designs without focus on breast or mammary cancer may be useful to functionally characterize breast cancer risk associated genetic variation at *5q11.2*.

In conclusion, rat *Mcs1b* contains a mammary gland-autonomous allele and a non-protein-coding genetic element that is orthologous to the GWAS-identified human *5q11.2* breast cancer susceptibility locus. We propose that *MIER3* is a strong candidate breast cancer susceptibility gene.

REFERENCES

(2005). A haplotype map of the human genome. *Nature* **437**, 1299-320.

(2012). Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res* **40**, D71-5.

Aasland, R., Stewart, A. F. and Gibson, T. (1996). The SANT domain: a putative DNA-binding domain in the SWI-SNF and ADA complexes, the transcriptional co-repressor N-CoR and TFIIIB. *Trends Biochem Sci* **21**, 87-8.

Ahmed, S. Thomas, G. Ghoussaini, M. Healey, C. S. Humphreys, M. K. Platte, R. Morrison, J. Maranian, M. Pooley, K. A. Luben, R. et al. (2009). Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. *Nat Genet* **41**, 585-90.

Aitman, T. J., Critser, J. K., Cuppen, E., Dominiczak, A., Fernandez-Suarez, X. M., Flint, J., Gauguier, D., Geurts, A. M., Gould, M., Harris, P. C. et al. (2008). Progress and prospects in rat genetics: a community view. *Nat Genet* **40**, 516-22.

Altekruse SF, K. C., Krapcho M, Neyman N, Aminou R, Waldron W, Ruhl J, Howlader N, Tatalovich Z, Cho H, Mariotto A, Eisner MP, Lewis DR, Cronin K, Chen HS, Feuer EJ, Stinchcomb DG, Edwards BK (eds). (2010). SEER Cancer Statistics Review, 1975-2007. In *National Cancer Institute.*, (ed. Bethesda, MD.

Altshuler, D., Daly, M. J. and Lander, E. S. (2008). Genetic mapping in human disease. *Science (New York, N Y)* **322**, 881-8.

American Cancer Society. (2011). Breast Cancer Facts & Figures 2011-2012, (ed. Atlanta: American Cancer Society.

American Cancer Society. (2012). Cancer Facts and Figures 2012, (ed. Atlanta: American Cancer Society.

Anderson, G. L., Chlebowski, R. T., Aragaki, A. K., Kuller, L. H., Manson, J. E., Gass, M., Bluhm, E., Connelly, S., Hubbell, F. A., Lane, D. et al. (2012). Conjugated equine oestrogen and breast cancer incidence and mortality in postmenopausal women with hysterectomy: extended follow-up of the Women's Health Initiative randomised placebo-controlled trial. *Lancet Oncol* **13**, 476-86.

Anderson, W. F. and Devesa, S. S. (2005). Breast carcinoma in men. *Cancer* **103**, 432-3; author reply 433.

Antoniou, A., Pharoah, P. D., Narod, S., Risch, H. A., Eyfjord, J. E., Hopper, J. L., Loman, N., Olsson, H., Johannsson, O., Borg, A. et al. (2003). 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* **72**, 1117-30.

Antoniou, A. C., Spurdle, A. B., Sinilnikova, O. M., Healey, S., Pooley, K. A., Schmutzler, R. K., Versmold, B., Engel, C., Meindl, A., Arnold, N. et al. (2008). Common breast cancer-predisposition alleles are associated with breast cancer risk in BRCA1 and BRCA2 mutation carriers. *Am J Hum Genet* **82**, 937-48.

Antoniou, A. C. Wang, X. Fredericksen, Z. S. McGuffog, L. Tarrell, R. Sinilnikova, O. M. Healey, S. Morrison, J. Kartsonaki, C. Lesnick, T. et al. (2010). A locus on 19p13 modifies risk of breast cancer in BRCA1 mutation carriers and is associated with hormone receptor-negative breast cancer in the general population. *Nat Genet* **42**, 885-92.

Badano, J. L. and Katsanis, N. (2002). Beyond Mendel: an evolving view of human genetic disease transmission. *Nat Rev Genet* **3**, 779-89.

Bahia, H., Ashman, J. N., Cawkwell, L., Lind, M., Monson, J. R., Drew, P. J. and Greenman, J. (2002). Karyotypic variation between independently cultured strains of the cell line MCF-7 identified by multicolour fluorescence in situ hybridization. *Int J Oncol* **20**, 489-94.

Barrett, J. C. and Cardon, L. R. (2006). Evaluating coverage of genome-wide association studies. *Nat Genet* **38**, 659-62.

Barrett, J. C., Fry, B., Maller, J. and Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263-5.

Bartella, V., Rizza, P., Barone, I., Zito, D., Giordano, F., Giordano, C., Catalano, S., Mauro, L., Sisci, D., Panno, M. L. et al. (2012). Estrogen receptor beta binds Sp1 and recruits a corepressor complex to the estrogen receptor alpha gene promoter. *Breast Cancer Res Treat.*

Beatson, G. T. (1896). On the treatment of inoperable cases of carcinoma of the mamma : suggestions for a new method of treatment, with illustrative cases. Glasgow: Printed by Alex. Macdougall.

Beral, V. (2003). Breast cancer and hormone-replacement therapy in the Million Women Study. *Lancet* **362**, 419-27.

Bevier, M., Sundquist, K. and Hemminki, K. (2011). Risk of breast cancer in families of multiple affected women and men. *Breast Cancer Res Treat.*

Blackmore, T., Mercer, C., Paterno, G. and Gillespie, L. (2008). The transcriptional cofactor MIER1-beta negatively regulates histone acetyltransferase activity of the CREB-binding protein. *BMC Research Notes* **1**, 68.

Boice, J. D., Jr., Bigbee, W. L., Mumma, M. T. and Blot, W. J. (2003). Cancer incidence in municipalities near two former nuclear materials processing facilities in Pennsylvania. *Health Phys* **85**, 678-90.

Bonde, J. P., Hansen, J., Kolstad, H. A., Mikkelsen, S., Olsen, J. H., Blask, D. E., Harma, M., Kjuus, H., de Koning, H. J., Olsen, J. et al. (2012). Work at night and breast cancer - report on evidence-based options for preventive actions. *Scand J Work Environ Health.*

Boonyaratanakornkit, V., Scott, M. P., Ribon, V., Sherman, L., Anderson, S. M., Maller, J. L., Miller, W. T. and Edwards, D. P. (2001). Progesterone receptor contains a proline-rich motif that directly interacts with SH3 domains and activates c-Src family tyrosine kinases. *Mol Cell* **8**, 269-80.

Boyd, N. F., Connelly, P., Byng, J., Yaffe, M., Draper, H., Little, L., Jones, D., Martin, L. J., Lockwood, G. and Tritchler, D. (1995). Plasma lipids, lipoproteins, and mammographic densities. *Cancer Epidemiol Biomarkers Prev* **4**, 727-33.

Broeks, A. Schmidt, M. K. Sherman, M. E. Couch, F. J. Hopper, J. L. Dite, G. S. Apicella, C. Smith, L. D. Hammet, F. Southey, M. C. et al. (2011). Low penetrance breast cancer susceptibility loci are associated with specific breast tumor subtypes: findings from the Breast Cancer Association Consortium. *Human Molecular Genetics* **20**, 3289-303.

Bulun, S. E., Chen, D., Moy, I., Brooks, D. C. and Zhao, H. (2012). Aromatase, breast cancer and obesity: a complex interaction. *Trends Endocrinol Metab* **23**, 83-9.

Burdall, S. E., Hanby, A. M., Lansdown, M. R. and Speirs, V. (2003). Breast cancer cell lines: friend or foe? *Breast Cancer Res* **5**, 89-95.

Campa, D., Kaaks, R., Le Marchand, L., Haiman, C. A., Travis, R. C., Berg, C. D., Buring, J. E., Chanock, S. J., Diver, W. R., Dostal, L. et al. (2011). Interactions between genetic variants and breast cancer risk factors in the breast and prostate cancer cohort consortium. *Journal of the National Cancer Institute* **103**, 1252-63.

Campeau, P. M., Foulkes, W. D. and Tischkowitz, M. D. (2008). Hereditary breast cancer: new genetic developments, new therapeutic avenues. *Hum Genet* **124**, 31-42.

Cardiff, R. D. and Kenney, N. (2007). Mouse mammary tumor biology: a short history. *Adv Cancer Res* **98**, 53-116.

Carmichael, A. R. (2006). Obesity and prognosis of breast cancer. *Obes Rev* **7**, 333-40.

Casey, M. L., MacDonald, P. C. and Andersson, S. (1994). 17 beta-Hydroxysteroid dehydrogenase type 2: chromosomal assignment and progestin regulation of gene expression in human endometrium. *J Clin Invest* **94**, 2135-41.

Cavalieri, E. L., Rogan, E. G. and Chakravarti, D. (2002). Initiation of cancer and other diseases by catechol ortho-quinones: a unifying mechanism. *Cell Mol Life Sci* **59**, 665-81.

Chlebowski, R. T., Anderson, G. L., Gass, M., Lane, D. S., Aragaki, A. K., Kuller, L. H., Manson, J. E., Stefanick, M. L., Ockene, J., Sarto, G. E. et al. (2010). Estrogen plus progestin and breast cancer incidence and mortality in postmenopausal women. *JAMA : the journal of the American Medical Association* **304**, 1684-92.

Chlebowski, R. T., Hendrix, S. L., Langer, R. D., Stefanick, M. L., Gass, M., Lane, D., Rodabough, R. J., Gilligan, M. A., Cyr, M. G., Thomson, C. A. et al. (2003). Influence of estrogen plus progestin on breast cancer and mammography in healthy postmenopausal women: the Women's Health Initiative Randomized Trial. *JAMA : the journal of the American Medical Association* **289**, 3243-53.

Christou, M., Moore, C. J., Gould, M. N. and Jefcoate, C. R. (1987). Induction of mammary cytochromes P-450: an essential first step in the metabolism of 7,12-dimethylbenz[a]anthracene by rat mammary epithelial cells. *Carcinogenesis* **8**, 73-80.

Christou, M., Savas, U., Schroeder, S., Shen, X., Thompson, T., Gould, M. N. and Jefcoate, C. R. (1995). Cytochromes CYP1A1 and CYP1B1 in the rat mammary gland: cell-specific expression and regulation by polycyclic aromatic hydrocarbons and hormones. *Mol Cell Endocrinol* **115**, 41-50.

Collaborative Group on Hormonal Factors in Breast Cancer. (2001). Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. *Lancet* **358**, 1389-99.

Costantino, J. P., Gail, M. H., Pee, D., Anderson, S., Redmond, C. K., Benichou, J. and Wieand, H. S. (1999). Validation studies for models projecting the risk of invasive and total breast cancer incidence. *J Natl Cancer Inst* **91**, 1541-8.

Cotroneo, M. S., Merry, G. M., Haag, J. D., Lan, H., Shepel, L. A. and Gould, M. N. (2006). The Mcs7 quantitative trait locus is associated with an increased susceptibility to mammary cancer in congenic rats and an allele-specific imbalance. *Oncogene* **25**, 5011-7.

Coussens, L., Yang-Feng, T. L., Liao, Y. C., Chen, E., Gray, A., McGrath, J., Seeburg, P. H., Libermann, T. A., Schlessinger, J., Francke, U. et al. (1985). Tyrosine kinase receptor with extensive homology to EGF receptor shares chromosomal location with neu oncogene. *Science (New York, N Y)* **230**, 1132-9.

Cox, A., Dunning, A. M., Garcia-Closas, M., Balasubramanian, S., Reed, M. W., Pooley, K. A., Scollen, S., Baynes, C., Ponder, B. A., Chanock, S. et al. (2007). A common coding variant in CASP8 is associated with breast cancer risk. *Nat Genet* **39**, 352-8.

Cui, J., Germer, K., Wu, T., Wang, J., Luo, J., Wang, S. C., Wang, Q. and Zhang, X. (2012). Cross-talk between HER2 and MED1 regulates tamoxifen resistance of human breast cancer cells. *Cancer Res* **72**, 5625-34.

Czerwinski, S. A., Lee, M., Choh, A. C., Wurzbacher, K., Demerath, E. W., Towne, B. and Siervogel, R. M. (2007). Genetic factors in physical growth and development and

their relationship to subsequent health outcomes. *American journal of human biology : the official journal of the Human Biology Council* **19**, 684-91.

Darbre, P. D. and Charles, A. K. (2010). Environmental oestrogens and breast cancer: evidence for combined involvement of dietary, household and cosmetic xenoestrogens. *Anticancer Res* **30**, 815-27.

Davidson, G., Wu, W., Shen, J., Bilic, J., Fenger, U., Stannek, P., Glinka, A. and Niehrs, C. (2005). Casein kinase 1 gamma couples Wnt receptor activation to cytoplasmic signal transduction. *Nature* **438**, 867-72.

de Jong, M. M., Nolte, I. M., te Meerman, G. J., van der Graaf, W. T., Oosterwijk, J. C., Kleibeuker, J. H., Schaapveld, M. and de Vries, E. G. (2002). Genes other than BRCA1 and BRCA2 involved in breast cancer susceptibility. *J Med Genet* **39**, 225-42.

de Lorimier, A. A. (2000). Alcohol, wine, and health. *Am J Surg* **180**, 357-61.

De Stefani, E., Ronco, A., Mendilaharsu, M., Guidobono, M. and Deneo-Pellegrini, H. (1997). Meat intake, heterocyclic amines, and risk of breast cancer: a case-control study in Uruguay. *Cancer Epidemiol Biomarkers Prev* **6**, 573-81.

Deitz, A. C., Zheng, W., Leff, M. A., Gross, M., Wen, W. Q., Doll, M. A., Xiao, G. H., Folsom, A. R. and Hein, D. W. (2000). N-Acetyltransferase-2 genetic polymorphism, well-done meat intake, and breast cancer risk among postmenopausal women. *Cancer Epidemiol Biomarkers Prev* **9**, 905-10.

den Dekker, A. D., Xu, X., Vaughn, M. D., Puckett, A. H., Gardner, L. L., Lambring, C. J., Deschenes, L. and Samuelson, D. J. (2012). Rat *Mcs1b* is concordant to the human GWAS-identified breast cancer risk locus at *5q11.2* and *Mier3* is a candidate susceptibility gene *Submitted.*

Ding, Z., Gillespie, L. L., Mercer, F. C. and Paterno, G. D. (2004a). The SANT domain of human MI-ER1 interacts with Sp1 to interfere with GC box recognition and repress transcription from its own promoter. *J Biol Chem* **279**, 28009-16.

Ding, Z., Gillespie, L. L., Mercer, F. C. and Paterno, G. D. (2004b). The SANT Domain of Human MI-ER1 Interacts with Sp1 to Interfere with GC Box Recognition and Repress Transcription from Its Own Promoter. *Journal of Biological Chemistry* **279**, 28009-28016.

Ding, Z., Gillespie, L. L. and Paterno, G. D. (2003). Human MI-ER1 alpha and beta function as transcriptional repressors by recruitment of histone deacetylase 1 to their conserved ELM2 domain. *Mol Cell Biol* **23**, 250-8.

Dinkel, H., Michael, S., Weatheritt, R. J., Davey, N. E., Van Roey, K., Altenberg, B., Toedt, G., Uyar, B., Seiler, M., Budd, A. et al. (2012). ELM--the database of eukaryotic linear motifs. *Nucleic Acids Res* **40**, D242-51.

Dobrzycka, K. M., Townson, S. M., Jiang, S. and Oesterreich, S. (2003). Estrogen receptor corepressors -- a role in human breast cancer? *Endocr Relat Cancer* **10**, 517-36.

Dunning, W. F. and Curtis, M. R. (1952). The incidence of diethylstilbestrol-induced cancer in reciprocal F hybrids obtained from crosses between rats of inbred lines that are susceptible and resistant to the induction of mammary cancer by this agent. *Cancer Res* **12**, 702-6.

Dunning, W. F., Curtis, M. R. and Segaloff, A. (1947). Strain Differences in Response to Diethylstilbestrol and the Induction of Mammary Gland and Bladder Cancer in the Rat. *Cancer Res* **7**, 511-521.

Dunning, W. F., Curtis, M. R. and Segaloff, A. (1953). Strain differences in response to estrone and the induction of mammary gland, adrenal, and bladder cancer in rats. *Cancer Res* **13**, 147-52.

Dutertre, M. and Smith, C. L. (2000). Molecular mechanisms of selective estrogen receptor modulator (SERM) action. *The Journal of pharmacology and experimental therapeutics* **295**, 431-7.

Dwinell, M. R., Worthey, E. A., Shimoyama, M., Bakir-Gungor, B., DePons, J., Laulederkind, S., Lowry, T., Nigram, R., Petri, V., Smith, J. et al. (2009). The Rat Genome Database 2009: variation, ontologies and pathways. *Nucleic Acids Res* **37**, D744-9.

Easton, D. F. (1999). How many more breast cancer predisposition genes are there? *Breast Cancer Res* **1**, 14-7.

Easton, D. F. Pooley, K. A. Dunning, A. M. Pharoah, P. D. Thompson, D. Ballinger, D. G. Struewing, J. P. Morrison, J. Field, H. Luben, R. et al. (2007). Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**, 1087-93.

Erkko, H., Xia, B., Nikkila, J., Schleutker, J., Syrjakoski, K., Mannermaa, A., Kallioniemi, A., Pylkas, K., Karppinen, S. M., Rapakko, K. et al. (2007). A recurrent mutation in PALB2 in Finnish cancer families. *Nature* **446**, 316-9.

Euhus, D. M., Leitch, A. M., Huth, J. F. and Peters, G. N. (2002). Limitations of the Gail model in the specialized breast cancer risk assessment clinic. *Breast J* **8**, 23-7.

Evans, D. G. and Howell, A. (2007). Breast cancer risk-assessment models. *Breast Cancer Res* **9**, 213.

Falany, J. L. and Falany, C. N. (1996). Regulation of estrogen sulfotransferase in human endometrial adenocarcinoma cells by progesterone. *Endocrinology* **137**, 1395-401.

Fernandez, S. V. and Russo, J. (2010). Estrogen and xenoestrogens in breast cancer. *Toxicol Pathol* **38**, 110-22.

Fisher, B., Costantino, J. P., Redmond, C. K., Fisher, E. R., Wickerham, D. L. and Cronin, W. M. (1994). Endometrial cancer in tamoxifen-treated breast cancer patients: findings from the National Surgical Adjuvant Breast and Bowel Project (NSABP) B-14. *J Natl Cancer Inst* **86**, 527-37.

Flemming, J., Madarnas, Y. and Franek, J. A. (2009). Fulvestrant for systemic therapy of locally advanced or metastatic breast cancer in postmenopausal women: a systematic review. *Breast Cancer Res Treat* **115**, 255-68.

Fletcher, O. and Houlston, R. S. (2010). Architecture of inherited susceptibility to common cancer. *Nat Rev Cancer* **10**, 353-61.

Flicek, P., Amode, M. R., Barrell, D., Beal, K., Brent, S., Chen, Y., Clapham, P., Coates, G., Fairley, S., Fitzgerald, S. et al. (2011). Ensembl 2011. *Nucleic Acids Research* **39**, D800-D806.

Ford, D., Easton, D. F. and Peto, J. (1995). Estimates of the gene frequency of BRCA1 and its contribution to breast and ovarian cancer incidence. *Am J Hum Genet* **57**, 1457-62.

Ford, D., Easton, D. F., Stratton, M., Narod, S., Goldgar, D., Devilee, P., Bishop, D. T., Weber, B., Lenoir, G., Chang-Claude, J. et al. (1998). Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium. *Am J Hum Genet* **62**, 676-89.

Ford, J., Hajibeigi, A., Long, M., Hahner, L., Gore, C., Hsieh, J. T., Clegg, D., Zerwekh, J. and Oz, O. K. (2011). GPR30 deficiency causes increased bone mass, mineralization, and growth plate proliferative activity in male mice. *J Bone Miner Res* **26**, 298-307.

Foulkes, W. D., Ghadirian, P., Akbari, M. R., Hamel, N., Giroux, S., Sabbaghian, N., Darnel, A., Royer, R., Poll, A., Fafard, E. et al. (2007). Identification of a novel truncating PALB2 mutation and analysis of its contribution to early-onset breast cancer in French-Canadian women. *Breast Cancer Res* **9**, R83.

Frazer, K. A., Murray, S. S., Schork, N. J. and Topol, E. J. (2009). Human genetic variation and its contribution to complex traits. *Nat Rev Genet* **10**, 241-51.

Gail, M. H., Brinton, L. A., Byar, D. P., Corle, D. K., Green, S. B., Schairer, C. and Mulvihill, J. J. (1989). Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst* **81**, 1879-86.

Gao, J., Lauer, F. T., Dunaway, S. and Burchiel, S. W. (2005). Cytochrome P450 1B1 is required for 7,12-dimethylbenz(a)-anthracene (DMBA) induced spleen cell immunotoxicity. *Toxicol Sci* **86**, 68-74.

Garcia-Closas, M. Hall, P. Nevanlinna, H. Pooley, K. Morrison, J. Richesson, D. A. Bojesen, S. E. Nordestgaard, B. G. Axelsson, C. K. Arias, J. I. et al. (2008a). Heterogeneity of Breast Cancer Associations with Five Susceptibility Loci by Clinical and Pathological Characteristics. *PLoS Genetics* **4**, e1000054.

Garcia-Closas, M. Hall, P. Nevanlinna, H. Pooley, K. Morrison, J. Richesson, D. A. Bojesen, S. E. Nordestgaard, B. G. Axelsson, C. K. Arias, J. I. et al. (2008b).

Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. *PLoS Genet* **4**, e1000054.

Geurts, A. M., Cost, G. J., Freyvert, Y., Zeitler, B., Miller, J. C., Choi, V. M., Jenkins, S. S., Wood, A., Cui, X., Meng, X. et al. (2009). Knockout rats via embryo microinjection of zinc-finger nucleases. *Science (New York, N Y)* **325**, 433.

Gossen, M. and Bujard, H. (1992). Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 5547-51.

Gottardis, M. M., Robinson, S. P., Satyaswaroop, P. G. and Jordan, V. C. (1988). Contrasting actions of tamoxifen on endometrial and breast tumor growth in the athymic mouse. *Cancer Res* **48**, 812-5.

Gould, K. A., Tochacek, M., Schaffer, B. S., Reindl, T. M., Murrin, C. R., Lachel, C. M., VanderWoude, E. A., Pennington, K. L., Flood, L. A., Bynote, K. K. et al. (2004). Genetic determination of susceptibility to estrogen-induced mammary cancer in the ACI rat: mapping of Emca1 and Emca2 to chromosomes 5 and 18. *Genetics* **168**, 2113-25.

Gould, M. N. (1986). Inheritance and site of expression of genes controlling susceptibility to mammary cancer in an inbred rat model. *Cancer Res* **46**, 1199-202.

Gould, M. N. (1995). Rodent models for the study of etiology, prevention and treatment of breast cancer. *Semin Cancer Biol* **6**, 147-52.

Gould, M. N. and Clifton, K. H. (1985). Clonogen transplantation assay of mammary and thyroid epithelial cells. In *Cell Clones: Manual of Mammalian Techniques*, (ed. C. S. Potten and J. H. Hendry), pp. 128-137. Edinburgh, London, Melbourne and New York: Churchill Livingstone.

Grune, T., Brzeski, J., Eberharter, A., Clapier, C. R., Corona, D. F., Becker, P. B. and Muller, C. W. (2003). Crystal structure and functional analysis of a nucleosome recognition module of the remodeling factor ISWI. *Mol Cell* **12**, 449-60.

Guy, C. T., Cardiff, R. D. and Muller, W. J. (1992a). Induction of mammary tumors by expression of polyomavirus middle T oncogene: a transgenic mouse model for metastatic disease. *Mol Cell Biol* **12**, 954-61.

Guy, C. T., Webster, M. A., Schaller, M., Parsons, T. J., Cardiff, R. D. and Muller, W. J. (1992b). Expression of the neu protooncogene in the mammary epithelium of transgenic mice induces metastatic disease. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 10578-82.

Haag, J. D., Hsu, L. C., Newton, M. A. and Gould, M. N. (1996). Allelic imbalance in mammary carcinomas induced by either 7,12-dimethylbenz[a]anthracene or ionizing radiation in rats carrying genes conferring differential susceptibilities to mammary carcinogenesis. *Mol Carcinog* **17**, 134-43.

Haag, J. D., Newton, M. A. and Gould, M. N. (1992). Mammary carcinoma suppressor and susceptibility genes in the Wistar-Kyoto rat. *Carcinogenesis* **13**, 1933-5.

Haag, J. D., Shepel, L. A., Kolman, B. D., Monson, D. M., Benton, M. E., Watts, K. T., Waller, J. L., Lopez-Guajardo, C. C., Samuelson, D. J. and Gould, M. N. (2003). Congenic rats reveal three independent Copenhagen alleles within the Mcs1 quantitative trait locus that confer resistance to mammary cancer. *Cancer Res* **63**, 5808-12.

Hall, J. M. and McDonnell, D. P. (1999). The estrogen receptor beta-isoform (ERbeta) of the human estrogen receptor modulates ERalpha transcriptional activity and is a key regulator of the cellular response to estrogens and antiestrogens. *Endocrinology* **140**, 5566-78.

Han, W., Woo, J. H., Yu, J. H., Lee, M. J., Moon, H. G., Kang, D. and Noh, D. Y. (2011). Common genetic variants associated with breast cancer in Korean women and differential susceptibility according to intrinsic subtype. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **20**, 793-8.

Hankinson, S. E., Willett, W. C., Manson, J. E., Hunter, D. J., Colditz, G. A., Stampfer, M. J., Longcope, C. and Speizer, F. E. (1995). Alcohol, height, and adiposity in relation to estrogen and prolactin levels in postmenopausal women. *J Natl Cancer Inst* **87**, 1297-302.

Hardell, L. and Sage, C. (2008). Biological effects from electromagnetic field exposure and public exposure standards. *Biomed Pharmacother* **62**, 104-9.

Heery, D. M., Kalkhoven, E., Hoare, S. and Parker, M. G. (1997). A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature* **387**, 733-6.

Hill, S. M., Blask, D. E., Xiang, S., Yuan, L., Mao, L., Dauchy, R. T., Dauchy, E. M., Frasch, T. and Duplesis, T. (2011). Melatonin and associated signaling pathways that control normal breast epithelium and breast cancer. *J Mammary Gland Biol Neoplasia* **16**, 235-45.

Hindorff, L., Institute), J. M. E. B., Wise, A., Junkins, H., Hall, P., Klemm, R. and Manolio, T. (2009). A Catalog of Published Genome-Wide Association Studies, (ed.

Hoover, R., Gray, L. A., Sr., Cole, P. and MacMahon, B. (1976). Menopausal estrogens and breast cancer. *The New England journal of medicine* **295**, 401-5.

Horwitz, K. B., Mockus, M. B. and Lessey, B. A. (1982). Variant T47D human breast cancer cells with high progesterone-receptor levels despite estrogen and antiestrogen resistance. *Cell* **28**, 633-42.

Hsu, L. C., Kennan, W. S., Shepel, L. A., Jacob, H. J., Szpirer, C., Szpirer, J., Lander, E. S. and Gould, M. N. (1994). Genetic identification of Mcs-1, a rat mammary carcinoma suppressor gene. *Cancer Res* **54**, 2765-70.

Hsueh, A. J., Peck, E. J., Jr. and Clark, J. H. (1976). Control of uterine estrogen receptor levels by progesterone. *Endocrinology* **98**, 438-44.

Hunter, D. J. (2005). Gene-environment interactions in human diseases. *Nat Rev Genet* **6**, 287-98.

Hutchinson, J. N. and Muller, W. J. (2000). Transgenic mouse models of human breast cancer. *Oncogene* **19**, 6130-7.

Isaacs, J. T. (1986). Genetic control of resistance to chemically induced mammary adenocarcinogenesis in the rat. *Cancer Res* **46**, 3958-63.

Isaacs, J. T. (1988). Inheritance of a genetic factor from the Copenhagen rat and the suppression of chemically induced mammary adenocarcinogenesis. *Cancer Res* **48**, 2204-13.

Jostins, L. and Barrett, J. C. (2011). Genetic risk prediction in complex disease. *Hum Mol Genet* **20**, R182-8.

Justenhoven, C., Hamann, U., Pierl, C. B., Baisch, C., Harth, V., Rabstein, S., Spickenheuer, A., Pesch, B., Bruning, T., Winter, S. et al. (2009). CYP2C19*17 is associated with decreased breast cancer risk. *Breast Cancer Res Treat* **115**, 391-6.

Kelsey, J. L., Gammon, M. D. and John, E. M. (1993). Reproductive factors and breast cancer. *Epidemiol Rev* **15**, 36-47.

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M. and Haussler, D. (2002). The human genome browser at UCSC. *Genome Res* **12**, 996-1006.

Kester, M. H., Bulduk, S., Tibboel, D., Meinl, W., Glatt, H., Falany, C. N., Coughtrie, M. W., Bergman, A., Safe, S. H., Kuiper, G. G. et al. (2000). Potent inhibition of estrogen sulfotransferase by hydroxylated PCB metabolites: a novel pathway explaining the estrogenic activity of PCBs. *Endocrinology* **141**, 1897-900.

Kleer, C. G., Cao, Q., Varambally, S., Shen, R., Ota, I., Tomlins, S. A., Ghosh, D., Sewalt, R. G., Otte, A. P., Hayes, D. F. et al. (2003). EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 11606-11.

Kleinberg, D. L., Ruan, W., Catanese, V., Newman, C. B. and Feldman, M. (1990). Non-lactogenic effects of growth hormone on growth and insulin-like growth factor-I messenger ribonucleic acid of rat mammary gland. *Endocrinology* **126**, 3274-6.

Klinge, C. M. (2000). Estrogen receptor interaction with co-activators and co-repressors. *Steroids* **65**, 227-51.

Kronenberg, H. and Williams, R. H. (2008). Williams textbook of endocrinology. Philadelphia, PA: Saunders/Elsevier.

Kuiper, G. G., Carlsson, B., Grandien, K., Enmark, E., Haggblad, J., Nilsson, S. and Gustafsson, J. A. (1997). Comparison of the ligand binding specificity and transcript tissue distribution of estrogen receptors alpha and beta. *Endocrinology* **138**, 863-70.

Kurz, S. G., Hansen, K. K., McLaughlin, M. T., Shivaswamy, V., Schaffer, B. S., Gould, K. A., McComb, R. D., Meza, J. L. and Shull, J. D. (2008). Tissue-specific actions of the Ept1, Ept2, Ept6, and Ept9 genetic determinants of responsiveness to estrogens in the female rat. *Endocrinology* **149**, 3850-9.

Lalloo, F. and Evans, D. G. (2012). Familial Breast Cancer. *Clin Genet.*

Lan, H., Kendziorski, C. M., Haag, J. D., Shepel, L. A., Newton, M. A. and Gould, M. N. (2001). Genetic loci controlling breast cancer susceptibility in the Wistar-Kyoto rat. *Genetics* **157**, 331-9.

Land, C. E., Tokunaga, M., Koyama, K., Soda, M., Preston, D. L., Nishimori, I. and Tokuoka, S. (2003). Incidence of female breast cancer among atomic bomb survivors, Hiroshima and Nagasaki, 1950-1990. *Radiat Res* **160**, 707-17.

Lange, C. A. (2008). Challenges to defining a role for progesterone in breast cancer. *Steroids* **73**, 914-21.

Lanigan, F., O'Connor, D., Martin, F. and Gallagher, W. M. (2007). Molecular links between mammary gland development and breast cancer. *Cell Mol Life Sci* **64**, 3159-84.

Lawther, P. J. and Waller, R. E. (1976). Coal fires, industrial emissions and motor vehicles as sources of environmental carcinogens. *IARC Sci Publ*, 27-40.

Lee, S. H., Wray, N. R., Goddard, M. E. and Visscher, P. M. (2011). Estimating missing heritability for disease from genome-wide association studies. *Am J Hum Genet* **88**, 294-305.

Lemoine, N. R., Staddon, S., Dickson, C., Barnes, D. M. and Gullick, W. J. (1990). Absence of activating transmembrane mutations in the c-erbB-2 proto-oncogene in human breast cancer. *Oncogene* **5**, 237-9.

Leygue, E., Dotzlaw, H., Watson, P. H. and Murphy, L. C. (1998). Altered estrogen receptor alpha and beta messenger RNA expression during human breast tumorigenesis. *Cancer Res* **58**, 3197-201.

Li, M., Chen, D., Shiloh, A., Luo, J., Nikolaev, A. Y., Qin, J. and Gu, W. (2002). Deubiquitination of p53 by HAUSP is an important pathway for p53 stabilization. *Nature* **416**, 648-53.

Lichtenstein, P., Holm, N. V., Verkasalo, P. K., Iliadou, A., Kaprio, J., Koskenvuo, M., Pukkala, E., Skytthe, A. and Hemminki, K. (2000). Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. *The New England journal of medicine* **343**, 78-85.

Liehr, J. G. (1990). Genotoxic effects of estrogens. *Mutat Res* **238**, 269-76.

Liehr, J. G. and Roy, D. (1990). Free radical generation by redox cycling of estrogens. *Free Radic Biol Med* **8**, 415-23.

Lydon, J. P., DeMayo, F. J., Funk, C. R., Mani, S. K., Hughes, A. R., Montgomery, C. A., Jr., Shyamala, G., Conneely, O. M. and O'Malley, B. W. (1995). Mice lacking progesterone receptor exhibit pleiotropic reproductive abnormalities. *Genes Dev* **9**, 2266-78.

MacMahon, B., Cole, P., Lin, T. M., Lowe, C. R., Mirra, A. P., Ravnihar, B., Salber, E. J., Valaoras, V. G. and Yuasa, S. (1970). Age at first birth and breast cancer risk. *Bull World Health Organ* **43**, 209-21.

Maddison, K. and Clarke, A. R. (2005). New approaches for modelling cancer mechanisms in the mouse. *J Pathol* **205**, 181-93.

Maguire, P., Margolin, S., Skoglund, J., Sun, X. F., Gustafsson, J. A., Borresen-Dale, A. L. and Lindblom, A. (2005). Estrogen receptor beta (ESR2) polymorphisms in familial and sporadic breast cancer. *Breast Cancer Res Treat* **94**, 145-52.

Marcotte, R. and Muller, W. J. (2008). Signal transduction in transgenic mouse models of human breast cancer--implications for human breast cancer. *J Mammary Gland Biol Neoplasia* **13**, 323-35.

Marian, A. J. (2012). Elements of 'missing heritability'. *Curr Opin Cardiol* **27**, 197-201.

Markel, P., Shu, P., Ebeling, C., Carlson, G. A., Nagle, D. L., Smutko, J. S. and Moore, K. J. (1997). Theoretical and empirical issues for marker-assisted breeding of congenic mouse strains. *Nat Genet* **17**, 280-4.

Martin, P. M., Horwitz, K. B., Ryan, D. S. and McGuire, W. L. (1978). Phytoestrogen interaction with estrogen receptors in human breast cancer cells. *Endocrinology* **103**, 1860-7.

Masson, L. F., Sharp, L., Cotton, S. C. and Little, J. (2005). Cytochrome P-450 1A1 gene polymorphisms and risk of breast cancer: a HuGE review. *Am J Epidemiol* **161**, 901-15.

Mathers, C., Fat, D. M., Boerma, J. T., World Health Organization. and ebrary Inc. (2008). The global burden of disease 2004 update, (ed. Geneva, Switzerland: World Health Organization.

McCarthy, P. L., Mercer, F. C., Savicky, M. W., Carter, B. A., Paterno, G. D. and Gillespie, L. L. (2008a). Changes in subcellular localisation of MI-ER1 alpha, a novel oestrogen receptor-alpha interacting protein, is associated with breast cancer progression. *Br J Cancer* **99**, 639-46.

McCarthy, P. L., Mercer, F. C., Savicky, M. W. J., Carter, B. A., Paterno, G. D. and Gillespie, L. L. (2008b). Changes in subcellular localisation of MI-ER1[alpha], a novel oestrogen receptor-[alpha] interacting protein, is associated with breast cancer progression. *Br J Cancer* **99**, 639-646.

McInerney, E. M., Rose, D. W., Flynn, S. E., Westin, S., Mullen, T. M., Krones, A., Inostroza, J., Torchia, J., Nolte, R. T., Assa-Munt, N. et al. (1998a). Determinants of coactivator LXXLL motif specificity in nuclear receptor transcriptional activation. *Genes & Development* **12**, 3357-68.

McInerney, E. M., Rose, D. W., Flynn, S. E., Westin, S., Mullen, T. M., Krones, A., Inostroza, J., Torchia, J., Nolte, R. T., Assa-Munt, N. et al. (1998b). Determinants of coactivator LXXLL motif specificity in nuclear receptor transcriptional activation. *Genes Dev* **12**, 3357-68.

Meijers-Heijboer, H., van den Ouweland, A., Klijn, J., Wasielewski, M., de Snoo, A., Oldenburg, R., Hollestelle, A., Houben, M., Crepin, E., van Veghel-Plandsoen, M. et al. (2002). Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. *Nat Genet* **31**, 55-9.

Mersch-Sundermann, V., Mochayedi, S., Kevekordes, S., Kern, S. and Wintermann, F. (1993). The genotoxicity of unsubstituted and nitrated polycyclic aromatic hydrocarbons. *Anticancer Res* **13**, 2037-43.

Migliaccio, A., Piccolo, D., Castoria, G., Di Domenico, M., Bilancio, A., Lombardi, M., Gong, W., Beato, M. and Auricchio, F. (1998). Activation of the Src/p21ras/Erk pathway by progesterone receptor via cross-talk with estrogen receptor. *EMBO J* **17**, 2008-18.

Milne, R. L., Benitez, J., Nevanlinna, H., Heikkinen, T., Aittomaki, K., Blomqvist, C., Arias, J. I., Zamora, M. P., Burwinkel, B., Bartram, C. R. et al. (2009). Risk of estrogen receptor-positive and -negative breast cancer and single-nucleotide polymorphism 2q35-rs13387042. *J Natl Cancer Inst* **101**, 1012-8.

Miyata, M., Furukawa, M., Takahashi, K., Gonzalez, F. J. and Yamazoe, Y. (2001). Mechanism of 7,12-dimethylbenz[a]anthracene-induced immunotoxicity: role of metabolic activation at the target organ. *Jpn J Pharmacol* **86**, 302-9.

Miyata, M., Kudo, G., Lee, Y. H., Yang, T. J., Gelboin, H. V., Fernandez-Salguero, P., Kimura, S. and Gonzalez, F. J. (1999). Targeted disruption of the microsomal epoxide hydrolase gene. Microsomal epoxide hydrolase is required for the carcinogenic activity of 7,12-dimethylbenz[a]anthracene. *J Biol Chem* **274**, 23963-8.

Modi, B. G., Neustadter, J., Binda, E., Lewis, J., Filler, R. B., Roberts, S. J., Kwong, B. Y., Reddy, S., Overton, J. D., Galan, A. et al. (2012). Langerhans cells facilitate epithelial DNA damage and squamous cell carcinoma. *Science (New York, N Y)* **335**, 104-8.

Mokbel, K. (2002). The evolving role of aromatase inhibitors in breast cancer. *Int J Clin Oncol* **7**, 279-83.

Moore, C. J., Bachhuber, A. J. and Gould, M. N. (1983). Relationship of mammary tumor susceptibility, mammary cell-mediated mutagenesis, and metabolism of polycyclic aromatic hydrocarbons in four types of rats. *J Natl Cancer Inst* **70**, 777-84.

Narod, S. A. and Salmena, L. (2011). BRCA1 and BRCA2 mutations and breast cancer. *Discov Med* **12**, 445-53.

Nathanson, K. L. and Weber, B. L. (2001). "Other" breast cancer susceptibility genes: searching for more holy grail. *Hum Mol Genet* **10**, 715-20.

Neuhausen, S., Gilewski, T., Norton, L., Tran, T., McGuire, P., Swensen, J., Hampel, H., Borgen, P., Brown, K., Skolnick, M. et al. (1996). Recurrent BRCA2 6174delT mutations in Ashkenazi Jewish women affected by breast cancer. *Nat Genet* **13**, 126-8.

Nevanlinna, H. and Bartek, J. (2006). The CHEK2 gene and inherited breast cancer susceptibility. *Oncogene* **25**, 5912-9.

Nijman, I. J., Kuipers, S., Verheul, M., Guryev, V. and Cuppen, E. (2008). A genome-wide SNP panel for mapping and association studies in the rat. *BMC Genomics* **9**, 95.

O'Malley, B. W. and Kumar, R. (2009). Nuclear receptor coregulators in cancer biology. *Cancer Res* **69**, 8217-22.

Oddoux, C., Struewing, J. P., Clayton, C. M., Neuhausen, S., Brody, L. C., Kaback, M., Haas, B., Norton, L., Borgen, P., Jhanwar, S. et al. (1996). The carrier frequency of the BRCA2 6174delT mutation among Ashkenazi Jewish individuals is approximately 1%. *Nat Genet* **14**, 188-90.

Ogino, T., Wei, S., Wei, K., Moralejo, D. H., Kose, H., Mizuno, A., Shima, K., Sasaki, Y., Yamada, T. and Matsumoto, K. (2000). Genetic evidence for obesity loci involved in the regulation of body fat distribution in obese type 2 diabetes rat, OLETF. *Genomics* **70**, 19-25.

Onami, S., Ozaki, M., Mortimer, J. E. and Pal, S. K. (2010). Male breast cancer: an update in diagnosis, treatment and molecular profiling. *Maturitas* **65**, 308-14.

Osborne, C. K., Hobbs, K. and Trent, J. M. (1987). Biological differences among MCF-7 human breast cancer cell lines from different laboratories. *Breast Cancer Res Treat* **9**, 111-21.

Osborne, C. K., Shou, J., Massarweh, S. and Schiff, R. (2005). Crosstalk between estrogen receptor and growth factor receptor pathways as a cause for endocrine therapy resistance in breast cancer. *Clin Cancer Res* **11**, 865s-70s.

Paech, K., Webb, P., Kuiper, G. G., Nilsson, S., Gustafsson, J., Kushner, P. J. and Scanlan, T. S. (1997). Differential ligand activation of estrogen receptors ERalpha and ERbeta at AP1 sites. *Science (New York, N Y)* **277**, 1508-10.

Pan, C., Kumar, C., Bohl, S., Klingmueller, U. and Mann, M. (2009). Comparative proteomic phenotyping of cell lines and primary cells to assess preservation of cell type-specific functions. *Mol Cell Proteomics* **8**, 443-50.

Paterno, G. D., Li, Y., Luchman, H. A., Ryan, P. J. and Gillespie, L. L. (1997). cDNA cloning of a novel, developmentally regulated immediate early gene activated by fibroblast growth factor and encoding a nuclear protein. *J Biol Chem* **272**, 25591-5.

Paterno, G. D., Mercer, F. C., Chayter, J. J., Yang, X., Robb, J. D. and Gillespie, L. L. (1998). Molecular cloning of human er1 cDNA and its differential expression in breast tumours and tumour-derived cell lines. *Gene* **222**, 77-82.

Pelucchi, C., Tramacere, I., Boffetta, P., Negri, E. and La Vecchia, C. (2011). Alcohol consumption and cancer risk. *Nutr Cancer* **63**, 983-90.

Peto, R. (2000). Cancer, genes, and the environment. *The New England journal of medicine* **343**, 1495; discussion 1495-6.

Pettersson, K., Delaunay, F. and Gustafsson, J. A. (2000). Estrogen receptor beta acts as a dominant regulator of estrogen signaling. *Oncogene* **19**, 4970-8.

Pettersson, K., Grandien, K., Kuiper, G. G. and Gustafsson, J. A. (1997). Mouse estrogen receptor beta forms estrogen response element-binding heterodimers with estrogen receptor alpha. *Mol Endocrinol* **11**, 1486-96.

Pharoah, P. D., Antoniou, A., Bobrow, M., Zimmern, R. L., Easton, D. F. and Ponder, B. A. (2002). Polygenic susceptibility to breast cancer and implications for prevention. *Nat Genet* **31**, 33-6.

Pharoah, P. D., Antoniou, A. C., Easton, D. F. and Ponder, B. A. (2008). Polygenes, risk prediction, and targeted prevention of breast cancer. *The New England journal of medicine* **358**, 2796-803.

Phillips, D. H. (1999). Polycyclic aromatic hydrocarbons in the diet. *Mutat Res* **443**, 139-47.

Pijpe, A., Andrieu, N., Easton, D. F., Kesminiene, A., Cardis, E., Nogues, C., Gauthier-Villars, M., Lasset, C., Fricker, J. P., Peock, S. et al. (2012). Exposure to diagnostic radiation and risk of breast cancer among carriers of BRCA1/2 mutations: retrospective cohort study (GENE-RAD-RISK). *BMJ (Clinical research ed)* **345**, e5660.

Pike, M. C., Krailo, M. D., Henderson, B. E., Casagrande, J. T. and Hoel, D. G. (1983). 'Hormonal' risk factors, 'breast tissue age' and the age-incidence of breast cancer. *Nature* **303**, 767-70.

Pike, M. C., Spicer, D. V., Dahmoush, L. and Press, M. F. (1993). Estrogens, progestogens, normal breast cell proliferation, and breast cancer risk. *Epidemiol Rev* **15**, 17-35.

Pollard, J. W. (2004). Tumour-educated macrophages promote tumour progression and metastasis. *Nat Rev Cancer* **4**, 71-8.

Prentice, R. L., Huang, Y., Hinds, D. A., Peters, U., Pettinger, M., Cox, D. R., Beilharz, E., Chlebowski, R. T., Rossouw, J. E., Caan, B. et al. (2009). Variation in

the FGFR2 gene and the effects of postmenopausal hormone therapy on invasive breast cancer. *Cancer Epidemiol Biomarkers Prev* **18**, 3079-85.

Preston-Martin, S., Pike, M. C., Ross, R. K. and Henderson, B. E. (1993). Epidemiologic evidence for the increased cell proliferation model of carcinogenesis. *Environ Health Perspect* **101 Suppl 5**, 137-8.

Pylkas, K., Vuorela, M., Otsukka, M., Kallioniemi, A., Jukkola-Vuorinen, A. and Winqvist, R. (2012). Rare copy number variants observed in hereditary breast cancer cases disrupt genes in estrogen signaling and TP53 tumor suppression network. *PLoS Genet* **8**, e1002734.

Qiu, C., Yu, M., Shan, L. and Snyderwine, E. G. (2003a). Allelic imbalance and altered expression of genes in chromosome 2q11-2q16 from rat mammary gland carcinomas induced by 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine. *Oncogene* **22**, 1253-60.

Qiu, C., Yu, M., Shan, L. and Snyderwine, E. G. (2003b). Allelic imbalance and altered expression of genes in chromosome 2q11-2q16 from rat mammary gland carcinomas induced by 2-amino-1-methyl-6-phenylimidazo[lsqb]4,5-b[rsqb]pyridine. *Oncogene* **22**, 1253-1260.

Rahman, N., Seal, S., Thompson, D., Kelly, P., Renwick, A., Elliott, A., Reid, S., Spanova, K., Barfoot, R., Chagtai, T. et al. (2007). PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nat Genet* **39**, 165-7.

Rhodes, D. R., Yu, J., Shanker, K., Deshpande, N., Varambally, R., Ghosh, D., Barrette, T., Pandey, A. and Chinnaiyan, A. M. (2004). ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* **6**, 1-6.

Ring, A. and Dowsett, M. (2004). Mechanisms of tamoxifen resistance. *Endocr Relat Cancer* **11**, 643-58.

Roa, B. B., Boyd, A. A., Volcik, K. and Richards, C. S. (1996). Ashkenazi Jewish population frequencies for common mutations in BRCA1 and BRCA2. *Nat Genet* **14**, 185-7.

Rogan, E. G., Badawi, A. F., Devanesan, P. D., Meza, J. L., Edney, J. A., West, W. W., Higginbotham, S. M. and Cavalieri, E. L. (2003). Relative imbalances in estrogen metabolism and conjugation in breast tissue of women with carcinoma: potential biomarkers of susceptibility to cancer. *Carcinogenesis* **24**, 697-702.

Ronckers, C. M., Erdmann, C. A. and Land, C. E. (2005). Radiation and breast cancer: a review of current evidence. *Breast Cancer Res* **7**, 21-32.

Rossouw, J. E., Anderson, G. L., Prentice, R. L., LaCroix, A. Z., Kooperberg, C., Stefanick, M. L., Jackson, R. D., Beresford, S. A., Howard, B. V., Johnson, K. C. et al. (2002). Risks and benefits of estrogen plus progestin in healthy postmenopausal

women: principal results From the Women's Health Initiative randomized controlled trial. *JAMA : the journal of the American Medical Association* **288**, 321-33.

Rozen, S. and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* **132**, 365-86.

Ruan, W., Monaco, M. E. and Kleinberg, D. L. (2005). Progesterone stimulates mammary gland ductal morphogenesis by synergizing with and enhancing insulin-like growth factor-I action. *Endocrinology* **146**, 1170-8.

Russo, J. and Russo, I. H. (1994). Toward a physiological approach to breast cancer prevention. *Cancer Epidemiol Biomarkers Prev* **3**, 353-64.

Saar, K., Beck, A., Bihoreau, M. T., Birney, E., Brocklebank, D., Chen, Y., Cuppen, E., Demonchy, S., Dopazo, J., Flicek, P. et al. (2008). SNP and haplotype mapping for genetic analysis in the rat. *Nat Genet* **40**, 560-6.

Safe, S. H. (1997). Is there an association between exposure to environmental estrogens and breast cancer? *Environ Health Perspect* **105 Suppl 3**, 675-8.

Saitoh, M., Ohmichi, M., Takahashi, K., Kawagoe, J., Ohta, T., Doshida, M., Takahashi, T., Igarashi, H., Mori-Abe, A., Du, B. et al. (2005). Medroxyprogesterone acetate induces cell proliferation through up-regulation of cyclin D1 expression via phosphatidylinositol 3-kinase/Akt/nuclear factor-kappaB cascade in human breast cancer cells. *Endocrinology* **146**, 4917-25.

Samuelson, D. J., Haag, J. D., Lan, H., Monson, D. M., Shultz, M. A., Kolman, B. D. and Gould, M. N. (2003). Physical evidence of Mcs5, a QTL controlling mammary carcinoma susceptibility, in congenic rats. *Carcinogenesis* **24**, 1455-60.

Samuelson, D. J., Hesselson, S. E., Aperavich, B. A., Zan, Y., Haag, J. D., Trentham-Dietz, A., Hampton, J. M., Mau, B., Chen, K. S., Baynes, C. et al. (2007). Rat Mcs5a is a compound quantitative trait locus with orthologous human loci that associate with breast cancer risk. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 6299-304.

Santen, R. J., Brodie, H., Simpson, E. R., Siiteri, P. K. and Brodie, A. (2009). History of aromatase: saga of an important biological mediator and therapeutic target. *Endocr Rev* **30**, 343-75.

Schechter, A. L., Hung, M. C., Vaidyanathan, L., Weinberg, R. A., Yang-Feng, T. L., Francke, U., Ullrich, A. and Coussens, L. (1985). The neu gene: an erbB-homologous gene distinct from and unlinked to the gene encoding the EGF receptor. *Science (New York, N Y)* **229**, 976-8.

Seal, S., Thompson, D., Renwick, A., Elliott, A., Kelly, P., Barfoot, R., Chagtai, T., Jayatilake, H., Ahmed, M., Spanova, K. et al. (2006). Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. *Nat Genet* **38**, 1239-41.

Shepel, L. A., Lan, H., Haag, J. D., Brasic, G. M., Gheen, M. E., Simon, J. S., Hoff, P., Newton, M. A. and Gould, M. N. (1998). Genetic identification of multiple loci that control breast cancer susceptibility in the rat. *Genetics* **149**, 289-99.

Shull, J. D., Pennington, K. L., Reindl, T. M., Snyder, M. C., Strecker, T. E., Spady, T. J., Tochacek, M. and McComb, R. D. (2001). Susceptibility to estrogen-induced mammary cancer segregates as an incompletely dominant phenotype in reciprocal crosses between the ACI and Copenhagen rat strains. *Endocrinology* **142**, 5124-30.

Shull, J. D., Spady, T. J., Snyder, M. C., Johansson, S. L. and Pennington, K. L. (1997). Ovary-intact, but not ovariectomized female ACI rats treated with 17beta-estradiol rapidly develop mammary carcinoma. *Carcinogenesis* **18**, 1595-601.

Sica, A., Schioppa, T., Mantovani, A. and Allavena, P. (2006). Tumour-associated macrophages are a distinct M2 polarised population promoting tumour progression: potential targets of anti-cancer therapy. *Eur J Cancer* **42**, 717-27.

Singletary, K. W. and McNary, M. Q. (1992). Effect of moderate ethanol consumption on mammary gland structural development and DNA synthesis in the female rat. *Alcohol* **9**, 95-101.

Sinha, R., Gustafson, D. R., Kulldorff, M., Wen, W. Q., Cerhan, J. R. and Zheng, W. (2000). 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine, a carcinogen in high-temperature-cooked meat, and breast cancer risk. *J Natl Cancer Inst* **92**, 1352-4.

Smith, P., McGuffog, L., Easton, D. F., Mann, G. J., Pupo, G. M., Newman, B., Chenevix-Trench, G., Szabo, C., Southey, M., Renard, H. et al. (2006). A genome wide linkage search for breast cancer susceptibility genes. *Genes Chromosomes Cancer* **45**, 646-55.

Smits, B., Sharma, D., Samuelson, D., Woditschka, S., Mau, B., Haag, J. and Gould, M. (2011a). 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 Research* **13**, R81.

Smits, B. M., Sharma, D., Samuelson, D. J., Woditschka, S., Mau, B., Haag, J. D. and Gould, M. N. (2011b). 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* **13**, R81.

Solari, F., Bateman, A. and Ahringer, J. (1999). The Caenorhabditis elegans genes egl-27 and egr-1 are similar to MTA1, a member of a chromatin regulatory complex, and are redundantly required for embryonic patterning. *Development* **126**, 2483-94.

Stacey, S. N., Manolescu, A., Sulem, P., Thorlacius, S., Gudjonsson, S. A., Jonsson, G. F., Jakobsdottir, M., Bergthorsson, J. T., Gudmundsson, J., Aben, K. K. et al. (2008). Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* **40**, 703-6.

Stewart, T. A., Pattengale, P. K. and Leder, P. (1984). Spontaneous mammary adenocarcinomas in transgenic mice that carry and express MTV/myc fusion genes. *Cell* **38**, 627-37.

Struewing, J. P., Abeliovich, D., Peretz, T., Avishai, N., Kaback, M. M., Collins, F. S. and Brody, L. C. (1995). The carrier frequency of the BRCA1 185delAG mutation is approximately 1 percent in Ashkenazi Jewish individuals. *Nat Genet* **11**, 198-200.

Thomas, D. (2010). Methods for investigating gene-environment interactions in candidate pathway and genome-wide association studies. *Annu Rev Public Health* **31**, 21-36.

Thomas, G., Jacobs, K. B., Kraft, P., Yeager, M., Wacholder, S., Cox, D. G., Hankinson, S. E., Hutchinson, A., Wang, Z., Yu, K. et al. (2009). A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). *Nat Genet* **41**, 579-84.

Thorne, L. B., Grant, A. L., Paterno, G. D. and Gillespie, L. L. (2005). Cloning and characterization of the mouse ortholog of mi-er1. *DNA Seq* **16**, 237-40.

Todorovic, R., Ariese, F., Devanesan, P., Jankowiak, R., Small, G. J., Rogan, E. and Cavalieri, E. (1997). Determination of benzo[a]pyrene- and 7,12-dimethylbenz[a]anthracene-DNA adducts formed in rat mammary glands. *Chem Res Toxicol* **10**, 941-7.

Travis, R. C., Reeves, G. K., Green, J., Bull, D., Tipper, S. J., Baker, K., Beral, V., Peto, R., Bell, J., Zelenika, D. et al. (2010). Gene-environment interactions in 7610 women with breast cancer: prospective evidence from the Million Women Study. *Lancet* **375**, 2143-51.

Trimboli, A. J., Cantemir-Stone, C. Z., Li, F., Wallace, J. A., Merchant, A., Creasap, N., Thompson, J. C., Caserta, E., Wang, H., Chong, J. L. et al. (2009). Pten in stromal fibroblasts suppresses mammary epithelial tumours. *Nature* **461**, 1084-91.

Tseng, L. and Gurpide, E. (1975). Effects of progestins on estradiol receptor levels in human endometrium. *J Clin Endocrinol Metab* **41**, 402-4.

Tsutsui, T., Maizumi, H., McLachlan, J. A. and Barrett, J. C. (1983). Aneuploidy induction and cell transformation by diethylstilbestrol: a possible chromosomal mechanism in carcinogenesis. *Cancer Res* **43**, 3814-21.

Tsutsui, T., Tamura, Y., Suzuki, A., Hirose, Y., Kobayashi, M., Nishimura, H., Metzler, M. and Barrett, J. C. (2000). Mammalian cell transformation and aneuploidy induced by five bisphenols. *Int J Cancer* **86**, 151-4.

Turnbull, C., Ahmed, S., Morrison, J., Pernet, D., Renwick, A., Maranian, M., Seal, S., Ghoussaini, M., Hines, S., Healey, C. S. et al. (2010). Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat Genet* **42**, 504-7.

Vahteristo, P., Bartkova, J., Eerola, H., Syrjakoski, K., Ojala, S., Kilpivaara, O., Tamminen, A., Kononen, J., Aittomaki, K., Heikkila, P. et al. (2002). A CHEK2 genetic variant contributing to a substantial fraction of familial breast cancer. *Am J Hum Genet* **71**, 432-8.

van Boxtel, R., Kuiper, R. V., Toonen, P. W., van Heesch, S., Hermsen, R., de Bruin, A. and Cuppen, E. (2011). Homozygous and Heterozygous p53 Knockout Rats Develop Metastasizing Sarcomas with High Frequency. *The American Journal of Pathology* **179**, 1616-1622.

Vargova, K., Curik, N., Burda, P., Basova, P., Kulvait, V., Pospisil, V., Savvulidi, F., Kokavec, J., Necas, E., Berkova, A. et al. (2011). MYB transcriptionally regulates the miR-155 host gene in chronic lymphocytic leukemia. *Blood* **117**, 3816-25.

Wang, B., Kennan, W. S., Yasukawa-Barnes, J., Lindstrom, M. J. and Gould, M. N. (1991). Overcoming the activity of mammary carcinoma suppressor gene in Copenhagen rats by v-H-ras oncogene transfer into mammary epithelial cells in situ. *Cancer Res* **51**, 5298-303.

Wang, S., Saboorian, M. H., Frenkel, E., Hynan, L., Gokaslan, S. T. and Ashfaq, R. (2000). Laboratory assessment of the status of Her-2/neu protein and oncogene in breast cancer specimens: comparison of immunohistochemistry assay with fluorescence in situ hybridisation assays. *J Clin Pathol* **53**, 374-81.

Wang, T. C., Cardiff, R. D., Zukerberg, L., Lees, E., Arnold, A. and Schmidt, E. V. (1994). Mammary hyperplasia and carcinoma in MMTV-cyclin D1 transgenic mice. *Nature* **369**, 669-71.

Williams, J. M. and Daniel, C. W. (1983). Mammary ductal elongation: differentiation of myoepithelium and basal lamina during branching morphogenesis. *Dev Biol* **97**, 274-90.

Wooster, R. and Weber, B. L. (2003). Breast and ovarian cancer. *The New England journal of medicine* **348**, 2339-47.

Yu, J., Rhodes, D. R., Tomlins, S. A., Cao, X., Chen, G., Mehra, R., Wang, X., Ghosh, D., Shah, R. B., Varambally, S. et al. (2007). A polycomb repression signature in metastatic prostate cancer predicts cancer outcome. *Cancer Res* **67**, 10657-63.

Zhang, R., Haag, J. D. and Gould, M. N. (1990). Site of expression and biological function of the rat mammary carcinoma suppressor gene. *Carcinogenesis* **11**, 1765-70.

Zhang, Y. and Chen, H. (2011). Genistein, an epigenome modifier during cancer prevention. *Epigenetics* **6**, 888-91.

Zheng, W., Deitz, A. C., Campbell, D. R., Wen, W. Q., Cerhan, J. R., Sellers, T. A., Folsom, A. R. and Hein, D. W. (1999). N-acetyltransferase 1 genetic polymorphism, cigarette smoking, well-done meat intake, and breast cancer risk. *Cancer Epidemiol Biomarkers Prev* **8**, 233-9.

Zheng, W., Gustafson, D. R., Sinha, R., Cerhan, J. R., Moore, D., Hong, C. P., Anderson, K. E., Kushi, L. H., Sellers, T. A. and Folsom, A. R. (1998). Well-done meat intake and the risk of breast cancer. *J Natl Cancer Inst* **90**, 1724-9.

Zheng, W. and Lee, S. A. (2009). Well-done meat intake, heterocyclic amine exposure, and cancer risk. *Nutr Cancer* **61**, 437-46.

Zheng, W., Wen, W., Gao, Y. T., Shyr, Y., Zheng, Y., Long, J., Li, G., Li, C., Gu, K., Cai, Q. et al. (2010). Genetic and clinical predictors for breast cancer risk assessment and stratification among Chinese women. *Journal of the National Cancer Institute* **102**, 972-81.

Zheng, W., Wen, W. Q., Gustafson, D. R., Gross, M., Cerhan, J. R. and Folsom, A. R. (2002). GSTM1 and GSTT1 polymorphisms and postmenopausal breast cancer risk. *Breast Cancer Res Treat* **74**, 9-16.

APPENDIX I

Introduction

During the course of my dissertation work some supporting experiments were not completed or were omitted into dissertation chapters as to not distract from the focus of the project. Despite this, some of the experiments and resulting data not included in the main body of this dissertation contributed to the conclusions of the work presented here and therefore are shown in the appendix. Below, these data will be explained in the context of the previous work described.

Endogenous Control for Taqman Gene Expression Studies

Introduction

Early attempts to measure gene expression of rat *Mcs1b* transcripts were performed using *Gapdh* as an endogenous control. However, the variability in *Gapdh* expression was high between samples and this was believed to be due to poor RNA quality. RNA quality was within acceptable limits (~1.8) for absorbance ratios at 260 nm and 280 nm when analyzed on a Nanodrop 2000 (Thermo Fisher Scientific); regardless, attempts were made to improve RNA isolation times and conditions but resulted in no change in the high variability in *Gapdh* expression between samples. I further analyzed RNA quality using RNA Pico Chips and an Agilent Bioanalyzer (Agilent, Santa Clara, CA). Based on RNA Integrity Numbers (RIN) determined by Bioanalyzer data, RNA was determined to not be compromised and that variability was due to variations in *Gapdh* expression itself and was not a good endogenous control candidate. Thus, I sought to identify a better endogenous control by measuring a panel of endogenous control gene candidates in WF and resistant WF.COP mammary glands.

Materials and Methods

RNA clean-up using sodium acetate

As stated above, *Gapdh* expression was highly variable between samples in initial gene expression assays. I initially believed this was due to degraded or contaminated RNA. To improve RNA quality I made some changes to the RNA isolation protocol. First, fewer samples were isolated in individual sessions to reduce the amount of time the RNA isolates were on ice. Total RNA was extracted with TRI-Reagent (Molecular Research Center) followed by standard chloroform/ethanol precipitation. To reduce possible solvent and DNA contamination RNA samples were further processed by a 1/10 v/v 3M sodium acetate and 2.5x v/v 100% ethanol wash on ice for 10 minutes followed by 80% ethanol wash followed by Turbo DNase (Life Technologies) treatment. Total RNA quantity and quality were measured with a Nanodrop 1000 (Fisher Scientific) and a Bioanalyzer with RNA 6000 NanoChips (Agilent).

Endogenous Control Array

To determine a good endogenous control candidate for the *Mcs1b* gene expression studies, I used an endogenous control array micro-fluidic card (Life Technologies). The card contains 16 different endogenous control probe/primer sets (Table S1) preloaded into the card. cDNA was prepared from four mammary carcinoma resistant WF.COP and four susceptible WF mammary gland samples and normalized to 2 μg. They were loaded in triplicate into the endogenous control array card according to manufacturer specifications, centrifuged and run on an ABI Prism 7900 HT using standard conditions.

Statistical Analysis

Endogenous control data were analyzed using SDS software (Life Technologies) and StatView (SAS Institute). Geometric means were calculated for each sample and Mann-Whitney tests were used to analyze expression between genotypes.

Results

3M sodium acetate results in higher quality RNA

Total RNA had higher RIN and 28S and 18S bands were stronger when analyzed with the Bioanalyzer. However, differences in *Gapdh* expression between WF and resistant mammary gland samples did not change following reducing time on ice or by sodium acetate cleanup.

Rplp2 is a good endogenous control gene for gene expression assays in rat mammary gland samples

Expression levels were highly variable between the genes in the panel (Figure S1). Ct values for most of the *Mcs1b* target transcripts were in a range of ~20 - 25 cycles (data not shown). Therefore, genes with mean Ct values outside the range of 20 - 25 cycles were omitted from further analysis. Candidate control genes were further analyzed for variance between replicates and between samples. Samples with SD higher than 0.5 Ct were excluded.

Seven genes in the panel fell inside the allowable Ct range: *Actbl2*, *Arbp*, *B2m*, *Ppia*, *Ppib*, *Rplp2*, and *Ubc* (Table S2). Of these, only *Rplp2* and *Ubc* had SDs below 0.5 cycles for both genotype groups. *Ubc* was differentially expressed between WF and

Table S1:

Endogenous control genes contained on ABI Endogenous Control Array MicroFluidic Card

Name (Alternate Name)	Gene Symbol	ABI product reference number
18S	18S	Hs99999901
Beta-actin	Actb	Rn00667869
Attachment region binding protein	Arbp	Rn00821065
Beta-2 microglobulin	B2m	Rn00560865
Glyceraldehyde-3-phosphate dehydrogenase	Gapdh	Rn99999916
Beta-glucuronidase	Gusb	Rn00566655
Hydroxymethylbilane synthase	Hmbs	Rn00565886
Hypoxanthine phosphoribosyltransferase 1	Hprt	Rn01527840
Phosphoglycerate kinase 1	Pgk1	Rn00821429
Peptidylprolyl isomerase A (Cyclophilin A)	Ppia	Rn00690933
Peptidylprolyl isomerase B (Cyclophilin B)	Ppib	Rn00574762
Ribosomal protein, large, P2	Rplp2	Rn014799271
TATA box binding protein	Tbp	Rn01455648
Transferrin receptor (p90, CD71)	Tfrc	Rn01474695
Ubiquitin C	Ubc	Rn017898121
Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide	Ywhaz	Rn00755072

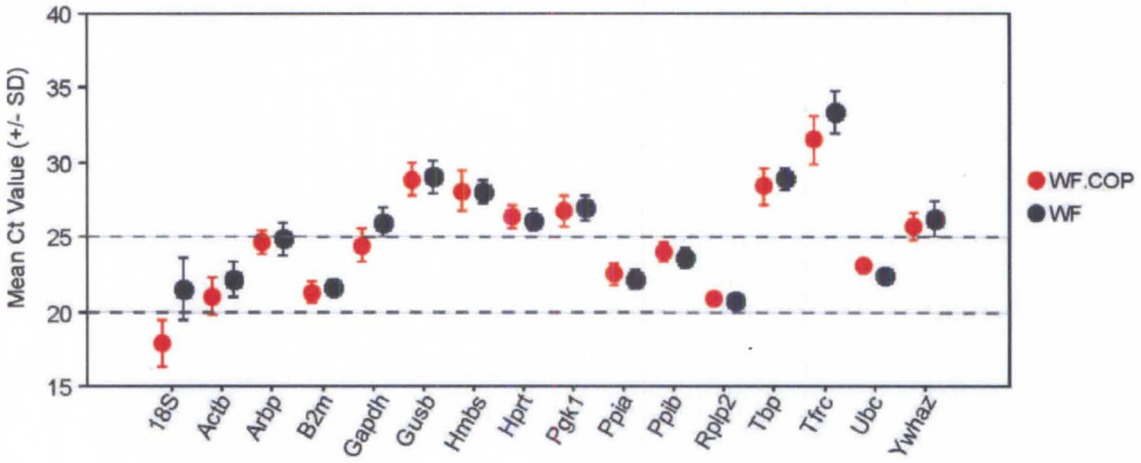


Figure S1: Mean Ct Values for Genes tested in ABI Endogenous Control Array. The geometric mean of the Ct value for each genotype is plotted. Error bars represent SD. The dashed line delineates the Ct value range desired for an optimal endogenous control.

Table S2:

Results of ABI Endogenous Control Array Micro-Fluidic Assay

Gene Symbol	Mean Ct value ± S.D.		P value[a]
	WF	WF.COP	
18S	21.5 ± 2.3	17.9 ± 1.8	0.0433
Actb	22.2 ± 1.3	21.0 ± 1.3	0.2482
Arbp	24.8 ± 1.2	24.6 ± 0.8	0.3865
B2m	21.6 ± 0.6	21.3 ± 0.8	0.7728
Gapdh	26.0 ± 1.1	24.4 ± 1.2	0.1489
Gusb	29.0 ± 1.2	28.8 ± 1.2	>.9999
Hmbs	28.0 ± 0.9	28.1 ± 1.5	0.7728
Hprt	26.1 ± 0.7	26.4 ± 0.9	0.7728
Pgk1	26.9 ± 1.0	26.7 ± 1.1	0.7728
Ppia	22.2 ± 0.7	22.5 ± 0.8	0.7728
Ppib	23.6 ± 0.7	24.0 ± 0.7	0.3865
Rplp2	20.7 ± 0.5	20.8 ± 0.4	0.7728
Tbp	28.9 ± 0.7	28.4 ± 1.3	>.9999
Tfrc	33.3 ± 1.6	31.5 ± 1.8	0.1489
Ubc	22.4 ± 0.5	23.0 ± 0.2	0.0433
Ywhaz	26.2 ± 1.3	25.7 ± 1.0	0.7728

[a] Non-parametric Mann-Whitney test

resistant WF.COP samples (P=0.0433); however, *Rplp2* was not different between genotypes (P=0.7728). Therefore, *Rplp2* was used in all subsequent gene expression experiments.

APPENDIX II

LIST OF ABBREVIATIONS AND SYMBOLS

17β-HSD	17β-hydroxysteroid dehydrogenase
18S	18S
ACF	Adherent cell fraction
ACI	ACI rat strain
Actb	Beta-actin
AI	Aromatase inhibitor
Ankrd55	Ankyrin repeat domain 55
Arbp	Attachment region binding protein
B[a]P	Benzo(a)pyrene
B2m	Beta-2 microglobulin
BAC	Bacterial artificial chromosome
BC	Backcross generation
BN	Brown Norway rat strain
bp	Base pairs
BPA	Bisphenol A
C5orf35	Chromosome 5 open reading frame 35 (SET domain-containing protein 9)
cDNA	Complementary DNA
gDNA	Genomic DNA
Chr	Chromosome
CI	Confidence Interval

CK1	Casein kinase 1
cM	CentiMorgan
CNV	Copy number variant
COP/Hsd	Copenhagen (Harlan Sprague Dawley) rat strain
CYP	Cytochrome P450
DCIS	Ductal carcinoma *in situ*
Ddx4	DEAD (Asp-Glu-Ala-Asp) box polypeptide 4
DES	Diethylstilbestrol
DMBA	7,12 Dimethylbenz(a)anthracene
DMEM	Dulbecco's Modified Eagle Medium
DNase	Deoxyribonuclease I
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytidine triphosphate
dGTP	Deoxyguanosine triphosphate
dTTP	Deoxythymidine triphosphate
dNTP	Deoxynucelotide triphosphate
E1	Estrone
E2	Estradiol
E3	Estriol
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
ELM2	Egl-27 and MTA1 homology 2 domain

ERE	Estrogen Response Element
ERα	Estrogen receptor alpha
ERβ	Estrogen receptor beta
Gapdh	Glyceraldehyde-3-phosphate dehydrogenase
Gpbp1	GC-rich promoter binding protein 1
Gusb	Beta-glucuronidase
GWAS	Genome-wide association study
HCA	Heterocyclic amine
HDAC	Histone deacetylase
Hmbs	Hydroxymethylbilane synthase
Hprt	Hypoxanthine phosphoribosyltransferase 1
HR	Hazard Ratio
HRT	Hormone replacement therapy
IACUC	Institutional Animal Care and Use Committee
IDBC	Invasive ductal breast carcinoma
Il31ra	Interleukin 31 receptor A
Il6st	Interleukin 6 signal transducer
ILBC	Invasive lobular breast carcinoma
indel	Nucleotide Insertion/Deletion variant
LD	Linkage disequilibrium
LTR	Long terminal repeat
LxxLL	Leu-Xaa-Xaa-Leu-Leu domain

MAF	Minor allele frequency
Map3k1	Mitogen-activated protein kinase kinase kinase 1
Mcs	Mammary carcinoma susceptibility locus
MEC	Mammary epithelial cell
MHC	Major histocompatibility complex
Mier3	Mesoderm induction early response 1, family member 3
miRNA	MicroRNA
MMTV	Murine mammary tumor virus
MTA1	Metastatic tumor antigen 1
MWS	Million Women Study
NCI	National Cancer Institute
nitro-PAH	Nitrated polycyclic aromatic hydrocarbon metabolite
NMU	*N*-Nitroso-*N*-methylurea
NSABP	National Surgical Adjuvant Breast and Bowel Project
NuRD	Nucleosome remodeling and histone deacetylase complex
OR	Odds ratio
ORF	Open reading frame
PAH	Polycyclic aromatic hydrocarbon
PCB	Polychlorinated biphenyl
PCR	Polymerase chain Reaction
Pgk1	Phosphoglycerate kinase 1
PhIP	2-amino-1-methyl-6-phenylimidazo(4,5-b)pyridine

PI3K	Phosphatidylinositol 3' kinase
Ppap2a	Phosphatidic acid phosphatase type 2A
Ppia	Peptidylprolyl isomerase A (Cyclophilin A)
Ppib	Peptidylprolyl isomerase B (Cyclophilin B)
PR-A	Progesterone receptor A
PR-B	Progesterone receptor B
PyV mT	Polyoma virus middle T antigen
QPCR	Quantitative polymerase chain reaction
QTL	Quantitative trait locus
R	Resistant mammary carcinoma phenotype
RF	Radiofrequency/microwave radiowave emissions
RIN	RNA integrity number
Rplp2	Ribosomal protein, large, P2
RR	Relative Risk
S	Susceptible mammary carcinoma phenotype
SANT	Swi3, Ada2, N-Cor, and TFIIIB domain
SD	Standard deviation
SERM	Selective Estrogen Receptor Modulator
SETD9	SET domain containing 9
Slc38a9	Solute carrier family 38 , member 9
SNP	Single Nucleotide Polymorphism
snRNA	Small nuclear RNA

SS	Dahl/Salt-Sensitive rat strain
STAR	Specific Targeted Research Project
Tbp	TATA box binding protein
TEB	Terminal end bud
tet	Tetracycline
Tfrc	Transferrin receptor (p90, CD71)
U	Undetermined mammary carcinoma phenotype
U6	Spliceosomal small non-coding nuclear RNA component
Ubc	Ubiquitin C
UTR	Untranslated region
WF/Hsd	Wistar Furth (Harlan Sprague Dawley) rat strain
(WF X COP) F1	First generation from a Wistar Furth/Copenhagen mating
WHI	Women's Health Initiative
WKy	Wistar Kyoto rat strain
WMG	Whole mammary gland transplant
Ywhaz	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide

CURRICULUM VITAE

Aaron denDekker
University of Louisville School of Medicine
Department of Biochemistry and Molecular Biology
Louisville, Kentucky 40201
502-852-7798
Addend01@exchange.louisville.edu

EDUCATION

University of Louisville, Louisville, KY
Ph.D. in Biochemistry and Molecular Biology **2013**
Mentor: David J. Samuelson, PhD
Dissertation Title: "Mapping and Characterization of Rat Mammary Carcinoma Susceptibility QTL *Mcs1b*"

University of Louisville, Louisville, KY
M.S. in Biochemistry and Molecular Biology **2012**

University of Louisville, Louisville, KY
B.S. in Chemistry **2001**
Area of Concentration: Biochemistry

RESEARCH EXPERIENCE

University of Louisville, Department of Biochemistry and Molecular Biology
Predoctoral Research with David Samuelson, PhD **2008-2012**

Dissertation Project: Genetic Mapping and Mechanism of Action of Rat QTL *Mcs1b*

Research Focus: My research focused on studying a quantitative trait locus in a congenic rat model that associates with an increased susceptibility to developing mammary carcinomas. This locus is identified as mammary carcinoma susceptibility locus 1b, or *Mcs1b*. This work was composed of three independent parts that focused on 1) fine-mapping the interval of the *Mcs1b* QTL on chromosome *2* using congenic rat lines containing a resistant *Mcs1b* allele on a susceptible background, 2) identifying whether the *Mcs1b*-conferred susceptibility phenotype is intrinsic to the mammary gland, and 3) developing a transcript map of genes lying in the locus by sequencing and analyzing gene expression profiles.

Methods/Techniques: Standard molecular biology techniques (i.e. DNA / RNA isolation from tissues and cells, PCR, western blots, northern blots, etc.), cloning, Sanger DNA sequencing, flow cytometry, histology, RT-PCR, gene expression assays, primary cell isolation and culture, radiolabeling, genotyping, animal usage and handling, survival and non-survival surgical techniques.

University of Louisville, Department of Biochemistry and Molecular Biology
Research Technologist in the Zebrafish Mapping Facility in the lab of Ronald Gregg, PhD **2006-2008**

Research Focus: My research focused on mapping and cloning mutations in Zebrafish in collaboration with outside investigators. Single point mutations were induced using N-ethyl-N-nitrosurea (ENU) and phenotypes were screened by collaborators. Selected mutations were mapped by our group using microsatellite and single nucleotide polymorphism (SNP) genotyping markers. Once intervals were fine mapped, gene candidates were obtained, cloned and sequenced to identify specific mutations leading to the phenotype.

Methods/Techniques: Positional mapping, PCR, microsatellite and SNP genotyping, cloning, DNA / RNA isolation, Sanger DNA sequencing, pyrosequencing, allelic discrimination assays (real time-PCR based).

University of Louisville, Department of Microbiology and Immunology
Research Technologist in the lab of Jill Suttles, PhD **2002-2004**

Research Focus: My research focused on two separate projects studying fatty-acid binding proteins (FABPs) and inflammation. The first project focused on the FABP adipocyte protein 2 (aP2) and a role in cholesterol trafficking and inflammatory activity of macrophages. The second project focused on macrophage FABPs and a role in the onset and progression of murine experimental auto-immune encephalitis (EAE), a mouse model of multiple sclerosis.

Methods/Techniques: Cell culture, primary cell isolation, animal usage and handling, murine bone marrow isolation, mouse brain and spinal cord micro-dissection, histology, immunohistochemistry, ELISAs, flow cytometry, western blots, northern blots, arginase assays, cell proliferation assays.

University of Louisville, Department of Ophthalmology

Research Technologist in the lab of Maureen McCall, PhD **2001-2002**

Research Focus: My work focused on a project studying the gamma aminobutyric acid C (GABA-C) receptor in the retina and its role in excitatory responses in vision.

Methods/Techniques: Animal usage and handling, mouse retina and brain micro-dissection, histology, immunohistochemistry, PCR, genotyping, electroretinograms, DNA / RNA isolation, northern blots.

University of Louisville, Department of Biochemistry and Molecular Biology

Undergraduate Research project in the lab of Thomas Wheeler, PhD **1999**

Research Focus: This undergraduate project focused on studying the ability of fructose-1, 6-bisphosphate (FBP) to permeate artificial multi-lamellar vesicles as a model for uptake in cardiomyocytes.

Methods/Techniques: Liposome preparation from phosphatidyl choline, radiolabeled-FBP uptake assays, enzymatic assays to measure inorganic phosphate.

RELATED EXPERIENCE

Baptist Hospital East, Louisville KY

Histotechnologist **2004-2006**

Duties: Assist pathologist with gross anatomy of surgical and dermatological specimens, basic histology of specimens, i.e. process, embed, cut and stain, provide general support for the clinical histology laboratory.

Jewish Hospital, Louisville KY

Histotechnologist **2004-2006**

Duties were the same as described above.

AWARDS

- Homberger Scholarship **2012**
- NIEHS T-32 Predoctoral Training Fellowship **2010-2012**
  Title: UofL Environmental Health Sciences Training Program
  PI/Project Leader: David W. Hein, PhD
- Integrated Program in Biological Sciences (IPIBS) Fellowship **2008-2010**

LEADERSHIP

- Biochemistry and Molecular Biology Student President **2009-2010**
- Biochemistry and Molecular Biology Student Vice-President **2010-2011**
- Graduate Executive Committee Representative **2009-2010**
- Training of Undergraduate Summer Students **2009-2011**

PUBLICATIONS

denDekker AD, Xu X, Vaughn MD, Puckett AH, Gardner LL, Lambring CJ, Deschenes L, Samuelson DJ. Rat *Mcs1b* is concordant to the genome wide association identified breast cancer locus at human *5q11.2* and *Mier3* is a candidate cancer susceptibility gene. *Cancer Research* 2012; 72(22):6002-12. PMID: 22993404

Thomas JL, Vihtelic TS, **denDekker AD**, Willer G, Luo X, Murphy TR, Gregg RG, Hyde DR, Thummel R. The loss of vacuolar protein sorting 11 (vps11) causes retinal pathogenesis in a vertebrate model of syndromic albinism. *Invest Ophthalmol Vis Sci.* 2011; 52(6):3119-28. PMID: 21622697.

Wheeler TJ, McCurdy JM, **denDekker AD**, Chien S. Permeability of fructose-1,6-bisphosphate in liposomes and cardiac myocytes. *Mol Cell Biochem.* 2004; 259(1-2):105-14. PMID: 15124914

INVITED TALKS

Cold Spring Harbor, NY
Rat Genomics and Models Meeting, 2009
Title: Rat *Mcs1b* is a mammary gland autonomous QTL and an ortholog of a breast cancer susceptibility locus on human chromosome 5q.

ABSTRACTS PRESENTED

Samuelson DJ, Xu X, **denDekker AD**, Sanders J, Kemper AF, Kalbfleisch T. Congenic Mapping and Functional Characterization of a Rat Mammary Carcinoma Susceptibility QTL, *Mcs1b*, Identifies *MAP3K1* and *MIER3* as Candidate Breast Cancer Susceptibility Genes. *Quantitative Genetics & Genomics Gordon Research Conference*, 2011.

Samuelson DJ, Xu X, **denDekker AD**, Sanders J, Vaughn MD, Lambring CJ, Puckett AH. Comparative Environmental Genomics of Common Breast Cancer Susceptibility Alleles. *NIH/NIEHS Center Director's Meeting*, 2010.

denDekker AD, Puckett AH, Vaughn MD, Lambring CJ, and Samuelson DJ. Rat *Mcs1b* is a mammary gland autonomous QTL and an ortholog of a breast cancer susceptibility locus on human chromosome 5q. *Rat Genomics & Models Annual Meeting, Cold Spring Harbor Laboratory*, 2009.

Samuelson DJ, **denDekker AD**, and Puckett AH. Rat *Mcs1b* is a comparative genetic model of a human breast cancer susceptibility locus. *Poster Presentation, Annual NIEHS P30 Directors' Meeting, University of Pennsylvania, Philadelphia*, 2008.