Mapping of Three Genetic Determinants of Susceptibility to Estrogen-Induced Mammary Cancer within the *Emca8* Locus on Rat Chromosome 5*

Beverly S. Schaffer¹,², Kristin M. Leland-Wavrin¹,³, Scott G. Kurz¹,², John A. Colletti³, Nicole L. Seiler³, Christopher L. Warren³ and James D. Shull¹,²,³,#

¹Department of Genetics, Cell Biology and Anatomy; ²Eppley Institute for Research in Cancer, University of Nebraska Medical Center, Omaha, Nebraska 68198-5805. ³McArdle Laboratory for Cancer Research, Department of Oncology, and UW Carbone Cancer Center, School of Medicine and Public Health, University of Wisconsin, Madison, Wisconsin 53706-1599.

Key words: ACI rat, Brown Norway rat, 17β-estradiol, mammary cancer, *Emca8*, congenic rat strain

Running Title: *Emca8* and Mammary Cancer Susceptibility

Conflict of interest. The authors have no conflicts of interest to disclose.
Footnotes to title page:

* This work was supported by grant R01-CA77876 from the National Cancer Institute. Shared research resources at the University of Nebraska Medical Center were supported by Cancer Center Support Grant P30-CA036727 and by grant P20-RR16469; those at the University of Wisconsin-Madison were supported by Cancer Center Support Grant P30-CA014520. B.S. was supported in part by postdoctoral fellowship DAMD17-03-1-0477 from the Department of Defense Breast Cancer Training Program and by training grant DAMD17-00-1-0361. J.C. was supported in part by training grant T32-CA009135.

# Address correspondence to: Dr. James Shull, McArdle Laboratory for Cancer Research, University of Wisconsin, 1400 University Avenue, Madison, Wisconsin (USA). Phone: (608) 265-5003. Fax: (608) 262-2824. Email: shull@oncology.wisc.edu
Abstract

The ACI rat model of 17β-estradiol (E2)-induced mammary cancer has gained wide use in the study of breast cancer etiology, prevention and genetics. *Emca8*, a QTL that determines susceptibility to E2-induced mammary cancer, was previously mapped to rat chromosome 5 (RNO5) in an intercross between resistant Brown Norway (BN) and susceptible ACI rats. In this study, a panel of congenic rat strains, each of which carries BN alleles across a defined segment of RNO5 on the ACI genetic background, was generated and used to map more precisely the *Emca8* determinants of mammary cancer susceptibility. Three distinct genetic determinants were localized within *Emca8*, and two of these were mapped to intervals of less than 15 megabases. *Emca8.1* harbors *Cdkn2a*, *Cdkn2b* and other genes and is orthologous to the 9p21 breast cancer locus identified in genome wide and candidate gene association studies. *Emca8.2* harbors *Cdkn2c* and other genes and is orthologous to the 1p32 locus in humans that is frequently deleted in breast cancers. Both *Emca8.1* and *Emca8.2* harbor copy number variants that are orthologous to copy number variant regions in humans. Gene expression profiles were defined for mammary tissues from E2 treated ACI and ACI.BN-Emca8 rats to define the impact of *Emca8* on gene expression and identify differentially expressed genes residing within *Emca8.1* and *Emca8.2*. This study further illustrates the relevance of the ACI rat model of E2-induced mammary cancer for identifying novel genetic determinants of breast cancer susceptibility and defining the mechanisms through which estrogens contribute to breast cancer development.
Introduction

Breast cancer remains the second leading cause of cancer related mortality among women in the United States in spite of numerous advances in diagnosis and treatment of this disease over the past three decades. Multiple factors contribute to or otherwise influence breast cancer development, including the female sex steroid hormones. The evidence implicating estrogens in breast cancer etiology is particularly convincing. Several population-based studies associate lifelong exposure to endogenous or exogenous estrogens with development of breast cancer (1-4). Moreover, data from the Women’s Health Initiative prospective study indicate that use by postmenopausal women of hormone replacement regimens that contain estrogens and a progestin significantly increases the risk of developing breast cancer (5, 6). Multiple studies indicate that inhibition of estrogen action through the use of selective estrogen receptor modulators (SERMs) reduces breast cancer risk by 30-50% in study populations comprised of either premenopausal or postmenopausal women (7-13). Data now emerging indicate that aromatase inhibitors, which block metabolism of androgen precursors to estrogens, are highly effective in preventing development of breast cancer in postmenopausal women (14). Although the association between estrogens and breast cancer is clear, the mechanisms through which estrogens contribute to development of breast cancer remain poorly defined but may include estrogen receptor-alpha (ERα)-mediated events and oxidative stress/damage resulting from estrogen metabolism (3).

A family history of breast cancer is strongly associated with increased risk of this disease, and the extent to which familial risk is elevated increases as a function of the number of affected first degree relatives and is inversely related to the age at which breast cancer is diagnosed in
those relatives (15). Monozygotic twins are more concordant with respect to a diagnosis of breast cancer when compared to dizygotic twins or non-twin siblings (16, 17). Together, these data indicate that shared genetic variants contribute to breast cancer risk independently of shared environment. Several genetic determinants of breast cancer risk have been identified, including mutant alleles of well-studied tumor suppressor genes such as BRCA1, BRCA2, TP53 and PTEN, which act as highly or moderately penetrant determinants of breast cancer risk in individuals inheriting those alleles (18, 19). However, because the frequency of mutant alleles of these genes in the general population is very low, these genes account for only approximately 25% of familial relative risk and less than 10% of all breast cancer cases. More recently, multiple loci have been identified in genome wide association studies (GWAS) that harbor as yet unidentified genetic determinants of breast cancer risk (20-26). The main effects of the individual GWAS loci on breast cancer risk are small, and together they are estimated to explain approximately 8% of familial relative risk (19). Therefore, it is probable that many genetic determinants of breast cancer risk remain to be identified.

We are using the ACI rat model of 17β-estradiol (E2)-induced mammary cancer to define the mechanisms through which estrogens contribute to breast cancer development; identify and functionally characterize the genetic variants that determine susceptibility; and define the hormone-gene-environment interactions that influence development of mammary cancer in this physiologically relevant rat model. Female ACI rats are uniquely susceptible to mammary cancer when treated continuously with physiologic levels of E2 (27, 28). Induction of mammary cancer in female ACI rats occurs through a mechanism that is largely dependent upon ERα (29, 30). Interestingly, progesterone is also required for development of mammary cancer in E2
treated ACI rats (27, 31). The mammary cancers that develop in E2 treated ACI rats are estrogen dependent and share multiple features in common with luminal-type breast cancers in humans, including chromosome copy number aberrations, expression of ERα and expression of progesterone receptor (Pgr) (32-34). Interval mapping analyses of progeny generated in intercrosses between susceptible ACI rats and resistant Copenhagen (COP) or Brown Norway (BN) rats revealed nine quantitative trait loci (QTL), designated Emca1 (Estrogen-induced mammary cancer) through Emca9, each of which harbors one or more genetic determinants of mammary cancer susceptibility (35-38). Interestingly, a subset of the loci identified in GWAS as determinants of breast cancer risk in humans appear to be orthologous to the previously mapped Emca loci, suggesting that these genetic determinants of breast cancer risk may be shared across species. The purpose of this study was to develop and employ congenic rat strains in a substitution mapping approach to confirm the existence of and to localize more precisely the genetic determinants of mammary cancer susceptibility that reside within Emca8, which was mapped to rat chromosome 5 (RNO5) in intercrosses between the ACI and BN rat strains (37). Multiple congenic strains, each of which carries a distinct segment of RNO5 from the resistant BN strain on the genetic background of the susceptible ACI strain, were developed and characterized to further map Emca8 and define its effects on mammary cancer phenotypes. Regions of RNO5 within the Emca8 locus that differ in copy number between the ACI and BN genomes were identified using array comparative genomic hybridization (aCGH). Genes that reside within Emca8 and were differentially expressed between E2 treated ACI and ACI.BN-Emca8 congenic rats were identified as Emca8 candidates. Data presented herein indicate that at least three distinct genetic determinants of mammary cancer susceptibility reside within Emca8.
Materials and Methods

Care, Treatment and Phenotypic Characterization of Animals. The Institutional Animal Care and Use Committee of the University of Nebraska Medical Center approved all procedures involving live animals. ACI and BN rats were obtained from Harlan Sprague Dawley, Inc. (Indianapolis, IN). The congenic rat strains discussed herein were generated as described below. The animals were housed in a barrier facility under controlled temperature, humidity, and 12-hour light/12-hour dark conditions. This facility was accredited by the American Association for Accreditation of Laboratory Animal Care and operated in accordance with the standards outlined in the Guide for the Care and Use of Laboratory Animals (DHHS Pub. 85-23). All procedures relating to care, propagation and treatment of the experimental animals have been described previously (27, 28, 32, 36, 37). Beginning 7 weeks after initiation of E2 treatment, each rat was examined once or twice weekly for the presence of palpable mammary tumors. The location and size of each palpable tumor were noted at each examination. The rats were killed by decapitation when the largest palpable mammary tumor reached approximately 2.0 cm in its largest dimension, if necessitated by treatment-related morbidity, or following 196 ± 4 days of E2 treatment. Mammary tumor number, size and location were recorded at necropsy. A portion of each tumor was subjected to histological examination to confirm the presence of mammary carcinoma. To evaluate the impact of Emca8 on gene expression profiles, another set of animals was euthanized following 12 weeks of E2 treatment and the mammary glands were harvested, frozen in liquid nitrogen and stored at -80° C.
Genotyping. Genomic DNA was isolated from tail clips or ear punches using DNeasy columns according to manufacturer’s protocol (QIAGEN, Valencia, CA). Genetic markers that are polymorphic between ACI and BN were selected from the SSLP database in the Rat Genome Database. Oligonucleotide primers for genotyping were obtained from Invitrogen (Carlsbad, CA). When necessary, additional SSLP markers for genotyping specific genome regions were developed in our laboratory using the published rat genome sequence (Version 3.4). Information on the primers used to genotype these markers is available in the Rat Genome Database. Genotyping was carried out by PCR amplification and agarose gel electrophoresis as previously described (36, 37, 39).

Generation and Characterization of Congenic Rat Strains. The ACI.BN-D5Mgh17:D5Mgh15 congenic line (abbreviated ACI.BN-Emca8 or Emca8 line) was generated using a marker assisted selection protocol as described previously (40). Briefly, BN female rats were crossed to ACI male rats to generate F1 progeny. The F1 males were then back crossed to ACI females to produce the N2 generation. Males from the N2 generation were genotyped at a panel of markers distributed across the Emca8 locus on RNO5 (Supplemental Table 1, Tier 1). Animals heterozygous across Emca8 were genotyped at a panel of 42 markers distributed across the other autosomes to identify the rat that carried the fewest BN alleles at these background markers (Supplemental Table 1, Tier 2). This male was mated to ACI females to generate N3 progeny, and the marker assisted selection protocol was repeated to generate N4 and subsequent generations of back cross progeny. A male was obtained at the N4 generation that was heterozygous for BN alleles across Emca8 but homozygous for ACI alleles at all 42 Tier 2 background markers used for the negative selection. Genotyping at 16 additional background
markers was initiated at the N5 generation (Supplemental Table 1, Tier 3). At this generation, a male was identified that remained heterozygous across Emca8, but was homozygous for ACI alleles at all 58 background markers. Heterozygous N6 progeny of this male were intercrossed to produce congenic rats that were homozygous for BN alleles from D5Mgh17 through D5Mgh15. Additionally, male rats that were heterozygous across this interval were backcrossed to ACI females to generate rats harboring recombinations within the region of interest on RNO5. Animals harboring recombinant chromosomes were again backcrossed to ACI to generate male and female siblings carrying the same recombinant chromosome and these siblings were intercrossed to generate eight additional congenic lines for fine mapping (Table 1). Ultimately, the genetic background of each of these congenic lines was confirmed to be that of the recipient ACI strain by genotyping at a panel of 121 markers, 80 of which were unique from those used for negative selection, distributed across all autosomes other than RNO5 (Supplemental Table 1, Tier 4). Nomenclature of the resultant congenic strains is in accordance with the guidelines of the International Rat Genome Nomenclature Committee; the recipient strain is indicated first followed by the donor strain and the markers that flank the region known to be homozygous for the donor strain, followed lastly by the ILAR laboratory code designation (Table 1). Female rats from each of the congenic lines were treated with E2 as described above to evaluate their susceptibility to mammary cancer. Groups of ACI rats were evaluated contemporaneously.

**Identification of Copy Number Variation between the ACI and BN Genomes.**

NimbleGen (Madison, WI) high density oligonucleotide arrays were utilized in aCGH experiments to catalog copy number variants (CNV) between the susceptible ACI rat strain and the resistant BN rat strain. These arrays cover the nonrepetitive portion of the rat genome at an
average density of one probe per 5 kb. DNAs were isolated from ACI and BN rats using DNeasy columns (QIAGEN, Valencia, CA). Two independent experiments were performed; DNA from spleen was used in the first experiment and DNA from liver of different individual rats was used in the second. Sample labeling, hybridization and data collection were performed by NimbleGen using conditions optimized to their platform. Duplicate labeling reactions and hybridizations were performed in each experiment, and the Cy3 and Cy5 dye labels were reversed in the technical duplicates to eliminate potential bias relating from dye labeling. The resulting data were analyzed using SignalMap software (NimbleGen). Select CNVs were confirmed using PCR. The aCGH data generated in these analyses comprised part of a larger study in which multiple approaches were used to catalog CNVs in the rat genome (41). These data have been deposited in the EMBL ArrayExpress repository under accession number E-MEXP-1517.

**Evaluation of Gene Expression in the Mammary Gland.** Absolutely RNA Miniprep Kits (Stratagene, La Jolla, CA) were used to isolate RNA from mammary tissues that were dissected from individual ACI and Emca8 rats (n = 5 each strain) that had been treated with E2 for 12 weeks, a time point that precedes development of palpable mammary cancer. The RNAs were labeled using One-cycle Target Labeling Kits (Affymetrix, Santa Clara, CA) following the manufacturer’s specifications. Briefly, 2.5 µg of total RNA was used to generate labeled cRNA that was then fragmented and hybridized to Affymetrix Rat Genome 230 version 2.0 GeneChips. All RNA labeling and microarray hybridizations were performed by the Microarray Core Facility at the University of Nebraska Medical Center. Assessment of data quality parameters was performed using Gene Chip Operating System (GCOS, Affymetrix) following the
manufacturer’s recommendations. The primary microarray data have been deposited in the Gene Expression Omnibus under accession numbers GSE39788. The data from the ten microarrays described herein were normalized as part of a larger set of arrays using the Robust Multiarray Average (RMA) algorithm with quantile normalization (gcrma function in Bioconductor gcrma library). The five arrays for each strain were treated as a set to make pairwise comparisons between the two rat strains. P-values for each gene in each comparison were calculated using an empirical Bayes moderated t-statistic method (eBayes function in the Bioconductor package limma) (42). Corrections for multiple hypothesis testing were made using Benjamini-Hochberg testing with a false discovery rate of 5% (43). Ontologizer 2.0 was used to identify over represented ontology terms associated with the differentially expressed genes identified in the ACI vs Emca8 comparison that exceeded the 5% false discovery rate (44). The analysis was conducted using Parent-Child Union with Benjamini-Hochberg multiple testing correction. The baseline gene population was obtained from the Affymetrix microarray annotation file (Rat230_2.na29.annot.csv). The ontology definition (.obo) and rat GO association file (.rgd) files used by Ontologizer were downloaded from NCBI on July 29, 2011.

Differential expression of three genes residing within Emca8 was confirmed by quantitative reverse transcription-polymerase chain reaction. Total RNA was isolated from mammary glands of five ACI and five Emca8 rats. Reverse transcription was performed using SuperScript™ III First-Strand Synthesis System (Invitrogen, Carlsbad, CA) and 900 ng total mammary RNA. Real time quantitative PCR (qPCR) was performed and the results analyzed using the ABI 7500 PCR System (Applied Biosystems, Inc., Forster City, CA). Primer and probe sets for Tspan1 (Rn01416779_m1), Cd52 (Rn02132983_s1) and Wnt4 (Rn00584577_m1) were
obtained from ABI Assays on Demand service (Carlsbad, CA). Standard curves were generated using cDNA reverse transcribed from lung RNA ($r^2 > 0.95$) and were utilized to calculate relative gene expression levels. Expression of Gapdh (4352338E, ABI) in each sample was analyzed simultaneously. Expression of each of the three genes was normalized to that of Gapdh.

**Statistical Analysis of Data.** Differences in latency to appearance of first palpable mammary tumor between strains were evaluated using the Logrank test (Mstat, v.5.4, University of Wisconsin, Madison, USA). The Wilcoxon rank-sum test was used to evaluate differences in the mean tumor number per rat (Mstat, v.5.4). Differences in gene expression were assessed using the Kruskal-Wallis ANOVA with Dunn’s post hoc tests (GraphPad Prism, San Diego California, USA). P values ≤ 0.05 were considered to be indicative of statistical significance.

**Results**

**ACI.BN-Emca8 congenic rats exhibit reduced susceptibility to E2-induced mammary cancer relative to ACI rats.** Emca8, a QTL on RNO5 for which the 95% confidence interval extends from $D5Rat134$ (52.43 Mb) to $D5Rat37$ (148.46 Mb), was identified by interval mapping analyses of progeny generated in a BN x ACI intercross (37). To confirm the existence of Emca8, we generated the ACI.BN-Emca8 congenic rat strain, which harbors BN alleles from $D5Mgh17$ (14.73 Mb) to $D5Mgh15$ (167.85 Mb) on the ACI genetic background, and characterized the susceptibility of this congenic rat strain to E2-induced mammary cancer. A group of 41 female ACI.BN-Emca8 rats was treated with E2 beginning at 9 weeks of age. The
median latency to appearance of the first palpable mammary tumor in these animals was 188
days, and the incidence of mammary cancer at the completion of the 28 week course of treatment
was 63% (Figure 1). The mean latency for those animals that developed grossly apparent
mammary cancer was 172 ± 21 days (Table 2). The number of mammary tumors per rat ranged
from 0 to 4, and the average tumor number per rat at necropsy was 1.4 ± 1.2. Age matched,
oviduct intact, female ACI.BN-Emca8 rats that received empty implants did not develop mammary
cancer within the 28 week course of this experiment.

A group of 16 E2 treated female ACI rats was evaluated contemporaneously with the
ACI.BN-Emca8 congenic rats. The median latency in this group of ACI rats was 149 days,
100% of the animals at risk developed mammary cancer by 196 days of treatment, and the
average number of mammary tumors observed at necropsy was 3.7 ± 3.2. The mammary cancer
phenotypes exhibited by this group of ACI rats did not differ significantly from other groups of
ACI rats evaluated during characterization of additional congenic rat strains described below.
Therefore, the individual groups of E2 treated ACI rats were merged into a single population (n
= 128) for further statistical analyses. Median latency to appearance of the first palpable
mammary tumor in the merged population of E2 treated ACI rats was 147 days, incidence in the
animals at risk was 94%, and the average number of tumors observed at necropsy was 5.3 ± 4.5
(Table 2). Relative to both the smaller and merged groups of ACI rats, the ACI.BN-Emca8
congenic rats exhibited significantly reduced susceptibility to E2-induced mammary cancer.
Latency to appearance of the first palpable mammary cancer was prolonged, mammary cancer
incidence was reduced and tumor number per rat was lower in the E2 treated ACI.BN-Emca8
congenic rats, relative to the treated ACI rats (Figure 1 and Table 2). These data confirm the
existence of Emca8 as a QTL on RNO5 that harbors at least one genetic determinant of susceptibility to E2-induced mammary cancer.

**Emca8 harbors multiple genetic determinants of mammary cancer susceptibility.**

Published interval mapping data revealed multiple local LOD peaks within the Emca8 95% confidence interval, suggesting that multiple determinants of mammary cancer susceptibility may reside on RNO5 (37). To localize more precisely these genetic determinants of mammary cancer susceptibility, we generated three additional congenic strains, each of which harbors BN alleles across an overlapping segment of the larger Emca8 congenic interval on the ACI genetic background (Table 1). The ACI.BN-Emca8a congenic strain harbors BN alleles from D5Mgh17 (14.73 Mb) to D5Rat98 (100.02 Mb); ACI.BN-Emca8b harbors BN alleles from D5Rat113 (76.71 Mb) to D5Rat36 (145.19 Mb); and ACI.BN-Emca8c harbors BN alleles from D5Uwm70 (128.99 Mb) to D5Mgh15 (167.85 Mb). Each of these congenic rat strains exhibited significantly reduced susceptibility to E2-induced mammary cancer when compared to E2 treated ACI rats (Figure 2 and Table 2). Latency to appearance of the first palpable mammary cancer was increased, the incidence of mammary cancer was decreased and the average tumor number per rat was lower in E2 treated ACI.BN-Emca8a, ACI.BN-Emca8b and ACI.BN-Emca8c rats, relative to treated ACI rats. Mammary cancer was not detected in sham treated, age matched, ovary intact, female congenic rats. These data demonstrate that RNO5 harbors a minimum of two physically distinct genetic determinants of susceptibility to E2-induced mammary cancer.
**Fine Mapping of Emca8.1 and Emca8.2.** Additional congenic strains were generated and characterized to fine map the genetic determinants of mammary cancer susceptibility residing within the *Emca8b* and *Emca8c* intervals. The ACI.BN-Emca8b1 and ACI.BN-Emca8b3 rat strains harbor BN alleles across the proximal and distal portions of the *Emca8b* interval, respectively. Both of these congenic rat strains exhibited significantly reduced susceptibility to E2-induced mammary cancer when compared to ACI rats, revealed by prolonged latency to appearance of palpable mammary cancer and reduced tumor number (Figure 3 and Table 2). The ACI.BN-Emca8d congenic strain was generated specifically to evaluate the impact of genetic variants within or near several candidate genes known to function in cell cycle regulation or mammary gland development, including *Cdkn2a, Cdkn2b, Nfia* and *Jun*. Although latency to appearance of palpable mammary cancer in E2 treated ACI.BN-Emca8d congenic rats did not differ from that observed in ACI rats, the number of mammary cancers observed at necropsy was significantly lower in ACI.BN-Emca8d congenic rats relative to ACI rats (Figure 3 and Table 2). Age matched, sham treated, ACI.BN-Emca8b1, ACI.BN-Emca8b3 and ACI.BN-Emca8d congenic rats did not develop mammary cancer. Together, these data localize a genetic determinant of mammary cancer susceptibility to a 14.6 Mb interval defined by *D5Rat151* (104.23 Mb) and *D5Rat25* (118.81 Mb) and designated *Emca8.1*.

The ACI.BN-Emca8c1 and ACI.BN-Emca8c2 congenic strains were generated to further evaluate the proximal portion of the *Ecma8c* interval, the region of RNO5 that generated the peak LOD score during interval mapping of *Emca8* (37). When treated with E2, ACI.BN-Emca8c1 rats developed mammary cancer in a manner that was indistinguishable from that of E2 treated ACI rats (Figure 4 and Table 2). By contrast, the ACI.BN-Emca8c2 congenic strain
exhibited dramatically reduced susceptibility to E2-induced mammary cancer. Latency to appearance of palpable mammary cancer was significantly prolonged and both the incidence of mammary cancer and tumor number per rat were significantly reduced in E2 treated ACI.BN-Emca8c2 rats, relative to treated ACI rats. Age matched, sham treated, ACI.BN-Emca8c1 and ACI.BN-Emca8c2 congenic rats did not develop mammary cancer. These data localize a second genetic determinant of mammary cancer susceptibility to a 10 Mb interval on RNO5 defined by markers D5Got42 (130.75 Mb) and D5Rat86 (140.74 Mb), designated Emca8.2.

**Histopathology of mammary tumors induced by 17β-estradiol.** The histological phenotypes of the mammary tumors that developed in each of the congenic rat strains in response to E2 treatment were similar to those that developed in ACI rats. All of the E2-induced tumors exhibited features of mammary carcinoma, including enlarged and pleomorphic nuclei with prominent nucleoli. Comedo features were frequently apparent. Although most of the tumors remained *in situ*, some were locally invasive and associated with a desmoplastic response. Overall, the tumors observed in this study did not differ discernibly from those described in previous studies (27, 32).

**Emca8.1 and Emca8.2 harbor copy number variant regions.** The genomes of ACI and BN rats were compared by aCGH using NimbleGen tiling path arrays to identify regions that vary in copy number between these inbred rat strains. These analyses revealed a total of 53 CNVs (data not shown), two of which reside within Emca8 on RNO5 (Figure 5). The first of these CNVs extends over an approximate 100 kb region located within the Emca8.1 interval and harbors four genes that encode members of the interferon alpha family of cytokines, including
*Ifna2*. The second CNV extends over approximately 500 kb within the *Emca8.2* interval. This CNV harbors 2 genes that encode members of the butyrophilin gene family and four pseudogenes. PCR analyses confirmed the existence of both of these CNVs and suggested the ACI genome lacks both of these genome segments when compared to the reference BN genome (data not shown). Interestingly, a search of the Database of Genomic Variants indicated the human genome similarly varies in copy number in the regions that are orthologous to these rat CNVs. Because the NimbleGen oligonucleotide array was constructed using the BN rat genome reference sequence (version 3.4) as the template, DNA segments that are present in the ACI genome but absent in the BN genome were not detectable using this platform.

**Impact of *Emca8* on gene expression in the mammary gland.** To define the impact of genetic variants residing within *Emca8* on gene expression in the mammary gland, microarray analyses were performed on RNA isolated from grossly normal mammary glands of ACI and ACI.BN-Emca8 rats that were treated with E2 for 12 weeks, a time point that generally precedes appearance of palpable mammary cancer. Transcripts corresponding to 40 probe sets were differentially expressed between mammary glands of the E2 treated ACI and ACI.BN-Emca8 rats (Supplemental Table 2). The differentially expressed genes include *Pgr* (progesterone receptor) and *Wnt4* (wingless-type MMTV integration site family, member 4), both of which are required for mammary gland development and were expressed at a higher level in ACI rats relative to ACI.BN-Emca8 congenic rats; and *Spp1* (secreted phosphoprotein 1, also known as osteopontin), which is secreted by the mammary gland as a component of milk and extracellular matrix and was expressed at a lower level in the mammary gland of E2 treated ACI rats compared to glands of ACI.BN-Emca8 rats. Gene ontology enrichment analyses of the
differentially expressed transcripts revealed a statistically significant association with the biological process tertiary branching involved in mammary gland duct morphogenesis. As expected, a disproportionately high fraction, 16 of 40 (40%), of the differentially expressed transcripts arise from genes that reside within the Emca8 congenic interval, including Tspan1, a putative mammary tumor suppressor gene that resides within Emca8.2 (45); Cd52, a protein expressed on the surface of lymphocytes and some epithelial cell types; Scl30a2, which encodes a zinc transporter and is regulated by prolactin-prolactin receptor signaling (46); and Wnt4, which functions downstream of progesterone-progesterone receptor in the regulation of ductal branching (47). Quantitative real-time PCR (qRT-PCR) was performed to confirm differential expression of a subset of the Emca8 candidate genes, Wnt4, Cd52 and Tspan1 (Supplemental Figure 1). The data from these analyses confirm the microarray data and indicate that each of these genes was expressed at a significantly lower level in the mammary glands of E2 treated ACI.BN-Emca8 congenic rats relative to glands of E2 treated ACI rats.

Discussion

Data presented herein confirm and extend a published interval mapping study of phenotypically defined (BNxACI)F2 rats that localized to RNO5 one or more genetic determinants of susceptibility to E2-induced mammary cancer (37). By generating and characterizing a panel of novel congenic rat strains, we have now demonstrated that at least three physically distinct genetic determinants of mammary cancer susceptibility reside on this chromosome. The first of these genetic determinants resides within the approximate 95 Mb segment of proximal RNO5 defined by markers D5Rat121 and D5Rat151; the second,
designated \textit{Emca8.1}, resides within the 14.6 Mb segment defined by markers \textit{D5Rat151} and \textit{D5Rat25}; and the third, designated \textit{Emca8.2}, resides within the 10 Mb segment of RNO5 defined by markers \textit{D5Got42} and \textit{D5Rat86} (Supplemental Figure 2).

Other studies have also genetically associated RNO5 with susceptibility to E2-induced mammary cancer. Interval mapping analyses of F2 progeny from reciprocal intercrosses between susceptible ACI and resistant COP rats localized \textit{Emca1} to distal RNO5 (36). Moreover, loss of RNO5 was the most frequently observed somatic genetic abnormality when a panel of mammary cancers induced by E2 in ACI rats was evaluated by comparative genomic hybridization (33). Together, these data suggest that loss of function of one or more tumor suppressor genes residing on RNO5 may contribute to mammary cancer development in this rat model. Loci on RNO5 have also been mapped as genetic determinants of susceptibility to mammary cancer induced by 7,12-dimethylbenz[a]anthracene (DMBA) in crosses between susceptible WF and resistant WKY rats (48-51), as well as in crosses between susceptible SPRD-Cu3 and WKY rats (52, 53). It is also noteworthy that mapping studies performed using multiple transgenic and knockout mouse models identified mammary cancer QTL on segments of mouse chromosome 4 that are orthologous to RNO5 (54-56). At this time, only one of these rodent mammary cancer QTL, \textit{Mcs5a}, has been defined to the nucleotide level (51). The region to which \textit{Mcs5a} was mapped resides within the \textit{Emca8a} congenic interval on RNO5 but is distinct from the \textit{Emca8.1} and \textit{Emca8.2} loci identified in this study.

\textit{Emca8.1} harbor over 60 annotated genes, including the tumor suppressor genes \textit{Cdkn2a}, which encodes p16\textsuperscript{Ink4a} and p19\textsuperscript{Arf} (p14\textsuperscript{Arf} in humans), and \textit{Cdkn2b}, which encodes p15\textsuperscript{Ink4b}. Also
residing within *Emca8.1* are the proto-oncogenes *Nfia* and *Jun*. Although the protein products of each of these genes play well defined roles in regulation of cell proliferation and/or mammary gland development, none of these genes have been strongly and directly implicated in breast cancer etiology. *Emca8.1* is orthologous to the 9p21 region of the human genome. Single nucleotide polymorphisms (SNPs) residing at 9p21, either within the intergenic region between *Cdkn2a* and *Cdkn2b* or upstream of these genes, have been associated with breast cancer risk in genome wide (GWAS) and candidate gene association studies (57, 58). Other SNPs within this same region have been associated with an assortment of other complex diseases, including melanoma, atherosclerosis and type 2 diabetes. A subset of these disease associated SNPs have been associated with the level of expression of *Cdkn2a*<sup>INK4a</sup>, *Cdkn2a*<sup>Arf</sup>, *Cdkn2b* and/or *ANRIL*, a long non-coding RNA that has been implicated in epigenetic regulation of the *Ink4* locus (59-61). Together, these data suggest that inherited risk of these diseases is determined by cis-acting genetic variants that impact expression of one or more of the genes residing within this region of the genome. Supporting this model is the observation that targeted deletion of the 70 kb interval of the mouse genome that is orthologous to the 9p21 risk locus dramatically decreased expression of *Cdkn2a* and *Cdkn2b* (62). Interestingly, this deletion also increased susceptibility to specific cancer types (62). Somatic copy number aberrations at 9p21 have also been associated with breast cancer etiology. A recent study of over 2000 breast cancers suggested that heterozygous or homozygous deletions at 9p21 affecting *MTAP*, a putative breast cancer suppressor gene, and *Cdkn2a* are driver events in breast cancer development (63). We are further exploring this apparent conservation of a breast cancer risk locus between humans and rats in comparative genetics and genomics based studies to identify the causal genetic variants and define the organ site and mechanism through which they impact breast cancer development.
Emca8.2 harbors nearly 150 annotated genes, including 124 protein coding genes, and is orthologous to the 1p32 region of the human genome. To our knowledge, no evidence exists to associate germ line variants at 1p32 with breast cancer risk. However, deletions at 1p32 have been reported to occur relatively frequently in breast cancers (64-66). In parallel studies, we have employed a congenic substitution approach to localize Emca1, a genetic determinant of susceptibility to E2-induced mammary cancer mapped to distal RNO5 in crosses between susceptible ACI rats and resistant COP rats (36), to the same region of RNO5 as Emca8.2, suggesting the BN and COP strains may share resistance conferring alleles at this locus (unpublished data). Among the candidate genes residing with Emca8.2 is Cdkn2c, which encodes the p18Ink4c tumor suppressor protein. Mice that are homozygous for mutant alleles of Cdkn2c spontaneously develop tumors in the intermediate lobe of the pituitary gland (67). Moreover, Cdkn2c<sup>-/-</sup> homozygotes and Cdkn2c<sup>+/+</sup> heterozygotes exhibit increased susceptibility, relative to Cdkn2c<sup>++/+</sup> littermates, to carcinogen-induced pituitary tumors and several other tumor types, indicating that haploinsufficiency of Cdkn2c predisposes to tumor development (68).

Interestingly, Cdkn2c mutations, when harbored on the Balb/c genetic background, predispose to development of luminal type mammary cancers and dysregulation of homeostasis within specific mammary epithelial cell populations (69), a phenotype that was not observed when the mutant Cdkn2c allele was harbored on the C57Bl/6 background.

An intriguing observation from this study is that Emca8.1 and Emca8.2 each contain a cluster of syntenic genes that appears to have arisen via duplication of an ancestral genome segment: Cdkn2b, Dmrt1 and Elavl2 within Emca8.1 and Cdkn2c, Dmrt2 and Elavl4 within
The Cdkn2 proteins regulate cell cycle progression via inhibition of the cyclin D dependent kinases Cdk4 and Cdk6. Because aberrant cyclin D signaling has been implicated in breast cancer etiology, we sequenced the ACI alleles of Cdkn2a, Cdkn2b and Cdkn2c over their respective protein coding regions. The Cdkn2a coding regions for both p16^{Ink4a} and p19^{Arf} were identical between the ACI and BN rat strains. A coding region SNP was identified in the ACI alleles of both Cdkn2b and Cdkn2c, relative to their respective reference BN alleles (Supplemental Table 3). However, neither SNP resulted in a change in amino acid sequence, strongly suggesting that these SNPs are not functionally significant with respect to mammary cancer susceptibility.

Both Emca8.1 and Emca8.2 were demonstrated to harbor CNVs. It is estimated that 5-20% of the genome is variant with respect to copy number within the human population (70). The distribution and functional significance of CNVs in rat appear similar to the human (41).

Although it is clear that CNVs dramatically impact gene expression and influence risk of specific complex diseases, the impact of CNVs on cancer risk, in particular breast cancer risk, is not well defined. One recent study of 68 cases of familial or early onset breast cancer revealed a total of 26 CNVs that occurred at least once in cases but not in two different control populations (71). A second recent study demonstrated that the percentage of the germline genome that varies in copy number is not associated with the extent to which the somatic genome is aberrant in copy number in a panel of 28 breast cancers (72). The CNVs residing within Emca8.1 and Emca8.2 harbor genes that encode modulators of the immune system, and the human orthologs to both of these genome segments also exhibit variation in copy number, suggesting that these CNVs may determine breast cancer risk via regulation of immune surveillance. The impact of the CNVs...
residing within Emca8.1 and Emca8.2 on susceptibility to E2-induced mammary cancer is currently under investigation.

The mechanisms and sites of action of the Emca8 genetic determinants of mammary cancer susceptibility are not known. Two observations suggest one or more of these determinants may function within the mammary epithelium. The first is that mammary cancers induced in ACI rats by E2 frequently exhibit loss of RNO5, suggesting that loss of function of one or more tumor suppressor genes on RNO5 contributes to mammary cancer development (33). The second is that the pattern of chromosome copy number changes exhibited by tumors induced in Emca8 congenic rats differs from that exhibited by tumors in ACI rats, suggesting that introgression of BN alleles across the Emca8 interval alters the pattern of somatic genetic events that occur during mammary cancer development (33). If the actions of the Emca8 determinants of susceptibility are mammary autonomous, then identification of those genes that are differentially expressed between the mammary glands of E2 treated ACI and Emca8 rats and reside within the Emca8 congenic interval would reveal Emca8 candidate genes whose expression is affected by genetic variants that reside within cis-acting regulatory elements. The genes that encode 40 percent of the differentially expressed transcripts identified in this experiment reside within the Emca8 congenic interval, including Cyp2j10, which resides in Emca8.1, and Tspan1, which resides in Emca8.2, and warrant further study to evaluate their potential role in mammary cancer development.

The ACI rat model of E2-induced mammary cancer exhibits many features in common with luminal type breast cancers in humans and is gaining wide use in the breast cancer research
community. The current study reveals the presence of at least three distinct genetic determinants of susceptibility to E2-induced mammary cancer on RNO5 (Supplemental Figure 2) and further illustrates the relevance of this model to human breast cancer by identifying Emca8.1 as a susceptibility locus that may be the functional ortholog to a genetic determinant of breast cancer risk in humans. Experiments are underway to further localize Emca8.1 and Emca8.2 to identify the causal genetic variants responsible for these two QTL and to define their sites and mechanisms of action.

Acknowledgements. The authors thank Maureen Hickman, Nyssa Becker, Dr. Wade Junker, Dr. Jim Eudy, Jason Huerta, Lynette Smith, Dr. Martin Tochack, Dr. Linda Buckles and Dr. Karen Gould for their contributions to this study.

Author contributions. Data generation and analysis: BS, KLW, SK, JC, NS, CW, JS. Manuscript preparation: BS, JC, JS. Study direction and oversight: JS.
Reference List


Table 1. Genetic Characteristics of Congenic Strains.

<table>
<thead>
<tr>
<th>Strain Designation</th>
<th>Strain Abbreviations</th>
<th>RGD ID</th>
<th>Proximal</th>
<th>Distal</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACL.BN-(D5Mgh17- D5Mgh15)/Shul</td>
<td>ACL.BN-Emca8a or Emca8b</td>
<td>5135479</td>
<td>D5Rat104</td>
<td>72.36 Mb D5Rat113</td>
</tr>
<tr>
<td>ACL.BN-(D5Rat372- D5Rat36)/Shul</td>
<td>ACL.BN-Emca8b or Emca8b3</td>
<td>5509086</td>
<td>D5Rat26</td>
<td>119.33 Mb D5Rat72</td>
</tr>
<tr>
<td>ACL.BN-(D5Uwm70- D5Mgh15)/Shul</td>
<td>ACL.BN-Emca8c or Emca8c</td>
<td>5135477</td>
<td>D5Uwm69</td>
<td>127.53 Mb D5Uwm70</td>
</tr>
<tr>
<td>ACL.BN-(D5Uwm70- D5Got42)/Shul</td>
<td>ACL.BN-Emca8c1 or Emca8c1</td>
<td>5135473</td>
<td>D5Uwm69</td>
<td>127.53 Mb D5Uwm70</td>
</tr>
<tr>
<td>ACL.BN-(D5Uwm70- D5Rat32)/Shul</td>
<td>ACL.BN-Emca8c2 or Emca8c2</td>
<td>5135472</td>
<td>D5Uwm69</td>
<td>127.53 Mb D5Uwm70</td>
</tr>
<tr>
<td>ACL.BN-(D5Rat60- D5Rat115)/Shul</td>
<td>ACL.BN-Emca8d or Emca8d</td>
<td>5135475</td>
<td>D5Rat131</td>
<td>104.23 Mb D5Rat60</td>
</tr>
</tbody>
</table>

1 Rat Genome Database.
Table 2. Mammary Cancer Phenotypes.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Median Latency (days)</th>
<th>P vs ACI</th>
<th>Hazard Ratio</th>
<th>Mean Latency (days)</th>
<th>Incidence at 196 days</th>
<th>Tumor Number per Rat</th>
<th>P vs ACI</th>
<th>N³</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACI</td>
<td>147</td>
<td>-</td>
<td>1.000</td>
<td>140 ± 27</td>
<td>94%</td>
<td>5.3 ± 4.5</td>
<td>-</td>
<td>128 (84)</td>
</tr>
<tr>
<td>ACI.BN-Emca8⁵</td>
<td>188</td>
<td>1.8 x 10⁻³</td>
<td>0.358</td>
<td>172 ± 21</td>
<td>63%</td>
<td>1.4 ± 1.2</td>
<td>4.1 x 10⁻⁴</td>
<td>41 (38)</td>
</tr>
<tr>
<td>ACI.BN-Emca8a</td>
<td>191</td>
<td>3.1 x 10⁻³</td>
<td>0.411</td>
<td>162 ± 29</td>
<td>68%</td>
<td>2.1 ± 1.9</td>
<td>5.4 x 10⁻³</td>
<td>30 (29)</td>
</tr>
<tr>
<td>ACI.BN-Emca8b</td>
<td>174</td>
<td>1.8 x 10⁻⁴</td>
<td>0.463</td>
<td>155 ± 26</td>
<td>76%</td>
<td>1.7 ± 2.0</td>
<td>2.2 x 10⁻⁷</td>
<td>35 (34)</td>
</tr>
<tr>
<td>ACI.BN-Emca8c</td>
<td>196</td>
<td>1.5 x 10⁻³</td>
<td>0.291</td>
<td>170 ± 25</td>
<td>56%</td>
<td>0.8 ± 0.9</td>
<td>2.8 x 10⁻¹¹</td>
<td>35 (31)</td>
</tr>
<tr>
<td>ACI.BN-Emca8b1</td>
<td>157</td>
<td>1.5 x 10⁻¹</td>
<td>0.541</td>
<td>143 ± 20</td>
<td>63%</td>
<td>2.1 ± 2.6</td>
<td>2.1 x 10⁻¹</td>
<td>24 (14)</td>
</tr>
<tr>
<td>ACI.BN-Emca8b3</td>
<td>170</td>
<td>4.4 x 10⁻³</td>
<td>0.411</td>
<td>154 ± 29</td>
<td>66%</td>
<td>1.2 ± 1.2</td>
<td>6.2 x 10⁻³</td>
<td>13 (11)</td>
</tr>
<tr>
<td>ACI.BN-Emca8d</td>
<td>155</td>
<td>0.0827</td>
<td>0.730</td>
<td>145 ± 18</td>
<td>100%</td>
<td>3.0 ± 2.9</td>
<td>0.0143</td>
<td>33 (18)</td>
</tr>
<tr>
<td>ACI.BN-Emca8c1</td>
<td>153</td>
<td>0.1196</td>
<td>0.781</td>
<td>143 ± 25</td>
<td>86%</td>
<td>4.4 ± 3.8</td>
<td>0.3098</td>
<td>23 (14)</td>
</tr>
<tr>
<td>ACI.BN-Emca8c2</td>
<td>NA³</td>
<td>6.8 x 10⁻⁷</td>
<td>0.240</td>
<td>161 ± 31</td>
<td>42%</td>
<td>0.5 ± 0.7</td>
<td>6.9 x 10⁻⁷</td>
<td>20 (18)</td>
</tr>
</tbody>
</table>

1. Calculated using log rank test.
2. Calculated for tumor positive rats.
3. Calculated for population at risk.
4. Calculated for all rats that were treated with E2 for at least 160 days and less than 200 days.
5. Total number of rats treated with E2 (number treated with E2 for at least 160 days and less than 200 days).
6. Median latency exceeds 196 days.
Figure 1. The ACI.BN-Emca8 congenic rat strain exhibits reduced susceptibility to mammary cancer relative to the parental ACI rat strain. Ovary intact female ACI and ACI.BN-Emca8 rats were treated with E2 beginning at 9 weeks of age. Each animal was thereafter examined at least once per week for the presence of palpable mammary cancer. Latency to appearance of mammary cancer was prolonged in treated ACI.BN-Emca8 rats (n = 41) relative to contemporaneously treated ACI rats (n = 16) or a population of ACI rats (n = 112) merged from other experiments described herein. Vertical ticks on each line represent censored animals.

Figure 2. Emca8 harbors multiple genetic determinants of mammary cancer susceptibility. The Emca8a, Emca8b and Emca8c congenic rat strains harbor BN alleles across distinct overlapping segments of the larger Emca8 interval (Table 1). Ovary intact female Emca8a (n = 30), Emca8b (n = 35), Emca8c (n = 35), and ACI (n = 128) rats were treated with E2 beginning at 9 weeks of age and were thereafter examined at least once per week for the presence of palpable mammary cancer. Latency to appearance of mammary cancer was prolonged in the Emca8a, Emca8b and Emca8c rats relative to the ACI rats, indicating that two or more physically distinct genetic determinants of mammary cancer susceptibility reside on RNO5. Vertical ticks represent censored animals.

Figure 3. Fine mapping of Emca8.1. Ovary intact female Emca8b1 (n = 21), Emca8b3 (n = 13) and Emca8d (n = 33) rats were treated with E2 beginning at 9 weeks of age and were thereafter examined at least once per week for the presence of palpable mammary cancer. Latency to appearance of mammary cancer was significantly prolonged in the Emca8b1 and Emca8b3 rats, but not Emca8d rats, relative to the ACI rats. Vertical ticks represent censored animals.
Figure 4. Fine mapping of *Emca8.2*. Ovary intact female *Emca8c1* (n = 23) and *Emca8c2* (n = 20) rats were treated with E2 beginning at 9 weeks of age and were thereafter examined at least once per week for the presence of palpable mammary cancer. Latency to appearance of mammary cancer was significantly prolonged in the *Emca8c2*, but not *Emca8c1* rats, relative to the ACI rats. Vertical ticks represent censored animals.

Figure 5. aCGH analyses reveal CNVs in *Emca8.1* and *Emca8.2*. DNA was isolated from spleen (A. top panel) or liver (A. bottom panel) of two individual ACI and two individual BN rats. The ACI and BN DNA samples were labeled with Cy5 and Cy3, respectively, and were hybridized to NimbleGen tiling path oligonucleotide arrays. The dye labels were swapped in parallel experiments that are not illustrated to eliminate potential labeling bias. The CNVs residing within *Emca8.1* and *Emca8.2* are illustrated along the X axis relative to the genome coordinates along RNO5. The log2 transformed Cy3/Cy5 ratio is illustrated on the y-axis. Panel B illustrates the locations of annotated genes residing on the two DNA strands in and around the CNVs.
Figure 1
Figure 2

![Graph showing tumor progression over days of E2 treatment](image-url)
Figure 3
Figure 5
Cancer Prevention Research

Mapping of Three Genetic Determinants of Susceptibility to Estrogen-Induced Mammary Cancer within the Emca8 Locus on Rat Chromosome 5

Beverly S. Schaffer, Kristin M. Leland-Wavrin, Scott G. Kurz, et al.


Updated version  Access the most recent version of this article at: doi:10.1158/1940-6207.CAPR-12-0346-T

Supplementary Material  Access the most recent supplemental material at: http://cancerpreventionresearch.aacrjournals.org/content/suppl/2012/11/16/1940-6207.CAPR-12-0346-T.DC1

Author Manuscript  Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.