Physical confirmation and comparative genomics of the rat *Mammary carcinoma susceptibility 3* quantitative trait locus

Saasha Le*, Zachary C. Martin*, David J. Samuels* ‡§

*Department of Biochemistry and Molecular Genetics, University of Louisville School of Medicine, Louisville, Kentucky, United States of America, ‡James Graham Brown Cancer Center, University of Louisville School of Medicine, Louisville, Kentucky, United States of America, §Center for Genetics and Molecular Medicine, University of Louisville School of Medicine, Louisville, Kentucky, United States of America.

§Corresponding author. Email: david.samuelson@louisville.edu.

Abstract

Human breast and rat mammary cancer susceptibility are complex phenotypes where complete sets of risk associated loci remain to be identified for both species. We tested multiple congenic rat strains to physically confirm and positionally map rat *Mammary carcinoma susceptibility 3* (*Mcs3*), a mammary cancer resistance allele previously predicted at *Rattus norvegicus* chromosome 1 (*RNO1*). The mammary cancer susceptible Wistar Furth (WF) strain was the recipient, and the mammary cancer resistant Copenhagen (Cop) strain was the *RNO1*-segment donor for congenics. Inbred WF females averaged 6.3 carcinogen-induced mammary carcinomas per rat. Two WF.Cop congenic strains averaged 2.8 and 3.4 mammary carcinomas per rat, which confirmed *Mcs3* as an independently acting allele. Two other WF.Cop congenic strains averaged 6.6 and 8.1 mammary carcinomas per rat, and thus, did not contain *Mcs3*. Rat *Mcs3* was delimited to 27.8 Mb of *RNO1* from rs8149408 to rs105131702 (*RNO1*:143700228-171517317 of RGSC 6.0/rn6). Human genetic variants with p values for association to breast cancer risk below $10^{-7}$ had not been reported for *Mcs3* orthologous loci; however, human variants located in *Mcs3*-orthologous regions with potential association to risk ($10^{-7} < p < 10^{-3}$) were listed in some population-based studies. Further, rat *Mcs3* contains sequence orthologous to human 11q13/14, a region frequently amplified in female breast cancer. We conclude that *Mcs3* is an independently acting mammary carcinoma resistance allele. Human population-based, genome-targeted association studies interrogating *Mcs3* orthologous loci may yield novel breast cancer risk associated variants and genes.

Introduction

Excluding skin cancer, breast cancer is the most commonly diagnosed cancer among women in the United States, accounting for 29% of cancers diagnosed and 14% of cancer related deaths in females (Siegel et al. 2016). The current population-based lifetime-risk of a woman being diagnosed with breast cancer is 12-13% (Siegel et al. 2016). Breast cancer susceptibility is a complex trait controlled by genetic, epigenetic, and environmental factors. The genetic component of breast cancer susceptibility is comprised of a spectrum of high, moderate, and low penetrance risk associated mutations and variants (Pharaoh et al. 2002). Genetic linkage analyses and genome-wide association studies (GWAS) have enabled the identification of several risk alleles, but a majority of susceptibility loci remain unidentified (Smith et al. 2006; Ioannidis et al. 2010; Mavaddat et al. 2010; Economopoulou et al. 2015). Furthermore, GWAS have yielded many potentially-associated variants. These are polymorphisms with p values for association above a stringent genome-wide significance threshold level of $1 \times 10^{-7}$, which is
commonly used to avoid false-positive associations (THOMAS et al. 2005). Comparative genomics approaches that integrate relevant QTLs of experimental organisms and human loci, containing potentially-associated variants, may be useful to identify additional risk associated alleles. To serve this approach, mammary cancer susceptibility QTLs identified in experimental organisms must be mapped to conserved syntenic regions to effectively guide human genome-targeted association studies (LESUEUR et al. 2005; SAMUELSON et al. 2007).

Laboratory rats (Rattus norvegicus) provide a good experimental model of female breast cancer, as rat mammary carcinomas are similar to human breast carcinomas with respect to histopathology and hormone responsiveness (GOULD 1995). Evidence suggests that both rat mammary and human breast carcinomas originate from mammary ductal cells (RUSSO et al. 1983). Inbred rat strains that differ in susceptibility to 7,12-Dimethylbenz[a]anthracene (DMBA)-, N-Methyl-N-Nitrosourea (NMU)-, or estrogen-induced mammary carcinogenesis have been used to predict locations of Mammary carcinoma susceptibility (Mcs) and Estrogen-induced mammary cancer (Emca) loci (HSU et al. 1994; SHEPEL et al. 1998; LAN et al. 2001; GOULD et al. 2004; QUAN et al. 2006; SCHAFFER et al. 2006; REN et al. 2013). Linkage analyses using crosses between the DMBA-induced mammary carcinoma susceptible Wistar-Furth (WF) and resistant Copenhagen (Cop) strains resulted in four predicted Mcs QTLs named Mcs1, Mcs2, Mcs3, and Mcs4 (HSU et al. 1994; SHEPEL et al. 1998). Resistance associated QTLs, Mcs1 and Mcs2, have been physically confirmed (HAAG et al. 2003b; SANDERS et al. 2011). In this article we report our congenic strain results that physically confirm rat Mcs3 and delimit this Cop resistance QTL to a 27.8 Mb segment of rat chromosome 1 (RNO1).

Materials and Methods

Congenic Breeding and Genotyping

Congenic lines were made by adapting previously described methods to RNO1 (SAMUELSON et al. 2003). Briefly, rats with selected Cop RNO1 segments from the predicted Mcs3 QTL were introgressed into a WF/NHsd genetic background by successive backcrossing to the WF strain. Congenic rat lines were maintained in an Association for the Assessment and Accreditation of Laboratory Animal Care (AAALAC) approved facility on a 12-hour light/dark cycle and provided LabDiet 5001 Rodent Diet (PMI Nutrition International) and water ad libitum. All animal protocols were approved by the University of Louisville Animal Care and Use Committee. Sequence information and locations of genetic markers defining the ends of COP alleles in each congenic line A, D, E, and G are available at the UCSC Genome Browser (www.genome.ucsc.edu), the Rat Genome Database (http://rgd.mcw.edu/), or Table S1. Animals were genotyped as described in (SAMUELSON et al. 2003). Briefly, tail clips were used for DNA extraction and genotyping by either gel electrophoresis or Sanger sequencing of PCR products. Informative genetic markers at respective congenic ends and at 10 Mb intervals between ends were used to determine genotypes. For microsatellite markers, each fast-PCR underwent denaturation of 95°C for 10 s, followed by 40 cycles of 94°C for 0 s and 63°C for 8 s, and an extension at 72°C for 30 s on an Applied Biosystems Veriti Fast Thermal Cycler. Amplified genomic DNA was run on a 3% high-resolution agarose gel. After electrophoresis, gels were stained in SYBR Gold and scanned using a Typhoon imager. Visible gel-bands of the appropriate size were analyzed along with
DNA from homozygous Cop and WF, and (WF x Cop)F₁ (heterozygous) rats. In regions where informative micro-satellite markers were limiting, Primer3 was used to design primers to PCR amplify genomic regions containing potentially WF/Cop informative single nucleotide variants (SNVs), identified using the SNPlotyper function available at the Rat Genome Database. Genomic DNA was extracted from spleen tissue and PCR amplified. These reactions were cleaned using Promega Wizard SV Gel and PCR clean–up system. Samples were sequenced by the University of Louisville, Center for Genetics and Molecular Medicine, sequencing core using an ABI PRISM 7700 Sequence Detection System. Primer sequences that define the ends of Mcs3 congenic lines A and E can be found in Table S2.

**Phenotyping**

Female WF.Cop congenic and WF/NHsd (Envigo) rats were gavaged with a single oral dose of DMBA (65 mg DMBA/kg body mass) in sesame oil at 50-55 days of age. Mammary carcinomas ≥3 × 3 mm² were counted at 15 weeks post-treatment.

**Comparative Genomics**

Human orthologous regions and transcripts mapping to the delimited rat Mcs3 and human orthologous loci were identified using the UCSC Genome Browser. The *Rattus norvegicus* reference genome sequence version RGSC 6.0/rm6 and *Homo sapiens* version GRCh38/hg38 were used. Breast cancer associated genes were identified by searching NCBI/PubMed using the respective Mcs3-nominated gene name and breast cancer as search terms. Databases at NHGRI-EBI (WELTER et al. 2014) and GWAS Central (BECK et al. 2014) were searched for variants located in Mcs3 orthologous regions that had p values of 1 x 10⁻³ or less for association to breast cancer.

**Statistical Analysis**

Mammary carcinoma multiplicity data were analyzed by comparing congenic strain phenotypes to a susceptible WF/NHsd phenotype. First, mammary carcinoma multiplicity phenotypes for all lines were compared using the Kruskal-Wallis nonparametric test to protect for multiple comparisons. Following a significant Kruskal-Wallis test, which was p < 0.0001, select group comparisons were made by two-tailed Mann-Whitney nonparametric tests corrected for ties. P values ≤ 0.05 were considered statistically significant.

**Data and Strain Availability**

Strain WF.Cop D is available upon request. Table S1 contains sequence and location information for primers to amplify microsatellite markers on RNO1. Table S2 contains sequence and location information for primers to amplify RNO1 SNV containing sequences. File S1 contains a list of Mcs3-nominated breast cancer susceptibility candidate genes.

**Results**

**Physical confirmation and positional mapping of Mcs3 with WF.Cop congenics**

The Mcs3 QTL was predicted to exist on RNO1 by Gould and colleagues in a previous study (SHEPEL et al. 1998). Figure 1 is a map of WF.Cop congenic strains used in the present study to physically confirm and positionally map the location of Mcs3. Table 1 contains DMBA-induced mammary carcinoma susceptibility phenotypes (mean ± SD mammary carcinomas per rat) of the congenic lines depicted in Figure 1. Congenic lines A (n=30) and D (n=19)
developed 3.4 ± 2.2 and 2.8 ± 2.3 mammary tumors per rat, respectively. Mammary cancer susceptible WF females (n=12) developed 6.3 ± 3.8 DMBA-induced mammary carcinomas per rat. Females from lines A and D developed significantly less mammary tumors per rat compared to susceptible WF control females (p-values = 0.019 and 0.015, respectively). Reduced susceptibility lines A and D contained partially overlapping segments of Cop RNO1 and were not different from each other with respect to DMBA-induced tumor multiplicity (p-value = 0.198). This suggests that these congenic lines likely contain the same QTL. Thus, the reduced susceptibility phenotypes of both lines, A and D, physically confirm the Mcs3 locus.

Congenic lines E (n=25) and G (n=29) developed 6.6 ± 3.4 and 8.1 ± 3.4 mammary tumors per rat, respectively. Neither line was significantly different in tumor multiplicity compared to susceptible WF rats (p-values = 0.935 and 0.334, respectively). Thus, neither line E nor line G contains an independently acting mammary carcinoma susceptibility QTL with an effect on the WF susceptibility phenotype.

The region of overlap between Cop RNO1 segments that indicated a presence of at least one independently acting Mcs3 allele (lines A and D) in combination with non-overlap between these segments and the Cop RNO1 segment contained in line E, which did not indicate the presence of an independently acting Cop allele, was used to delimit the Mcs3 QTL interval (Figure 1). Congenic Mcs3-resistance associated lines A and D overlapped each other minimally from markers D1Rat321 to rs105307119, and maximally from D1Rat320 to rs105131702. Susceptible line E minimally overlapped resistance-associated line A from rs105409983 to rs8149408 and resistance-associated line D from D1Rat320 to rs8149408. Considering the phenotypes and relevant regions of overlap and non-overlap of lines A, D, and E, the Mcs3 QTL was
delimited to a 27.8 Mb region of RNO1 that spans from SNVs rs8149408 to rs105131702. Respectively, these SNVs physically mapped to RNO1 base positions 143,700,228 and 171,517,317 of the rat reference genome build RGSC 6.0/rn6.

Results of the original QTL scan predicted two LOD score peaks on RNO1 at markers D1Mit11 and D1Wox6 (SHEPEL et al.). These peak markers are denoted with asterisks in Figure 1. Congenic line E had the Cop genotype at both peak markers, but did not have an Mcs3-associated resistance phenotype, as would be predicted based on the original QTL scan. The phenotype of congeneric line G, which was derived from line E and contained the Cop allele at D1Mit11 and the WF allele at D1Wox6, was measured to determine if epistatic QTLs might be responsible for the susceptibility phenotype of line E. The susceptibility phenotypes of lines E and G were not significantly different when compared to each other (p-value = 0.182). Thus, we did not uncover evidence of an epistatic interaction to potentially explain the susceptibility phenotype of line E.

Human orthologous regions to the rat Mcs3 QTL

The delimited rat Mcs3 QTL from rs8149408 to rs105131702 of RNO1 was found to align to four syntenic regions of human chromosomes 11 and 15 and four other regions of the human genome that contained a single gene (Table 2). The UCSC genome browser (https://genome.ucsc.edu) was used to identify known genes at rat Mcs3 and orthologous human loci (KENT et al. 2002; KAROLCHIK et al. 2004). These genes are listed in File S1. There were 310 rat and 287 human genes annotated in the respective genomic intervals. At least 198 of these genes were in common or orthologous between these species.

Public databases, namely the NHGRI-EBI catalog (https://www.ebi.ac.uk/gwas) (WELTER et al. 2014), GWAS Central (www.gwascentral.org) (BECK et al. 2014), NCBI PubMed (https://www.ncbi.nlm.nih.gov/pubmed) (PUBMED), and the Cancer Portal of Rat Genome Database (https://rgd.mcw.edu) (RGD), were searched to identify genes and genetic variation at Mcs3 orthologous regions.
human genome regions that have been associated with breast carcinoma susceptibility and development. Table 3 contains a list of female breast cancer associated genes and amplified regions that have \textit{Mcs3}-nominated rat orthologs. Rat \textit{Mcs3} was found to contain sequence orthologous to human \textit{11q13/14}, which contains multiple genes and is amplified in a subset of female breast carcinomas with poor prognosis (TSUDA \textit{et al.} 1989). In addition to a functional correlation with breast cancer development, studies reporting genetic associations of \textit{CYP2C19} variants with breast cancer susceptibility have been published (JUSTENHOVEN \textit{et al.} 2009; SANGRAJRANG \textit{et al.} 2009; GAN \textit{et al.} 2011).

Table 2. Minimum-number or names of genes at human loci containing rat \textit{Mcs3}$^1$ orthologous sequence

<table>
<thead>
<tr>
<th>Human Locus</th>
<th>Position$^2$</th>
<th>Gene Transcripts (minimum number or name)</th>
<th>Human</th>
<th>Rat</th>
<th>Human/Rat Orthologs</th>
</tr>
</thead>
<tbody>
<tr>
<td>15q25.2</td>
<td>15:83134545-84130720</td>
<td></td>
<td>6</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>15q25.1-15q25.2</td>
<td>15:80005820-82285404</td>
<td></td>
<td>22</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>11q13.4-11q13.5,11q14.1-11q14.3</td>
<td>11:71915764-89617253</td>
<td></td>
<td>142</td>
<td>103</td>
<td>87</td>
</tr>
<tr>
<td>11p15.4</td>
<td>11:3609839-7196087</td>
<td></td>
<td>113</td>
<td>186</td>
<td>90</td>
</tr>
<tr>
<td>Xq28</td>
<td>X:154886349-154888061</td>
<td></td>
<td>F8A1</td>
<td>F8a1</td>
<td>1</td>
</tr>
<tr>
<td>10q23.33</td>
<td>10:94762624-94853260</td>
<td></td>
<td>CYP2C19</td>
<td>Cyp2c6v1</td>
<td>1</td>
</tr>
<tr>
<td>11p11.12</td>
<td>11:49146635-49208670</td>
<td></td>
<td>FOLH1</td>
<td>Folh1</td>
<td>1</td>
</tr>
<tr>
<td>2p16.3</td>
<td>2:51696250-51787033</td>
<td></td>
<td>LOC730100</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Totals</td>
<td></td>
<td></td>
<td>287</td>
<td>310</td>
<td>198</td>
</tr>
</tbody>
</table>

$^1$Rat \textit{Mcs3} from SNVs \textit{rs8149408} to \textit{rs105131702} or \textit{RNO1}:143700228-171517317 (RGSC 6.0/rn6)

$^2$Human genome reference GRCh38/hg38
Table 3. Rat Mcs3-nominated human genes and amplified regions correlated with breast cancer development.

<table>
<thead>
<tr>
<th>Genes</th>
<th>Locus</th>
<th>RefSeq on Function</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ILK</strong> integrin-linked kinase</td>
<td>11p15.4</td>
<td>Serine/threonine kinase that regulates integrin-mediated signaling</td>
<td>(YANG et al. 2013)</td>
</tr>
<tr>
<td><strong>IL18BP</strong> interleukin 18 binding protein</td>
<td>11q13.4</td>
<td>Inhibitor of proinflammatory cytokine IL18</td>
<td>(YANG et al. 2015)</td>
</tr>
<tr>
<td><strong>EMSY</strong> EMSY, BRCA2 interacting transcriptional repressor</td>
<td>11q13.5</td>
<td>Repression of BRCA2-mediated DNA repair</td>
<td>(HUGHES-DAVIES et al. 2003; MADJD et al. 2014)</td>
</tr>
<tr>
<td><strong>PAK1</strong> p21 (RAC1) activated kinase 1</td>
<td>11q13.5-q14.1</td>
<td>Serine/threonine kinase that regulates cell motility and morphology</td>
<td>(ADAM et al. 2000; WANG et al. 2005; BOSTNER et al. 2007)</td>
</tr>
<tr>
<td><strong>RSF1</strong> remodeling and spacing factor 1</td>
<td>11q14.1</td>
<td>Histone chaperone involved in RSF chromatin-remodeling complex</td>
<td>(REN et al. 2014)</td>
</tr>
<tr>
<td><strong>GAB2</strong> GRB2 associated binding protein 2</td>
<td>11q14.1</td>
<td>Adapter protein involved in signal transduction with a known role in PI3K activation</td>
<td>(BOCANEGRA et al. 2010)</td>
</tr>
<tr>
<td><strong>CYP2C19</strong> cytochrome P450 family 2 subfamily C member 19</td>
<td>10q23.33</td>
<td>Monoxygenase involved in xenobiotic and estrogen metabolism</td>
<td>(JUSTENHOVE N et al. 2009; SANGRAJRAN G et al. 2009; GAN et al. 2011)</td>
</tr>
</tbody>
</table>

Genome-wide significant associations (p values $< 10^{-7}$) between human genetic variants located in Mcs3-orthologous multi-gene regions and female breast cancer risk have not been reported, but potentially associated variants with P values $< 10^{-3}$ have been reported (Table 4).
Discussion

Rat *Mcs3* was physically confirmed and delimited to a 27.8 Mb segment of rat chromosome 1. Rat *Mcs3* is the last of the known Cop rat mammary carcinoma resistance-associated QTLs (*Mcs1*-3) to be physically confirmed. In their linkage analysis predicting the *Mcs3* QTL, Gould and colleagues reported that *Mcs3* heterozygous females had, on average, a 42% reduction in mammary carcinoma multiplicity compared to the WF phenotype; and, females homozygous for the *Mcs3* Cop allele had an 84% reduction in mammary tumor number (SHEPEL et al. 1998). Thus, they appropriately concluded that there was no dominance effect at the *Mcs3* QTL.

While we did not test heterozygous females in our study, we observed that the *Mcs3* Cop allele reduced the mammary carcinoma susceptibility phenotype of the highly susceptible WF strain from 46-65% when homozygous. The discrepancy between homozygous genotypes in these two studies may be due to effects of Cop resistance alleles at other *Mcs* resistance-associated QTLs. Another possibility is *Mcs3* contains multiple independent QTLs, and at least one of these was not contained in the congenic strains tested for our study.

Table 4. Rat *Mcs3*-nominated human variants potentially associated with breast cancer susceptibility

<table>
<thead>
<tr>
<th>Variant</th>
<th>Locus</th>
<th>Base Position (hg38)</th>
<th>Genomic Context/Position</th>
<th>P-Value for Association</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1435808</td>
<td>2p16.3</td>
<td>chr2:52,036,315</td>
<td>Intronic/LOC730100</td>
<td>0.000642</td>
<td>(HUNTER et al. 2007)</td>
</tr>
<tr>
<td>rs6733295</td>
<td></td>
<td>chr2:52,040,189</td>
<td>Intronic/LOC730100</td>
<td>0.000959</td>
<td>(HUNTER et al. 2007)</td>
</tr>
<tr>
<td>rs4352262</td>
<td></td>
<td>chr2:53,374,593</td>
<td>Intergenic/between LOC730100 and ASB3</td>
<td>0.000893</td>
<td>(HUNTER et al. 2007)</td>
</tr>
<tr>
<td>rs10168550</td>
<td></td>
<td>chr2:53,385,857</td>
<td>Intergenic/between LOC730100 and ASB3</td>
<td>0.000381</td>
<td>(HUNTER et al. 2007)</td>
</tr>
<tr>
<td>rs7126870</td>
<td>11p15.4</td>
<td>chr11:3868829</td>
<td>Intronic/STIM1</td>
<td>0.000898</td>
<td>(HUNTER et al. 2007)</td>
</tr>
<tr>
<td>rs1459952</td>
<td>11q1.1</td>
<td>chr11:81871697</td>
<td>Intergenic/between LOC101928944 and MIR4300HG</td>
<td>0.000562</td>
<td>(KIBRIYA et al. 2009)</td>
</tr>
<tr>
<td>rs3793949</td>
<td>11q14.1</td>
<td>chr11:83847695</td>
<td>Intronic/DLG2</td>
<td>0.000702</td>
<td>(HUNTER et al. 2007)</td>
</tr>
<tr>
<td>rs11235127</td>
<td>11q14.2</td>
<td>chr11:87377502</td>
<td>Intergenic/between THEM135 and LOC105369423</td>
<td>0.00005</td>
<td>(EASTON et al. 2007)</td>
</tr>
<tr>
<td>rs12269979</td>
<td></td>
<td>chr11:87960152</td>
<td>Intergenic/between LOC105369423 and RAB38</td>
<td>0.000439</td>
<td>(HUNTER et al. 2007)</td>
</tr>
<tr>
<td>rs6495623</td>
<td>15q25.2</td>
<td>chr15:81848308</td>
<td>Intergenic/between TMC3 and MEX3B</td>
<td>0.000871</td>
<td>(HUNTER et al. 2007)</td>
</tr>
</tbody>
</table>
The 27.8 Mb region of RNO1 from SNVs rs8149408 to rs105131702 is the segment most likely to contain Mcs3; however, it is possible that Mcs3 is more complex. For example, an interaction between one or more elements in the delimited region and elements outside of this region could be required to confer the Mcs3 associated phenotype. The delimited Mcs3 segment is distal to both D1Mit11 and D1Wox6, the markers with peak LOD scores from the original QTL scan, but within the predicted Mcs3 QTL interval (SHEPEL et al. 1998). This is not the first time a rat Mcs locus has been mapped to a genomic segment that did not contain a peak marker of the original QTL scan. For example, neither Mcs1b nor Mcs1c QTLs contain the peak LOD marker of the predicted Mcs1 QTL (HAAG et al. 2003a; DENDEKKER et al. 2012). The Mcs3 QTL, as delimited here, overlaps the distal third of mouse Mammary tumor susceptibility modifier 1 (Mtsm1) (KOCH et al. 2007). Interestingly, the mouse Mtsm1 segment overlapping rat Mcs3 does not contain the peak LOD score marker defining the Mtsm1 QTL. Furthermore, it is worth noting that Mcs3 had one of the lowest LOD scores of Mcs QTLs identified in linkage analyses using WF females as the susceptible strain (SHEPEL et al.; LAN et al.). Thus, interval mapping effectively predicted rat Mcs QTLs; and, as expected, a wide genomic region of interrogation was required in congenic studies to pinpoint the location of each QTL for fine-mapping.

Most importantly, this work provides information that is translatable to the genetic component of human breast cancer by using a versatile experimental organism that is highly relevant to female breast cancer. Different segments of the rat Mcs3 locus align to four multi-gene human syntenic regions on chromosomes 11 and 15 and four single-gene loci at Xq28, 2p16.3, 10p23.33, and 11p11.12. The rat Mcs3 locus contains 310 annotated genes and the orthologous human regions contain 287 annotated genes. To reduce the number of Mcs3-nominated candidate susceptibility genes, it will be necessary to fine-map the Mcs3 locus. However, the human orthologous regions to rat Mcs3 could be targeted for deep genetic analysis to determine if these regions contain risk associated variants. In support of this direction, population-based genetic association studies have identified potentially associated variants at Mcs3 orthologous loci 2p16.3, 11p15.4, 11q14.1, 11q14.2, and 15q25.2 (EASTON et al. 2007; HUNTER et al. 2007; KIBRIYA et al. 2009). This provides strong rationale for additional human studies to identify risk associated variants at these loci. Human genome targeted association studies require considerable resources, as a high density of variants and a large sample size are required to properly test each targeted locus. Guidance from both rat genomics and previous human genetic studies stands to increase the likelihood of finding positive associations. Thus, these human loci are excellent candidates for further genetic analyses to test a high density of variants for association to risk.

Discussion of the entire list of genes located within the Mcs3 QTL, as currently defined, is not practical; however, it is notable that some Mcs3-nominated candidate susceptibility genes have known or suggested roles in breast cancer progression, diagnosis, or prognosis. This list includes p21 (RAC1) activated kinase 1 (PAK1), EMSY, BRCA2 interacting transcriptional repressor (EMSY), remodeling and spacing factor 1 (RSF1), GRB2 associated binding protein 2 (GAB2), integrin-linked protein kinase (ILK), interleukin-18-binding protein (IL18BP), and cytochrome P450 family 2 subfamily C member 19 (CYP2C19).

The EMSY, PAK1, RSF1, and GAB2 genes are located within the 11q13/14 breast
cancer amplicon. Interestingly, the breast cancer oncogene cyclin D1 (CCND1), which is also located within the 11q13/14 amplicon, is not contained within the rat Mcs3 delimited region. The EMSY gene encodes a BRCA2 binding partner that silences a potential transcription activation domain of BRCA2; and thereby, represses BRCA2-mediated DNA repair (Hughes-Davies et al. 2003; Moelans et al. 2010). Amplification of EMSY has been reported in 13% of breast cancers and is associated with poor survival, especially in node negative breast cancer (Hughes-Davies et al. 2003). A recent study reported a significant positive association between EMSY expression and lymph node metastasis, as well as a larger tumor size (Madid et al.). The PAK1 gene product has many roles in cancer, including breast cancer progression, development and maintenance of a metastatic phenotype of breast cancer cells, and a predictor of recurrence and tamoxifen resistance in post-menopausal breast cancer (Adam et al. 2000; Bostner et al. 2007; Kumar and Li 2016). Ectopic expression of activated Pak1 has been shown to induce mouse mammary tumors (Wang et al. 2005). High Rsf-1 expression is associated with breast cancer sub-type and poor prognosis (Ren et al. 2014). Amplification of GAB2, independent of CCND1, has been observed in breast carcinoma samples (Daly et al. 2002; Bocanegra et al. 2010).

A direct interaction between PAK1 and ILK proteins has been identified. The activity of ILK is regulated by PAK1 phosphorylation (Acciconia et al. 2007). Higher transcript levels of ILK have been found in human breast cancer tissue compared to adjacent non-diseased breast tissue (Yang et al. 2013). In the same study, breast cancer patients with more intense immunostaining for ILK had lower 5-year survival than patients with low ILK levels (Yang et al. 2013).

Interleukin-18-binding protein (IL18BP) regulates the activity of IL-18, which has been shown to be higher in breast carcinoma tissue compared to tissue from patients with benign breast disease (Srabovic et al. 2011). Another study suggests that IL-18 enhances breast cancer cell migration (Yang et al. 2015). Due to a potential role of IL-18 in cancer progression and metastasis, an IL18BP-Fc has been developed to antagonize the effects of IL-18 (Ca et al.). Because of an established role in estrogen metabolism, CYP2C19 variants have been tested for association with breast cancer risk. There is evidence that the CYP2C19*3 variant may be associated with breast cancer risk in Chinese Han women (Gan et al. 2011). Neither the CYP2C19*17 variant, which results in a rapid metabolizer phenotype, nor the CYP2C19*2 variant were associated with breast cancer risk in population-based studies of women of European descent; however, CYP2C19*17 was associated with decreased risk in women using hormone replacement therapy for greater than 10 years (Justenhoven et al. 2009; Marie-Genica Consortium on Genetic Susceptibility for Menopausal Hormone Therapy Related Breast Cancer Risk 2010; Justenhoven et al. 2012). In a study of cancer recurrence in ER+ post-menopausal breast cancer cases, it was concluded that patients with a CYP2C19*2 allele may benefit more from tamoxifen therapy (BeeLEN et al. 2013).

In summary, rat Mcs3 is an independently acting QTL located in a 27.8 Mb region on RNO1 between rs8149408 and rs105131702. Additional congenic studies and other genomic approaches will be necessary to reduce the Mcs3-nominated candidate susceptibility gene list and determine mechanisms of Mcs3-associated reduced susceptibility to mammary carcinomas. While an overwhelming list of
Mcs3-nominated breast cancer susceptibility candidate genes was identified, a manageable number of orthologous human syntenic regions were found to warrant deeper analysis of these loci in human population-based genetic association studies. The fact that some of these regions contain variants potentially-associated with breast cancer risk further supports the need to ultra-fine map these loci to determine if true positive associations to susceptibility exist.

**Acknowledgments**

This work was supported by the National Institutes of Health grant R01CA137052, the University of Louisville, School of Medicine, Department of Biochemistry & Molecular Genetics, and the University of Louisville, School of Medicine Research Office. Authors thank Jennifer Sanders, Huy T. Le, and Aaron Puckett for their technical assistance.
References:


RGD, Cancer Portal, Rat Genome Database Web Site, pp. Medical College of Wisconsin, Milwaukee, Wisconsin.


Samuelson, D. J., S. E. Hesselson, B. A. Aperovich, Y. Zan, J. D. Haag et al., 2007 Rat Mcs5a is a compound quantitative trait locus with orthologous human loci that associate with breast cancer risk. Proc Natl Acad Sci U S A 104: 6299-6304.


