

RESEARCH ARTICLE

Open Access

Significant overlap between human genome-wide association-study nominated breast cancer risk alleles and rat mammary cancer susceptibility loci

Jennifer Sanders¹ and David J Samuelson^{1,2,3,4*}

Abstract

Introduction: Human population-based genome-wide association (GWA) studies identify low penetrance breast cancer risk alleles; however, GWA studies alone do not definitively determine causative genes or mechanisms. Stringent genome-wide statistical significance level requirements, set to avoid false-positive associations, yield many false-negative associations. Laboratory rats (*Rattus norvegicus*) are useful to study many aspects of breast cancer, including genetic susceptibility. Several rat mammary cancer associated loci have been identified using genetic linkage and congenic strain based-approaches. Here, we sought to determine the amount of overlap between GWA study nominated human breast and rat mammary cancer susceptibility loci.

Methods: We queried published GWA studies to identify two groups of SNPs, one that reached genome-wide significance and one comprised of SNPs failing a validation step and not reaching genome-wide significance. Human genome locations of these SNPs were compared to known rat mammary carcinoma susceptibility loci to determine if risk alleles existed in both species. Rat genome regions not known to associate with mammary cancer risk were randomly selected as control regions.

Results: Significantly more human breast cancer risk GWA study nominated SNPs mapped at orthologs of rat mammary cancer loci than to regions not known to contain rat mammary cancer loci. The rat genome was useful to predict associations that had met human genome-wide significance criteria and weaker associations that had not.

Conclusions: Integration of human and rat comparative genomics may be useful to parse out false-negative associations in GWA studies of breast cancer risk.

Introduction

Breast cancer is a complex disease characterized by environmental, genetic, and epigenetic factors. Due to the complexity of developing this disease a woman's individual risk may vary greatly from population risk estimates. The familial relative risk of developing breast cancer increases with the number of affected relatives, suggesting that there is a strong genetic component associated with this disease [1,2]. High-penetrance breast cancer risk mutations such as those of *BRCA1* and *BRCA2* have been identified [3,4]. Population frequencies of mutations with high-penetrance

toward risk are rare due to their severe effects on individuals and, thus, these mutations account for only a small percentage of population risk. Risk alleles with moderate penetrance and minor allele population frequencies of 0.005 to 0.01 (for example, *PALB2*) are estimated to account for approximately 3% of risk. Therefore, a majority of population-based breast cancer risk is likely explained by low penetrance alleles with rare to common population frequencies [5].

Genome-wide association (GWA) studies have been used to identify several low penetrance breast cancer risk alleles [6]. Due to a need to control for numerous multiple comparisons made in GWA studies, a Bonferroni correction based *P*-value cut-off of $\leq 1 \times 10^{-7}$ is typically required for an association to be considered genome-wide significant. It has been suggested that this approach is too stringent as it may result in many false negative

* Correspondence: david.samuelson@louisville.edu

¹Department of Biochemistry and Molecular Biology, University of Louisville School of Medicine, Louisville, KY 40292, USA

²James Graham Brown Cancer Center, University of Louisville School of Medicine, Louisville, KY 40292, USA

Full list of author information is available at the end of the article

associations [7]. Furthermore, while GWA studies are unbiased approaches to identify genomic regions associated with breast cancer risk, these epidemiology-based approaches cannot easily determine risk genes or genetically determined mechanisms of susceptibility. Currently, only a small percentage of breast cancer heritability is explained by published studies suggesting that considerable genetic variation associated with breast cancer risk remains to be identified [5,8].

Comparative genetics between rats and humans has also been used to identify breast cancer risk alleles [9]. In general, the laboratory rat is a good experimental organism to model breast cancer. Compared to induced mammary tumors in mice, rats develop mammary carcinomas of ductal origin, which is similar to a majority of human breast cancers. Also, rat mammary tumors are responsive to estrogen, just as are a majority of human breast tumors [10,11]. Most importantly, the laboratory rat is a versatile organism to study breast cancer susceptibility, as experiments can be controlled at genetic and environmental levels. Inbred rat strains exhibit differential susceptibility to chemically induced carcinogenesis using 7,12-dimethylbenz[a]anthracene (DMBA) [10,12-14]. Copenhagen (COP) and Wistar-Kyoto (WKY) rat strains are resistant to DMBA, *N*-Nitroso-*N*-methylurea (NMU), and oncogene induced mammary carcinomas, while the Wistar-Furth (WF) rat strain is susceptible.

Previous genetic studies using rats have identified eight *Mammary carcinoma susceptibility (Mcs)* loci, named *Mcs 1-8* [15-18]. A (WF \times COP)_{F1} \times WF backcross design was used to identify *Mcs 1-4*. Copenhagen alleles at *Mcs 1-3* are associated with decreased mammary tumor multiplicity, while the *Mcs 4* COP allele is associated with increased tumor development [15]. Further analysis of the *Mcs1* locus using WF.COP congenic lines, spanning different regions of the quantitative trait locus (QTL), identified three independent loci associated with mammary carcinoma susceptibility, named *Mcs 1a-c* [17]. Another linkage analysis study using WF and WKY rat strains revealed four additional QTLs associated with mammary carcinoma susceptibility, named *Mcs 5-8*. Additionally, a modifier of *Mcs8*, *Mcsm1*, partially counteracts the resistance phenotype conferred by *Mcs8* [16,18]. Further analysis of the *Mcs5* locus using WF.WKY congenic rat lines resulted in the identification of four subloci named *Mcs5a1*, *Mcs5a2*, *Mcs5b* and *Mcs5c* [9,19]. Additional linkage analysis using the SPRD-Cu3 rat strain (DMBA-induced mammary carcinogenesis susceptible) and the resistant WKY rat strain resulted in the identification of three more rat QTLs associated with mammary cancer named *Mcstm1*, *Mcstm2/Mcsta2* and *Mcsta1* [20,21]. Several rat genomic regions that associate with mammary cancer susceptibility were identified using beta-estradiol instead of DMBA to induce carcinogenesis.

These QTLs were identified using the August Copenhagen Irish (ACI) rat strain, which is susceptible to beta-estradiol carcinogenesis and the COP and Brown Norway (BN) rat strains, which are resistant. These loci are named *Estrogen-induced mammary cancer* loci or *Emca 1-2* and *Emca 4-8* [22,23].

Comparative genomics between human breast and rat mammary cancer risk alleles will continue to be warranted, especially if appreciable overlap in genetic susceptibility exists between these species. In this study, genomic locations of human breast cancer risk GWA study-identified polymorphisms were compared to the rat genome to determine if positive associations were more often located at orthologs to rat mammary cancer risk loci than at randomly selected regions not known to be associated with rat mammary cancer susceptibility.

Methods

Converting rat mammary cancer associated loci to human orthologous regions

No research animals were used in this work. Previously published information on rat mammary cancer associated loci was used. Human orthologous regions of rat regions that associate with mammary cancer susceptibility listed in Table 1 were determined using the 'In other genomes (convert)' function available at the University of California Santa Cruz (UCSC) genome browser [24]. Rat Nov. 2004 (Baylor 3.4/rn4) and human Feb. 2009 (GRCh37/hg19) genome assemblies were used. If a rat mammary cancer locus split into multiple human orthologous regions, we noted all orthologous regions until they reached less than 1% of the bases and spanned less than 1% of the original rat mammary cancer locus using the UCSC genome browser.

Random rat regions

To determine if human GWA study-identified polymorphisms map to rat mammary cancer loci more frequently than to random regions of the rat genome, we selected rat genome segments that have not shown an association with mammary cancer risk. These rat genomic regions were named 'random rat regions' and are listed in Table 2. We initially focused on fourteen *Mcs/Mcsm* regions with an average size of 22,322,710 bps as these are generally smaller in size than other rat mammary cancer associated loci identified. Fourteen random rat genome regions, each 22,322,710 bps in size were used for comparison. Random rat regions were selected by picking a chromosome using a random number generator function of Microsoft Excel. The range of chromosomes entered into the random number generator function was 1 through 21 (rats have 21 chromosomes, including a sex chromosome). The start position for each random rat region was determined using a random

Table 1 Location of rat mammary cancer susceptibility loci and human orthologous regions used in this study

Rat <i>Mcs</i> locus (Overlap)	Boundary markers	Rat chr	Region(UCSC rat assembly 2004)	Reference	Human orthologous region (UCSC human assembly 2009)
DMBA induced mammary carcinogenesis					
<i>Mcs1a</i>	D2Mit29 to D2Uwm14	RNO2	5,601,528- 10,539,344	Haag <i>et al.</i> [17]	<i>Chr5</i> : 89,216,702-93,113,337
<i>Mcs1b</i>	ENSRNOSNP2740854 to g2UI2-27	RNO2	42,364,155-44,195,382	DenDekker <i>et al.</i> [25]	<i>Chr5</i> : 54,816,178-57,003,049
<i>Mcs1c</i>	D2M13Mit286 to D2Uia5	RNO2	13,909,383- 20,666,092	Haag <i>et al.</i> [17]	<i>Chr5</i> : 81,891,633-86,857,442 <i>Chr5</i> : 86,171,198-86,251,067
<i>Mcs2</i> (overlaps <i>Mcs6</i> , <i>Emca4</i>)	D7rat39 to D7Uwm12	RNO7	4,936,704-86,028,057	Sanders <i>et al.</i> [18]	<i>Chr12</i> : 57,316,160-108,177,690 <i>Chr8</i> : 97,242,984-115,650,989 <i>Chr19</i> : 281,161-2,497,331 <i>Chr19</i> : 15,059,910-15,808,112
<i>Mcs3</i>	D1Rat27 to D1Mit12	RNO1	90,282,174-156,954,117	Shepel <i>et al.</i> [15]	<i>Chr15</i> : 80,282,370-102,265,870 <i>Chr15</i> : 25,574,935-28,567,541 <i>Chr11</i> : 17,403,456-22,898,646 <i>Chr11</i> : 74,958,193-89,350,902 <i>Chr19</i> : 48,799,986-51,921,957 <i>Chr19</i> : 28,701,413-30,656,003
<i>Mcs4</i>	D8Rat164 to D8Rat108	RNO8	28,414,100-72,403,639	Shepel <i>et al.</i> [15]	<i>Chr11</i> : 107,453,990-132,383,506 <i>Chr15</i> : 62,105,069-76,028,735 <i>Chr15</i> : 76,091,658-78,185,872 <i>Chr15</i> : 78,380,119-78,998,961 <i>Chr15</i> : 51,349,646-51,942,505
<i>Mcs5a1</i> (overlaps <i>Mcstm1</i> , <i>Emca8</i>)	SNP-61634906 to SNP- 61666918	RNO5	61,634,727-61,666,739	Samuelson <i>et al.</i> [9]	<i>Chr9</i> : 37,562,516-37,589,491
<i>Mcs5 a2</i> (overlaps <i>Mcstm1</i> , <i>Emca8</i>)	SNP-61667232 to gUwm23-29	RNO5	61,667,053-61,751,614	Samuelson <i>et al.</i> [9]	<i>Chr9</i> : 37,590,988-37,654,512
<i>Mcs5b</i> (overlaps <i>Mcstm1</i> , <i>Emca8</i>)	gUwm50-20 to D5Got9	RNO5	65,498,190-67,464,050	Samuelson <i>et al.</i> [19]	<i>Chr9</i> : 103,492,712-105,220,552
<i>Mcs5c</i> (overlaps <i>Mcstm1</i> , <i>Emca8</i>)	gUwm74-1 to gUwm54-8	RNO5	81,118,457-81,295,367	Veillet <i>et al.</i> [26]	<i>Chr9</i> : 118,231,525-118,416,951 <i>Chr12</i> : 72,033,141-72,033,263
<i>Mcs6</i> (overlaps <i>Mcs2</i>)	D7Rat171 to gUwm64-3	RNO7	22,382,725-55,384,873	Sanders <i>et al.</i> [18]	<i>Chr12</i> : 71,270,266-105,502,699
<i>Mcs7</i> (overlaps <i>Mcsta1</i>)	D10Got124 to gUwm58-136	RNO10	89,575,060-100,335,500	Cotroneo <i>et al.</i> [27]	<i>Chr17</i> : 40,183,547-67,946,104
<i>Mcs8</i>	D14Mit1 to D14Rat99	RNO14	12,386,493-26,416,791	Lan <i>et al.</i> [16]	<i>Chr4</i> : 65,556,457-81,559,483
<i>Mcsm1</i> (overlaps <i>Emca7</i>)	D6Mit9 to D6Rat12	RNO6	34,039,303-114,032,192	Lan <i>et al.</i> [16]	<i>Chr14</i> : 25,151,530-80,417,386 <i>Chr2</i> : 334,41-18,603,019 <i>Chr7</i> : 12,561,599-19,619,365 <i>Chr7</i> : 107,770,320-111,916,436

Table 1 Location of rat mammary cancer susceptibility loci and human orthologous regions used in this study (Continued)

Mcstm1 (overlaps <i>Mcs5, Emca1, Emca8</i>)	D5rat124 to <i>Pla2g2a</i>	RNO5	19,206,257-157,657,360	Piessevaux <i>et al.</i> [21]	<i>Chr1</i> : 20,301,931-59,012,763 <i>Chr1</i> : 59,119,520-67,602,141 <i>Chr9</i> : 27,325,071-123,488,955 <i>Chr6</i> : 87,792,854-100,245,025 <i>Chr8</i> : 87,055,841-97,247,307 <i>Chr8</i> : 58,994,818-62,700,945
Mcstm2 (overlaps <i>Emca2</i>)	D18Wox8 to D18Rat44	RNO18	32,458,819-86,863,412	Piessevaux <i>et al.</i> [21]	<i>Chr18</i> : 10,202,644-13,129,349 <i>Chr18</i> : 41,356,963-54,158,113 <i>Chr18</i> : 54,267,924-58,201,561 <i>Chr18</i> : 66,912,039-78,010,606 <i>Chr5</i> : 112,300,500-130,363,372 <i>Chr5</i> : 142,780,151-147,624,793 <i>Chr5</i> : 147,647,196-150,176,352
Mcsta1 (overlaps <i>Mcs7</i>)	D10Rat91 to <i>D10Rat97</i>	RNO10	9,762,188-108,776,963	Piessevaux <i>et al.</i> [21]	<i>Chr5</i> : 130,482,861-173,663,969 <i>Chr5</i> : 177,530,539-180,675,650 <i>Chr17</i> : 690,639-15,624,409 <i>Chr17</i> : 16,916,926-20,222,700 <i>Chr17</i> : 25,525,650-78,247,249 <i>Chr16</i> : 78,402-6,094,950
β-estradiol induced mammary carcinogenesis					
Emca1 (overlaps <i>Mcstm1, Emca8</i>)	D5Rat53 to D5Rat57	RNO5	103,677,474-155,121,024	Gould <i>et al.</i> [22]	<i>Chr1</i> : 23,607,020-59,012,763 <i>Chr1</i> : 59,119,520-67,602,141 <i>Chr9</i> : 17,037,252-27,300,264
Emca2 (overlaps <i>Mcstm2</i>)	D18Rat27 to D18Rat43	RNO18	18,562,643-66,652,947	Gould <i>et al.</i> [22]	<i>Chr5</i> : 110,259,180-130,363,372 <i>Chr5</i> : 137,224,929-147,624,793 <i>Chr5</i> : 147,647,196-150,176,352 <i>Chr18</i> : 10,202,644-13,129,349 <i>Chr18</i> : 35,982,130-41,016,602 <i>Chr18</i> : 52,597,120-54,158,113 <i>Chr18</i> : 54,267,924-58,201,561 <i>Chr2</i> : 127,805,417-128,786,719
Emca4 (overlaps <i>Mcs2</i>)	D7Rat44 to D7Rat15	RNO7	66,201,980-107,428,439	Schaffer <i>et al.</i> [23]	<i>Chr8</i> : 97,242,984-137,409,536 <i>Chr12</i> : 57,316,160-59,093,375

Table 1 Location of rat mammary cancer susceptibility loci and human orthologous regions used in this study (Continued)

Emca5	D3Rat227 to D3Rat210	RNO3	41,054,012-171,063,335	Schaffer <i>et al.</i> [23]	<i>Chr20</i> : 1,746,912-62,907,504 <i>Chr2</i> : 110,841,402-113,650,057 <i>Chr2</i> : 159,530,076-188,395,371 <i>Chr11</i> : 26,296,319-57,753,858 <i>Chr15</i> : 32,905,485-34,664,466 <i>Chr15</i> : 34,933,152-51,298,144
Emca6	D4Rat14 to D4Rat202	RNO4	41,729,583-159,115,617	Schaffer <i>et al.</i> [23]	<i>Chr7</i> : 23,252,368-33,103,107 <i>Chr7</i> : 115,026,301-150,558,396 <i>Chr3</i> : 88,756-12,883,445 <i>Chr3</i> : 13,004,609-15,163,132 <i>Chr3</i> : 64,018,604-75,322,612 <i>Chr3</i> : 125,977,400-128,219,297 <i>Chr2</i> : 68,713,643-89,165,869 <i>Chr4</i> : 89,504,626-95,273,083 <i>Chr4</i> : 120,978,632-122,320,931 <i>Chr12</i> : 156,786-2,821,588 <i>Chr10</i> : 43,277,230-46,218,580
Emca7 (overlaps <i>Mcsm1</i>)	D6Rat68 to D6Rat81	RNO6	2,802,670-111,967,837	Schaffer <i>et al.</i> [23]	<i>Chr14</i> : 25,151,530-78,362,253 <i>Chr2</i> : 33,441-35,642,893 <i>Chr2</i> : 38,644,737-51,698,454 <i>Chr7</i> : 12,561,599-19,619,365 <i>Chr7</i> : 105,197,211-111,916,436
Emca8 (overlaps <i>Mcs5</i> , <i>Mcstm1</i> , <i>Emca1</i>)	D5Rat134 to D5Rat37	RNO5	52,434,178-148,460,381	Schaffer <i>et al.</i> [23]	<i>Chr9</i> : 6,756,013-27,300,264 <i>Chr9</i> : 27,925,947-123,488,955 <i>Chr1</i> : 33,159,021-59,012,763 <i>Chr1</i> : 59,119,520-67,602,141

Table 2 Random rat genomic segments and human orthologous regions used in this study

Rat <i>Mcs</i> locus	Rat <i>chr</i>	Region (UCSC rat assembly 2004)	Human orthologous region (UCSC human assembly 2009)
Random rat region 1	RNO9	20,000,000-44,322,711	<i>Chr2</i> : 97,158,323-106,711,249 <i>Chr6</i> : 56,223,874-73,919,999 <i>Chr2</i> : 189,182,486-189,878,065 <i>Chr13</i> : 103,235,577-103,556,495 <i>Chr2</i> : 128,848,569-129,254,860
Random rat region 2	RNO15	60,000,001-84322711	<i>Chr13</i> : 53,226,266-74,878,291 <i>Chr13</i> : 42,064,282-42,529,444
Random rat region 3	RNO16	68,621,246-92,943,956	<i>Chr13</i> : 103,539,456-115,092,822 <i>Chr8</i> : 36,627,241-42,308,840 <i>Chr8</i> : 638,582-6,693,649 <i>Chr8</i> : 42,690,588-43,056,179 <i>Chr13</i> : 52,753,969-53,050,606
Random rat region 4	RNO9	91,398,460-115,721,170	<i>Chr5</i> : 98,385,946-110,062,886 <i>Chr18</i> : 612,848-9,957,727 <i>Chr2</i> : 240,340,012-242,806,427
Random rat region 5	RNO13	55,373,307-79,696,017	<i>Chr1</i> : 169,844,936-194,938,667
Random rat region 6	RNO11	39,408,000-63,730,710	<i>Chr3</i> : 95,108,010-118,895,417
Random rat region 7	RNO17	68,384,015-92,706,72	<i>Chr10</i> : 138,740-22,530,353 <i>Chr1</i> : 236,673,870-240,084,642
Random rat region 8	RNO3	12,585,543-36,908,253	<i>Chr2</i> : 140,246,548-155,465,845 <i>Chr9</i> : 123,526,091-129,443,210
Random rat region 9	RNO19	34,130,390-58,453,100	<i>Chr16</i> : 66,968,878-90,107,058 <i>Chr10</i> : 33,502,588-35,153,585 <i>Chr1</i> : 229,402,942-235,324,796 <i>Chr4</i> : 150,548,912-150,855,848
Random rat region 10	RNO12	18,203,110-42,525,820	<i>Chr12</i> : 110,503,298-120,870,994 <i>Chr12</i> : 121,578,435-132,335,900 <i>Chr7</i> : 66,878,689-71,941,664 <i>Chr7</i> : 101,137,811-102,184,451 <i>Chr7</i> : 99,995,220-100,350,712 <i>Chr7</i> : 72,707,443-74,223,683 <i>Chr7</i> : 75,027,443-76,145,496
Random rat region 11	RNO20	30,416,373-54,739,083	<i>Chr6</i> : 101,086,446- 116,620,662 <i>Chr6</i> : 117,266,139-123,147,126 <i>Chr2</i> : 109,065,537-109,613,060 <i>Chr6</i> : 116,688,407-116,905,609
Random rat region 12	RNO13	955,085-25,277,795	<i>Chr18</i> : 58,351,906-63,553,937 <i>Chr2</i> : 124,758,685-125,682,595
Random rat region 13	RNO1	1,136,860- 25,459,569	<i>Chr6</i> : 128,011,342-150,185,813 <i>Chr6</i> : 123,315,387-124,317,854
Random rat region 14	RNO2	182,078,762-206,401,472	<i>Chr1</i> : 107,259,608-154,441,176

number generator function of Excel. The rat genome is 2.75 Gb in size [28] or 130,952,381 bp per chromosome if divided equally across chromosomes. Therefore, values for the rat genome start-position were chosen from 1 to 130,952,381 using a random number generator. Following, 22,322,709 bps were added to each random start position to obtain the desired full length. The 14 random rat genome regions were then entered into the UCSC genome browser, and the human orthologous regions were determined using the ‘in other genomes (convert)’ function, as described above [24]. Randomly-generated rat genome segments were used as controls if the human orthologous segment did not contain a sequence that was also orthologous to a known rat mammary cancer associated locus. We also verified, using the UCSC genome browser, that human orthologous regions to random rat regions were not located at human centromeric regions, as genetic variation in these chromosomal regions is underrepresented in GWA studies [29,30]. Total sizes and percentages of rat genome covered by rat mammary cancer loci and random genome regions used are in Table 3.

Determining human GWA study nominated polymorphisms

Human breast cancer risk GWA studies considered were published through March 2013. In the first round of analysis we picked GWA studies with a clearly defined study population of subjects of European descent. In the second round of analysis, the defined population was broader and included studies that tested populations of non-European descent. Studies that included non-European descent populations were subdivided into respective populations used. The GWA studies evaluated are listed in Tables 4 and 5. Results from GWA studies used consisted of multiple stages (two to four stages) to evaluate breast cancer risk association. In our analysis, all SNPs that entered the final stage of their respective study were compared in the rat genome. A tested SNP was called either ‘associated’ if it reached genome wide significance in its respective study or ‘potentially associated’ if it failed to meet the respective study statistical criterion following the final stage of analysis. Conventionally, a *P*-value level for

an association to be considered statistically significant in a GWA study is 1×10^{-7} . This stringency is to protect from false-positives due to multiple comparisons on a genome-wide scale. It has been argued that this low *P*-value requirement results in numerous false negative associations [7]. Therefore, we queried supplemental material of each published GWA study considered and picked polymorphisms that failed the validation stage in the respective study. We also included polymorphisms that did reach genome-wide significance. We considered 533 SNPs from studies that included populations of European descent, and 285 SNPs from studies of non-European descent populations. All SNPs used in this analysis are listed in Additional file 1. Human genomic locations of polymorphisms were found using dbSNP (GRCh37 assembly) [31]. These were compared to locations of the human orthologous regions of rat mammary cancer loci and random rat regions. If a polymorphism mapped to a region of interest, the name, location, odds ratio, 95% confidence interval, and *P*-value were noted.

Statistics

The number of human polymorphisms that mapped to orthologous regions containing rat mammary cancer loci (observed) was compared to the number of human polymorphisms that mapped to random rat regions (expected) using a chi-square analysis with one degree of freedom. Several rat mammary cancer loci overlap extensively and subsequently several human polymorphisms mapped to multiple rat loci. Currently, it is not known if these overlapping rat mammary cancer loci would fine-map to the same locus or independent loci. For this study, human polymorphisms mapping to overlapping rat mammary cancer susceptibility associated sequences were counted only once. For analysis of associated (passed genome-wide significance level) versus potentially associated (did not pass genome-wide significance level) associations, a logistic regression was performed using SYSTAT 13 statistical software. A threshold of associated or potentially associated was used as the independent variable and outcome was either the SNP mapped to a rat mammary cancer locus or it mapped to a random rat region.

Table 3 Total size and percentage of rat genome covered by rat mammary cancer loci and random rat regions

Region	Loci	Total size (bases)	Total overlapping bases	Total unique bases	Rat genome portion (based on total unique bases)
<i>Mcs/Mcsm</i> only	14	345,323,605	33,002,148	312,321,457	11.4%
All known ratmammary cancer loci	24	1,230,487,116	325,386,323	905,100,793	32.9%
Random rat regions	14	312,517,940	-	312,517,940	11.4%

Table 4 Breast cancer risk genome-wide association studies using populations of European descent

GWAS	Population	Stages	Cases/controls stage 1	Cases/controls stage 2	Cases/controls stage 3	Cases/controls stage 4	Study P-value cut-off for significance
Ahmed <i>et al.</i> [32]	European descent	4	390/364	3,990/3,928	3,878/3,928	33,134/36,141	$P < E-07$
Antoniou <i>et al.</i> [33]	European descent	2	1,193/1,190	5,986/2,974			$P < E-07$
Easton <i>et al.</i> [34]	European descent	3	408/400	3,990/3,916	21,860/22,578		$P < E-07$
Fletcher <i>et al.</i> [35]	European descent	3	3,981/2,365	4,804/3,936	4,237/5,044		-
Garcia-Closas <i>et al.</i> [36]	European descent	2	4,193/35,194	6,514/41,455			$P < 5E-08$
Gaudet <i>et al.</i> [37]	European descent	2	899/804	1,264/1,222			$P < E-05$
Ghoussaini <i>et al.</i> [38]	European descent	2	56,989/58,098	69,564/68,150			$P < E-04$
Haiman <i>et al.</i> [39]	European descent/African descent	2	African descent (1,004/2,745), European descent (1,718/3,670)	European descent (2,292/16,901)			-
Hunter <i>et al.</i> [40]	European descent	2	1,145/1,142	1,776/2,072			$P < 2E-05$
Li <i>et al.</i> [41]	European descent	2	617/4,583	1,011/7,604			$P < E-05$
Li <i>et al.</i> [42]	European descent	2	2,702/ 5,726	?			$P < E-06$
Mavaddat <i>et al.</i> [43]	European descent	2	4,470/4,560	?			$P < 5E-02/6.25E-03$
Michailidou <i>et al.</i> [44]	European descent	2	10,052/12,575	45,290/41,880			$P < 5E-08$
Murabito <i>et al.</i> [45]	European descent	1	250/1,345				$P < 5E-08$
Sehrawat <i>et al.</i> [46]	European descent	2	348/348	1,153/1,215			$P < 6.4E-08$
Stacey <i>et al.</i> [47]	European descent	2	1,600/11,563	4,554/17,577			$P < E-07/P < 6.8E-08$
Stacey <i>et al.</i> [48]	European descent	2	6,145/33,016	5,028/32,090			-
Thomas <i>et al.</i> [49]	European descent	3	1,145/1,142	4,547/4,434	4,078/5,223		$P < 5E-07$
Turnbull <i>et al.</i> [50]	European descent	2	3,659/4,897	12,576/12,223			$P < E-04$

Table 5 Breast cancer risk genome-wide association studies of non-European descent populations

GWAS	Population	Stages	Cases/controls stage 1	Cases/controls stage 2	Cases/controls stage 3	Cases/controls stage 4	Study P-value cut-off for significance
Cai <i>et al.</i> [51]	Asian descent	4	2,062/2,066	4,146/1,823	6,436/6,716	4,509/6,338	-
Chen <i>et al.</i> [52]	African- American descent	2	3,153/2,831	3,607/11,330			$P < 5E-08$
Gold <i>et al.</i> [53]	Ashkenazi Jewish descent	3	249/299	950/979	243/187		$P < E-05$
Haiman <i>et al.</i> [39]	African descent/ European descent	2	African descent (1,004/2,745), European descent (1,718/3,670)	European descent (2,292/16,901)			-
Kim <i>et al.</i> [54]	Asian descent	3	2,273/2,052	2,052/2,169	1,997/1,676		$P < 5E-04$
Long <i>et al.</i> [55]	Asian descent/ European descent	3	2,073/2,084	4,425/1,915	Asian descent (6,173/6,340), European descent (2,797/2,662)		-
Long <i>et al.</i> [56]	Asian descent	4	2,918/2,324	3,972/3,852	5,203/5,138	7,489/9,934	$P < 5E-08$
Zheng <i>et al.</i> [57]	Asian descent	3	1,505/1,522	1,554,1,576	3,472/900		$P < 5E-08$
Zheng <i>et al.</i> [58]	Asian descent	2	23,637/25,579				$P < 1.5E-03$

Results

Significantly more breast cancer risk GWA study nominated SNPs are located at orthologs of rat *Mcs/Mcsm* loci compared to random rat genomic regions

We picked 28 GWA studies of breast cancer risk in which well-defined populations were analyzed (Table 4). Physical locations of polymorphisms that failed the final validation step and polymorphisms that reached genome-wide significance were determined using dbSNP [31]. We included SNPs that failed the final validation step of the respective study, because it has been suggested that many true associations are ruled out in a GWA study due to stringent statistical analysis methods [7]. We determined if sequences containing these polymorphisms were located at either a human genome region orthologous to a known rat mammary cancer locus or to a randomly selected region of the rat genome. Our goal was to determine if GWA study-nominated potentially-associated (did not pass final validation) and associated (genome-wide significant) risk polymorphisms, together map more often to human orthologous regions of rat mammary cancer susceptibility loci than to randomly selected rat genome segments of similar size. If yes, it would suggest that human GWA information combined with rat genetic susceptibility information is broadly useful to determine true genetic associations. Overall, rat *Mcs/Mcsm* loci are mapped to shorter genomic segments than other rat mammary cancer risk loci; therefore, we first compared overlap between human GWA study nominated breast cancer risk SNPs and rat *Mcs/Mcsm* loci to overlap of human associated SNPs with randomly selected rat genomic regions not known to contain mammary cancer susceptibility loci (Figure 1). Human GWA studies were grouped by population of descent for comparison. There was a significant difference between the number of GWA study nominated SNPs mapping to rat *Mcs/Mcsm* loci compared to random rat regions in studies analyzing populations of European descent (66 SNPs to 51 SNPs respectively, P -value <0.05). Human GWA study identified polymorphisms located at orthologs of rat loci are listed in Additional file 2. This result supports previous studies indicating rat genetic susceptibility is useful to predict and study human breast cancer risk loci. There was no difference in Asian or African-American descent populations. This is likely due to a limited number of published population-based breast cancer risk genetic-association studies using these populations.

Breast cancer risk GWA study nominated polymorphisms map more often to orthologs of all known rat mammary cancer loci than to randomly selected regions

Next, we included additional rat mammary cancer susceptibility loci that have been identified, but span large

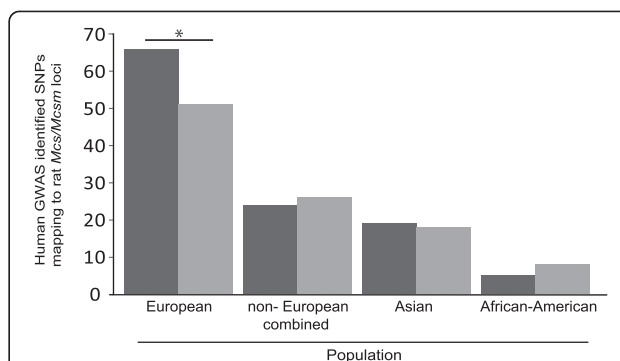
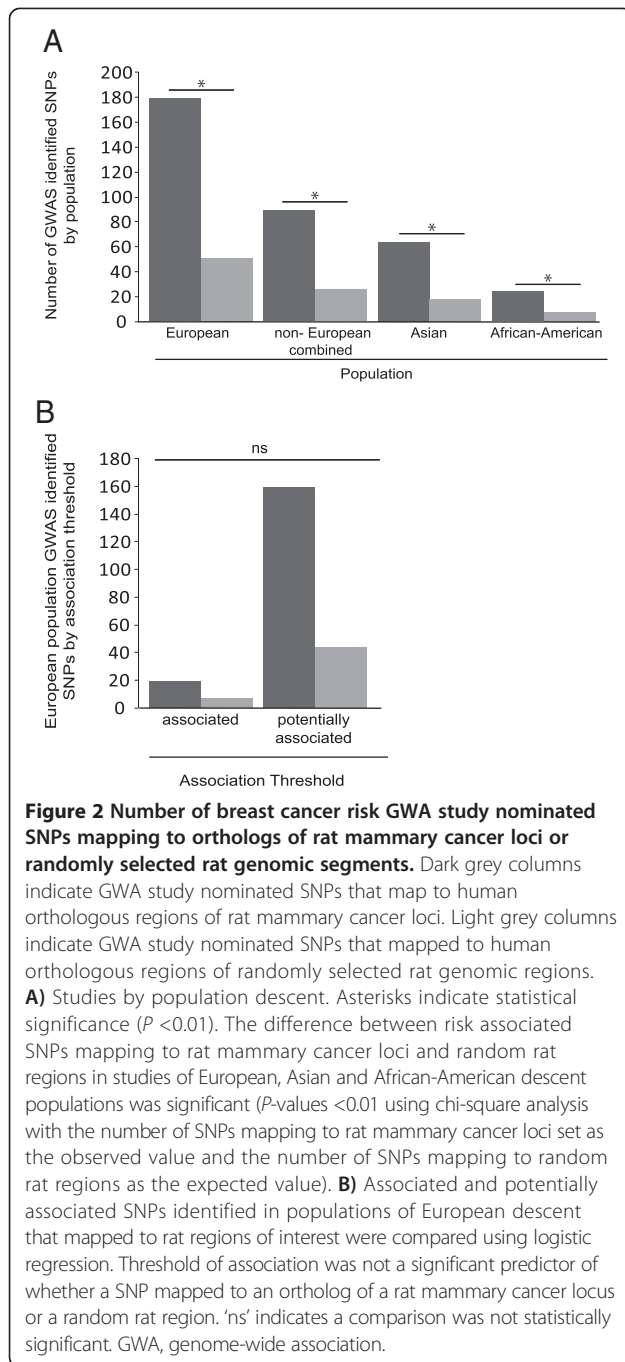


Figure 1 Number of breast cancer risk GWA study nominated SNPs mapping to rat *Mcs/Mcsm* regions. Number of GWA study nominated SNPs mapping to orthologs of rat *Mcs/Mcsm* loci and rat random regions. Dark grey columns represent the number of GWA study nominated human SNPs mapping to the human orthologous regions of the *Mcs/Mcsm* loci. Light grey columns represent the number of GWA study nominated human SNPs mapping to the human orthologous regions of the random rat control regions. The difference between risk associated SNPs mapping to rat *Mcs/Mcsm* and random rat regions was statistically significant for European populations. Asterisk indicates P -value <0.05 using chi-square analysis with number of SNPs mapping to *Mcs/Mcsm* set as the observed value and number of SNPs mapping to random rat regions as the expected value. GWA, genome-wide association.

genomic segments. Loci added were *Mcstm1*, *Mcstm2*, *Mcsta1*, *Emca1-2* and *Emca4-8* [20-23]. The same random rat genomic regions used previously were used in this analysis to be consistent. Respectively, 179 and 51 GWA study nominated polymorphisms were located in human orthologous regions to rat mammary cancer loci and randomly selected rat regions (Figure 2A) when studies using populations of European descent were considered. This difference was statistically significant ($P < 0.01$). Note, some rat mammary cancer loci identified in independent studies have long regions of overlap. Consequently, several human GWA study identified polymorphisms mapped to human sequence orthologous to overlapping rat susceptibility loci. As it is not known if these rat loci contain unique sub-loci, human risk associated polymorphisms mapping to overlapping rat regions were counted only once. The size of the rat genome covered by all known rat mammary cancer susceptibility loci compared to control loci was disproportionate (Table 3). However, the ratio of breast cancer risk associated human SNPs at orthologs to rat mammary cancer susceptibility loci to SNPs at random segments was higher than the ratio of susceptibility loci bases to random bases (3.5 versus 2.9). This result was relatively proportionate to the previous result when only rat *Mcs/Mcsm* loci were considered (1.29 for *Mcs/Mcsm* and 1.21 for all susceptibility loci), suggesting that a potential bias was not introduced by the increase in total genomic coverage.



Not surprisingly, only 179 of 533 or 33.6% of the total human GWA study identified SNPs using populations of European descent were located at orthologs to rat mammary cancer associated loci. It is notable that 57 of the 533 total SNPs evaluated were reported in more than one GWA study; a majority of these were potential associations that failed the final validation step of the respective study. These results further suggest that there are several breast cancer risk associated SNPs not reaching

genome-wide statistical significance in human population-based genetic studies.

Since more breast cancer risk polymorphisms nominated from GWA studies of populations of European descent mapped to orthologs of rat mammary cancer loci than to randomly selected regions of the rat genome, we determined if this was the case for association studies using non-European descent populations. We queried the nine GWA studies of populations of non-European ancestry that are listed in Table 5. These were GWA studies using populations of African, African-American, Ashkenazi Jewish, and Asian descent; however, only polymorphisms from studies using African-American, Ashkenazi Jewish and Asian descent populations mapped to any of the human orthologous segments to rat genomic regions picked for this study. First, results from all studies of non-European descent populations were combined (Figure 2A). Eighty-nine risk associated SNPs mapped to orthologs of rat mammary cancer loci and 26 SNPs were located at randomly selected rat regions. Next, studies using populations of Asian, Ashkenazi Jewish and African-American descent were considered separately. This resulted in 64 Asian descent population SNPs mapping to orthologs of rat mammary cancer loci and 18 SNPs to random rat regions. Twenty-four SNPs identified in studies of African-American descent populations were located at orthologs to rat mammary cancer loci and eight SNPs in random rat regions. The difference between rat mammary cancer loci and random regions was statistically significant ($P < 0.01$) for both populations (Figure 2A). Interestingly, one SNP from a study of an Ashkenazi Jewish population mapped to the human orthologous region of rat *Mcsta1*, but no GWA study nominated SNP from that study mapped to a rat random region [53]. The lack of human SNPs mapping to orthologs of rat mammary cancer loci from populations of African and Ashkenazi Jewish descent may be due to a limited number of studies conducted on these populations. On the other hand, it may indicate that susceptibility alleles different from those currently identified in laboratory rats are segregating in these populations. Out of 285 SNPs considered from studies using populations of non-European descent, 89 SNPs or 31% mapped to orthologs of rat mammary cancer loci (Additional file 2). Fifteen risk-associated SNPs were represented in more than one human GWA study.

Next, GWA-study nominated variants from populations of European descent were separated by associated (reached genome-wide significance) and potentially associated (did not reach genome-wide significance after the final stage) variants (Figure 2B). Nineteen associated SNPs were located at rat mammary cancer loci compared to seven SNPs that mapped to random rat regions. Comparatively, 160 potentially associated SNPs mapped

to rat mammary cancer susceptibility loci compared to 44 SNPs that mapped to random rat regions. A logistic regression was performed using threshold of association (associated or potentially associated) as the independent variable and rat genome location (ortholog of a rat mammary cancer risk locus or a randomly selected locus) as the dependent variable. Threshold of association was not a significant effect (P -value = 0.54). This result, that both associated and potentially associated breast cancer risk variants map more often to orthologs of rat mammary cancer risk loci than rat regions not associated with susceptibility, strongly supports that comparative genomics between humans and rats may be an effective integrative approach to determine which potential associations nominated by human association studies are true positives.

Human populations have been studied more extensively for breast cancer genetic risk than have rat populations; therefore, it is not surprising that human studies have yielded a considerable number of genome-wide significantly associated SNPs in alleles where it is not known if the rat genome contains a concordant allele. Interestingly, seven strongly associated human SNPs were in sequences orthologous to the randomly selected rat genome regions that are not known to associate with rat mammary cancer based on studies evaluating specific rat strains; thus, it is possible that a portion of the rat genome used in this study as rat random-genome control regions may actually associate with unidentified rat mammary cancer susceptibility loci. Thus, more rat genomic regions associated with mammary cancer risk may be identified with additional rat genetic studies. To date, only six inbred rat strains have been used to identify rat genomic regions associated with mammary cancer risk [15,16,20-23]. Therefore, it is highly likely that more mammary cancer susceptibility loci may be identified by incorporating additional rat strains. It is also possible that more extensive analysis of previously studied rat strains may yield additional susceptibility loci by using a higher density of genetic markers for example.

Twenty-one of the 24 known rat mammary cancer associated loci are orthologous to human loci containing SNPs that are either associated or potentially associated with breast cancer risk. Fourteen of the known rat mammary cancer associated loci are orthologous to human risk alleles marked by GWA study nominated SNPs reaching genome-wide significance. Human GWA study designs do not definitively determine causative genes or mechanisms. The laboratory rat is a versatile experimental organism to complement human studies of breast cancer. For example, inbred rat strains provide a model with reduced genetic variation that can be genetically manipulated and environmentally controlled. The overlap between human breast and rat mammary cancer

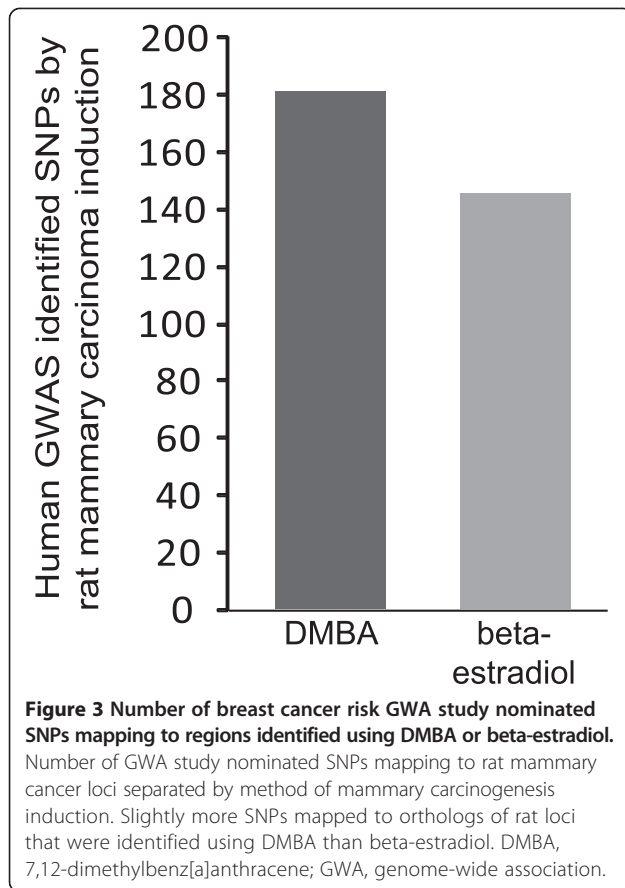
susceptibility associated loci suggests rats can be used extensively to study genetically determined mechanisms and environment interactions that will translate directly to human breast cancer risk and prevention.

Human GWA study nominated breast cancer risk SNPs map similarly to rat mammary cancer associated loci identified using 7,12-dimethylbenz[a]anthracene or beta-estradiol

Several rat mammary cancer loci used in this study were identified using DMBA to induce mammary tumors. These are *Mcs1a-c*, *Mcs2-4*, *Mcs5a1*, *Mcs5a2*, *Mcs5b-c*, *Mcs6-Mcs8*, *Mcsm1*, *Mcstm1-2* and *Mcsta1*. The remaining rat mammary cancer loci considered were identified using beta-estradiol to induce mammary carcinogenesis. Estradiol-associated susceptibility loci are *Emca1-2* and *Emca4-8*. While DMBA is representative of environmental polycyclic aromatic hydrocarbons, this synthesized mammary carcinogen is not found in nature. Conversely, estradiol is an endogenous environmental exposure associated with breast cancer risk. Human GWA study nominated SNPs mapping to orthologs of rat mammary cancer loci identified using DMBA were compared to those identified using beta-estradiol. We considered SNPs from all GWA studies, irrespective of the population used. We noted that many DMBA and beta-estradiol identified rat mammary cancer loci overlap. In fact, seven of the fourteen DMBA associated rat mammary cancer loci overlap at least one beta-estradiol associated rat mammary cancer risk locus, and five of the seven beta-estradiol loci overlap rat mammary cancer loci identified using DMBA. To account for this overlap, human SNPs mapping to overlapping rat mammary cancer loci, one identified using DMBA and the other using beta-estradiol, were included once in the 'DMBA' group and once in the 'beta-estradiol' group. These results are shown in Figure 3. A relatively similar number of GWA study nominated SNPs mapped to orthologs of rat mammary cancer loci that were identified using DMBA (181 SNPs) and beta estradiol (146 SNPs). This suggests that different mammary carcinoma induction methods can effectively identify rat susceptibility loci relevant to human disease risk, and it also suggests that a plethora of carcinogenesis mechanisms may be genetically determined.

Discussion

It has been suggested that the use of Bonferroni-based correction procedures to protect against multiple comparisons in GWA studies is too stringent and results in an abundance of false negative associations with little recourse to sort these from true-negative associations. Therefore, we considered associated and potentially associated human SNPs from breast cancer risk GWA studies to determine if SNPs that failed validation and SNPs that reached genome-wide significance map to



respective regions of the rat genome known to associate with rat mammary cancer risk more often than to regions of the rat genome that are not known to associate with susceptibility. Results presented here indicate that the rat genome is useful to prioritize and rank human alleles potentially associated with risk. The rat genome is useful regardless of the human population studied. Significantly more SNPs from GWA studies of populations of European, Asian, and African-American descent map to human orthologous regions of rat mammary cancer loci than to human orthologous regions of randomly selected rat genomic regions not known to associate with mammary cancer susceptibility. This supports the general idea that there are SNPs associated with breast cancer risk that are missed due to conservative statistical methods used in GWA studies, and that the rat is useful to parse out important genetic variation in susceptibility to mammary carcinogenesis.

Interestingly, we were unable to map GWA study nominated SNPs to 3 of the 24 known rat mammary cancer loci. These were *Mcs1a*, *Mcs5a1* and *Mcs5c*. However, using a genome-targeted population-based genetic association study, a human SNP associated with breast cancer risk has been identified at human *MCS5A1* [9]. The risk-associated SNPs at *MCS5A1* are adjacent to a breast

cancer risk-associated SNP at *MCS5A2*, which was identified in two independent human population based studies [9,43]. Taken together, there is a high correlation between genetics of breast cancer susceptibility in humans and mammary cancer susceptibility in rats. Interestingly, there are several human genomic regions that are human GWA study nominated hotspots (for example, 19q13, *FGFR2*) that are not known to have concordant rat orthologs. An explanation is that human breast and rat mammary cancer susceptibility are controlled by overlapping and non-overlapping genetic mechanisms. Another explanation is that there are rat genomic regions associated with mammary cancer risk yet to be discovered by using additional inbred strains, more extensive analysis of strains previously studied, and different methods of carcinogenesis induction.

Conclusions

There is extensive genomic overlap between human breast and rat mammary cancer susceptibility. The rat genome may provide utility to identify true-positive associations regardless of the human population used for a GWA study. The laboratory rat will continue to be an important model organism for researching genetically determined mechanisms of mammary cancer susceptibility that may translate directly to human susceptibility. An appreciable number of GWA study nominated SNPs not meeting genome-wide significance levels have genomic overlap with rat mammary cancer susceptibility loci. This supports the general idea that Bonferroni-based multiple-comparison correction procedures are too stringent and complementary approaches that integrate rat genomics would be highly efficacious to prioritize breast cancer risk associated alleles.

Additional files

Additional file 1: Table S1. List of GWAS nominated SNPs used in Analysis.

Additional file 2: Table S2. Breast cancer risk associated polymorphisms from studies of European descent populations that map to rat mammary cancer loci and random rat regions. **Table S3.** Breast cancer risk associated polymorphisms from studies of non-European descent populations that map to rat mammary cancer loci and random rat regions.

Abbreviations

ACI: August Copenhagen Irish; BN: Brown Norway; bp: base pair; BRCA 1 or 2: breast cancer 1 or 2; COP: Copenhagen; DMBA: 7,12-dimethylbenz[a]anthracene; EMCA: estrogen-induced mammary cancer; GWA study: genome-wide association study; Mcs: mammary carcinoma susceptibility; Mcsm: mammary carcinoma susceptibility modifier; Mcsta: mammary cancer susceptibility tumor aggressiveness; Mcstm: mammary cancer susceptibility tumor multiplicity; NMU: *N*-nitroso-*N*-methylurea; PALB 2: partner and localizer of BRCA2; QTL: quantitative trait locus; SNP: single nucleotide polymorphism; WF: Wistar-Furth; WKY: Wistar-Kyoto.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

JS and DJS contributed to study conception and design, data analysis, and interpretation of results. JS and DJS drafted this manuscript and revised it. Both authors read and approved the final manuscript.

Acknowledgements

This work was supported by grant CA137052 from the National Institutes of Health. JS was supported by the University of Louisville Integrated Programs in Biomedical Sciences and NIH/NIEHS training grant T32-ES011564.

Author details

¹Department of Biochemistry and Molecular Biology, University of Louisville School of Medicine, Louisville, KY 40292, USA. ²James Graham Brown Cancer Center, University of Louisville School of Medicine, Louisville, KY 40292, USA. ³Center for Genetics and Molecular Medicine, University of Louisville School of Medicine, Louisville, KY 40292, USA. ⁴Health Sciences Center, University of Louisville School of Medicine, Louisville, KY 40292, USA.

Received: 8 July 2013 Accepted: 10 January 2014

Published: 27 January 2014

References

- Hemminki K, Vaittinen P, Kyriönen P: **Age-specific familial risks in common cancers of the offspring.** *Int J Cancer* 1998, **78**:172–175.
- Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A, Hemminki K: **Environmental and heritable factors in the causation of cancer — analyses of cohorts of twins from Sweden, Denmark, and Finland.** *New Engl J Med* 2000, **343**:78–85.
- Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, Collins N, Gregory S, Gumbs C, Micklem G: **Identification of the breast cancer susceptibility gene BRCA2.** *Nature* 1995, **378**:789–792.
- Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, King MC: **Linkage of early-onset familial breast cancer to chromosome 17q21.** *Science* 1990, **250**:1684–1689.
- Mavaddat N, Antoniou AC, Easton DF, Garcia-Closas M: **Genetic susceptibility to breast cancer.** *Mol Oncol* 2010, **4**:174–191.
- Hindorf LA, Gillanders EM, Manolio TA: **Genetic architecture of cancer and other complex diseases: lessons learned and future directions.** *Carcinogenesis* 2011, **32**:945–954.
- Thomas DC, Haile RW, Duggan D: **Recent developments in genomewide association scans: a workshop summary and review.** *Am J Human Genet* 2005, **77**:337–345.
- Ioannidis JP, Castaldi P, Evangelou E: **A compendium of genome-wide associations for cancer: critical synopsis and reappraisal.** *J Natl Cancer Inst* 2010, **102**:846–858.
- Samuelson DJ, Hesselson SE, Aperavich BA, Zan Y, Haag JD, Trentham-Dietz A, Hampton JM, Mau B, Chen K-S, Baynes C, Khaw KT, Luben R, Perkins B, Shah M, Pharoah PD, Dunning AM, Easton DF, Ponder BA, Gould MN: **Rat Mcs5a is a compound quantitative trait locus with orthologous human loci that associate with breast cancer risk.** *Proc Natl Acad Sci U S A* 2007, **104**:6299–6304.
- Russo IH, Russo J: **Role of hormones in mammary cancer initiation and progression.** *J Mammary Gland Biol Neoplasia* 1998, **3**:49–61.
- Russo J, Trait L, Russo I: **Susceptibility of the mammary gland to carcinogenesis. III. The cell of origin of rat mammary carcinoma.** *Am J Pathol* 1983, **113**:50–66.
- Shull JD: **The rat oncogene: comparative genetics and genomics of rat models of mammary carcinogenesis.** *Breast Dis* 2007, **28**:69–86.
- Gould MN: **Rodent models for the study of etiology, prevention and treatment of breast cancer.** *Semin Cancer Biol* 1995, **6**:147–152.
- Russo J, Russo IH: **Atlas and histologic classification of tumors of the rat mammary gland.** *J Mammary Gland Biol Neoplasia* 2000, **5**:187–200.
- Shepel LA, Lan H, Haag JD, Brasic GM, Gheen ME, Simon JS, Hoff P, Newton MA, Gould MN: **Genetic identification of multiple loci that control breast cancer susceptibility in the rat.** *Genetics* 1998, **149**:289–299.
- Lan H, Kendziorski CM, Haag JD, Shepel LA, Newton MA, Gould MN: **Genetic loci controlling breast cancer susceptibility in the Wistar-Kyoto rat.** *Genetics* 2001, **157**:331–339.
- Haag JD, Shepel LA, Kolman BD, Monson DM, Benton ME, Watts KT, Waller JL, Lopez-Guajardo CC, Samuelson DJ, Gould MN: **Congenic rats reveal three independent Copenhagen alleles within the mcs1 quantitative trait locus that confer resistance to mammary cancer.** *Cancer Res* 2003, **63**:5808–5812.
- Sanders J, Haag JD, Samuelson DJ: **Physical confirmation and mapping of overlapping rat mammary carcinoma susceptibility QTLs, Mcs2 and Mcs6.** *PLoS One* 2011, **6**:e19891.
- Samuelson DJ, Aperavich BA, Haag JD, Gould MN: **Fine mapping reveals multiple loci and a possible epistatic interaction within the mammary carcinoma susceptibility quantitative trait locus, Mcs5.** *Cancer Res* 2005, **65**:9637–9642.
- Stieber D, Piessevaux G, Rivière M, Laes JF, Quan X, Szpirer J, Szpirer C: **Isolation of two regions on rat chromosomes 5 and 18 affecting mammary cancer susceptibility.** *Int J Cancer* 2007, **120**:1678–1683.
- Piessevaux G, Lella V, Rivière M, Stieber D, Drèze P, Szpirer J, Szpirer C: **Contrasting epistatic interactions between rat quantitative trait loci controlling mammary cancer development.** *Mamm Genome* 2009, **20**:43–52.
- Gould KA, Tochacek M, Schaffer BS, Reindl TM, Murrin CR, Lachel CM, VanderWoude EA, Pennington KL, Flood LA, Bynote KK, Meza JL, Newton MA, Shull JD: **Genetic determination of susceptibility to estrogen-induced mammary cancer in the ACI Rat: mapping of Emca1 and Emca2 to chromosomes 5 and 18.** *Genetics* 2004, **168**:2113–2125.
- Schaffer BS, Lachel CM, Pennington KL, Murrin CR, Strecker TE, Tochacek M, Gould KA, Meza JL, McComb RD, Shull JD: **Genetic bases of estrogen-induced tumorigenesis in the rat: mapping of loci controlling susceptibility to mammary cancer in a brown Norway × ACI Intercross.** *Cancer Res* 2006, **66**:7793–7800.
- Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D: **The human genome browser at UCSC.** *Genome Res* 2002, **12**:996–1006.
- den Dekker 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 risk locus at human 5q11.2 and MIER3 is a candidate cancer susceptibility gene.** *Cancer Res* 2012, **72**:6002–6012.
- Veillet AL, Haag JD, Remfert JL, Meilahn AL, Samuelson DJ, Gould MN: **Mcs5c: a mammary carcinoma susceptibility locus located in a gene desert that associates with tenascin C expression.** *Cancer Prev Res (Phila)* 2011, **4**:97–106.
- Cotroneo MS, Merry GM, Haag JD, Lan H, Shepel LA, Gould MN: **The Mcs7 quantitative trait locus is associated with an increased susceptibility to mammary cancer in congenic rats and an allele-specific imbalance.** *Oncogene* 2006, **25**:5011–5017.
- Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergrin EJ, Scherer S, Scott G, Steffen D, Worley KC, Burch PE, Okwuonu G, Hines S, Lewis L, DeRamo C, Delgado O, Dugan-Rocha S, Miner G, Morgan M, Hawes A, Gill R, Celera, Holt RA, Adams MD, Amanatides PG, Baden-Tillson H, Barnstead M, Chin S, Evans CA, Ferriera S, Fosler C, et al: **Genome sequence of the Brown Norway rat yields insights into mammalian evolution.** *Nature* 2004, **428**:493–521.
- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, et al: **The sequence of the human genome.** *Science* 2001, **291**:1304–1351.
- Lander E, Lauren L, Bruce B, Chad N, Michael Z, Jennifer B, Keri D, Ken D, Michael D, Fitz HW, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczyk J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, et al: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409**:860.
- Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K: **dbSNP: the NCBI database of genetic variation.** *Nucleic Acids Res* 2001, **29**:308–311.
- Ahmed S, Thomas G, Ghousaini M, Healey CS, Humphreys MK, Platte R, Morrison J, Maranian M, Pooley KA, Luben R, Eccles D, Evans DG, Fletcher O, Johnson N, dos Santos Silva I, Peto J, Stratton MR, Rahman N, Jacobs K, Prentice R, Anderson GL, Rajkovic A, Curb JD, Ziegler RG, Berg CD, Buys SS, McCarty CA, Feigelson HS, Calle EE, et al: **Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2.** *Nat Genet* 2009, **41**:585–590.
- Antoniou AC, Wang X, Fredericksen ZS, McGuffog L, Tarrell R, Sinilnikova OM, Healey S, Morrison J, Kartsonaki C, Lesnick T, Ghousaini M, Barrowdale D, EMBRACE, Peock S, Cook M, Oliver C, Frost D, Eccles D, Evans DG, Eeles R, Izatt L, Chu C, Douglas F, Paterson J, Stoppa-Lyonnet D, Houdayer C,

- Mazoyer S, Giraud S, Lasset C, Remenieras A, et al: **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* 2010, **42**:885–892.
34. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, Struwing JP, Morrison J, Field H, Luben R, Wareham N, Ahmed S, Healey CS, Bowman R, SEARCH collaborators, Meyer KB, Haiman CA, Kolonel LK, Henderson BE, Le Marchand L, Brennan P, Sangrajrang S, Gaborieau V, Odefrey F, Shen CY, Wu PE, Wang HC, Eccles D, Evans DG, Peto J, et al: **Genome-wide association study identifies novel breast cancer susceptibility loci.** *Nature* 2007, **447**:1087–1093.
35. Fletcher O, Johnson N, Orr N, Hosking FJ, Gibson LJ, Walker K, Zelenika D, Gut I, Heath S, Palles C, Coupland B, Broderick P, Schoemaker M, Jones M, Williamson J, Chilcott-Burns S, Tomczyk K, Simpson G, Jacobs KB, Chanock SJ, Hunter DJ, Tomlinson IP, Swerdlow A, Ashworth A, Ross G, dos Santos Silva I, Lathrop M, Houlston RS, Peto J: **Novel breast cancer susceptibility locus at 9q31.2: results of a genome-wide association study.** *J Natl Cancer Inst* 2011, **103**:425–435.
36. Garcia-Closas M, Couch FJ, Lindstrom S, Michailidou K, Schmidt MK, Brook MN, Orr N, Rhee SK, Riboli E, Feigelson HS, Le Marchand L, Buring JE, Eccles D, Miron P, Fasching PA, Brauch H, Chang-Claude J, Gaudin A, Gagnon AK, Nevanlinna H, Giles GG, Cox A, Hopper JL, Bolla MK, Wang Q, Dennis J, Dicks E, Howat WJ, Schoof N, Bojesen SE, et al: **Genome-wide association studies identify four ER negative-specific breast cancer risk loci.** *Nat Genet* 2013, **45**:392–398.
37. Gaudet MM, Kirchoff T, Green T, Vijai J, Korn JM, Guiducci C, Segre AV, McGee K, McGuffog L, Kartsonaki C, Morrison J, Healey S, Sinilnikova OM, Stoppa-Lyonnet D, Mazoyer S, Gauthier-Villars M, Sobol H, Longy M, Frenay M, GEMO Study Collaborators, Hogeworst FB, Rookus MA, Collée JM, Hoogerbrugge N, van Roozendaal KE, Study Collaborators HEBON, Piedmonte M, Rubinstein W, Nerenstone S, Van Le L, et al: **Common genetic variants and modification of penetrance of BRCA2-associated breast cancer.** *PLoS Genet* 2010, **6**:e1001183.
38. Ghousaini M, Fletcher O, Michailidou K, Turnbull C, Schmidt MK, Dicks E, Dennis J, Wang Q, Humphreys MK, Luccarini C, Baynes C, Conroy D, Maranian M, Ahmed S, Driver K, Johnson N, Orr N, dos Santos Silva I, Waisfisz Q, Meijers-Heijboer H, Uitterlinden AG, Rivadeneira F, Netherlands Collaborative Group on Hereditary Breast and Ovarian Cancer (HEBON), Hall P, Czene K, Irwanto A, Liu J, Nevanlinna H, Aittomäki K, Blomqvist C, et al: **Genome-wide association analysis identifies three new breast cancer susceptibility loci.** *Nat Genet* 2012, **44**:312–318.
39. Haiman CA, Chen GK, Vachon CM, Canzian F, Dunning A, Millikan RC, Wang X, Ademyiwa F, Ahmed S, Ambrosone CB, Baglietto L, Balleine R, Bandera EV, Beckmann MW, Berg CD, Bernstein L, Blomqvist C, Blot WJ, Brauch H, Buring JE, Carey LA, Carpenter JE, Chang-Claude J, Chanock SJ, Chasman DJ, Clarke CL, Cox A, Cross SS, Deming SL, Diasio RB, et al: **A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer.** *Nat Genet* 2011, **43**:1210–1214.
40. Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, Wacholder S, Wang Z, Welch R, Hutchinson A, Wang J, Yu K, Chatterjee N, Orr N, Willett WC, Colditz GA, Ziegler RG, Berg CD, Buys SS, McCarty CA, Feigelson HS, Calle EE, Thun MJ, Hayes RB, Tucker M, Gerhard DS, Fraumeni Jr JF, Hoover RN, Thomas G, Chanock SJ: **A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer.** *Nat Genet* 2007, **39**:870–874.
41. Li J, Humphreys K, Darabi H, Rosin G, Hannelius U, Heikkinen T, Aittomäki K, Blomqvist C, Pharoah P, Dunning A, Ahmed S, Hooning MJ, Hollestelle A, Oldenburg RA, Alfredsson L, Palotie A, Peltonen-Palotie L, Irwanto A, Low HQ, Teoh GH, Thalamuthu A, Kere J, D'Amato M, Easton DF, Nevanlinna H, Liu J, Czene K, Hall P: **A genome-wide association scan on estrogen receptor-negative breast cancer.** *Breast Cancer Res* 2010, **12**:R93.
42. Li J, Humphreys K, Heikkinen T, Aittomäki K, Blomqvist C, Pharoah P, Dunning A, Ahmed S, Hooning MJ, Martens J, van den Ouweland AM, Alfredsson L, Palotie A, Peltonen-Palotie L, Irwanto A, Low HQ, Teoh GH, Thalamuthu A, Easton DF, Nevanlinna H, Liu J, Czene K, Hall P: **A combined analysis of genome-wide association studies in breast cancer.** *Breast Cancer Res Treat* 2011, **126**:717–727.
43. Mavaddat N, Dunning AM, Ponder BA, Easton DF, Pharoah PD: **Common genetic variation in candidate genes and susceptibility to subtypes of breast cancer.** *Cancer Epidemiol Biomarkers Prev* 2009, **18**:255–259.
44. Michailidou K, Hall P, Gonzalez-Neira A, Ghousaini M, Dennis J, Milne RL, Schmidt MK, Chang-Claude J, Bojesen SE, Bolla MK, Wang Q, Dicks E, Lee A, Turnbull C, Rahman N, Breast and Ovarian Cancer Susceptibility Collaboration, Fletcher O, Peto J, Gibson L, Dos Santos Silva I, Nevanlinna H, Muranen TA, Aittomäki K, Blomqvist C, Czene K, Irwanto A, Liu J, Waisfisz Q, Meijers-Heijboer H, Adank M: **Large-scale genotyping identifies 41 new loci associated with breast cancer risk.** *Nat Genet* 2013, **45**:353–361.
45. Murabito J, Rosenberg C, Finger D, Kreger B, Levy D, Splansky G, Antman K, Hwang SJ: **A genome-wide association study of breast and prostate cancer in the NHLBI's Framingham Heart Study.** *BMC Med Genet* 2007, **8**:S6.
46. Sehrawat B, Sridharan M, Ghosh S, Robson P, Cass C, Mackey K, Greiner R, Damaraju S: **Potential novel candidate polymorphisms identified in genome-wide association study for breast cancer susceptibility.** *Hum Genet* 2011, **130**:529–537.
47. Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, Gudjonsson SA, Masson G, Jakobsdottir M, Thorlacius S, Helgason A, Aben KK, Strobbe LJ, Albers-Akkers MT, Swinkels DW, Henderson BE, Kolonel LN, Le Marchand L, Millastre E, Andres R, Godino J, Garcia-Prats MD, Polo E, Tres A, Mouy M, Saemundsdottir J, Backman VM, Gudmundsson L, Kristjansson K, Bergthorsson JT, Kostic J, et al: **Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer.** *Nat Genet* 2007, **39**:865–869.
48. Stacey SN, Manolescu A, Sulem P, Thorlacius S, Gudjonsson SA, Jonsson GF, Jakobsdottir M, Bergthorsson JT, Gudmundsson J, Aben KK, Strobbe LJ, Albers-Akkers MT, Swinkels DW, Henderson BE, Kolonel LN, Le Marchand L, Millastre E, Andres R, Godino J, Garcia-Prats MD, Polo E, Tres A, Mouy M, Saemundsdottir J, Backman VM, Gudmundsson L, Kristjansson K, Bergthorsson JT, Kostic J, Frigge ML, et al: **Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer.** *Nat Genet* 2008, **40**:703–706.
49. Thomas G, Jacobs KB, Kraft P, Yeager M, Wacholder S, Cox DG, Hankinson SE, Hutchinson A, Wang Z, Yu K, Chatterjee N, Garcia-Closas M, Gonzalez-Bosquet J, Prokunina-Olsson L, Orr N, Willett WC, Colditz GA, Ziegler RG, Berg CD, Buys SS, McCarty CA, Feigelson HS, Calle EE, Thun MJ, Diver R, Prentice R, Jackson R, Kooperberg C, Chlebowski R, Lissowska J, et al: **A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1).** *Nat Genet* 2009, **41**:579–584.
50. Turnbull C, Ahmed S, Morrison J, Pernet D, Renwick A, Maranian M, Seal S, Ghousaini M, Hines S, Healey CS, Hughes D, Warren-Perry M, Tapper W, Eccles D, Evans DG, Breast Cancer Susceptibility Collaboration (UK), Hooning M, Schutte M, van den Ouweland A, Houlston R, Ross G, Langford C, Pharoah PD, Stratton MR, Dunning AM, Rahman N, Easton DF: **Genome-wide association study identifies five new breast cancer susceptibility loci.** *Nat Genet* 2010, **42**:504–507.
51. Cai Q, Long J, Lu W, Qu S, Wen W, Kang D, Lee JY, Chen K, Shen H, Shen CY, Sung H, Matsuo K, Haiman CA, Khoo US, Ren Z, Iwasaki M, Gu K, Xiang YB, Choi JY, Park SK, Zhang L, Hu Z, Wu PE, Noh DY, Tajima K, Henderson BE, Chan KY, Su F, Kasuga Y, Wang W, et al: **Genome-wide association study identifies breast cancer risk variant at 10q21.2: results from the Asia Breast Cancer Consortium.** *Hum Mol Genet* 2011, **20**:4991–4999.
52. Chen F, Chen G, Stram D, Millikan R, Ambrosone C, John E, Bernstein L, Zheng W, Palmer J, Hu J, Rebbeck TR, Ziegler RG, Nyante S, Bandera EV, Ingles SA, Press MF, Ruiz-Narvaez EA, Deming SL, Rodriguez-Gil JL, Demichele A, Chanock SJ, Blot W, Signorello L, Cai Q, Li G, Long J, Huo D, Zheng Y, Cox NJ, Olopade OI, et al: **A genome-wide association study of breast cancer in women of African ancestry.** *Hum Genet* 2013, **132**:39–48.
53. Gold B, Kirchoff T, Stefanov S, Lautenberger J, Viale A, Garber J, Friedman E, Narod S, Olshen AB, Gregersen P, Kosarin K, Olsh A, Bergeron J, Ellis NA, Klein RJ, Clark AG, Norton L, Dean M, Boyd J, Offit K: **Genome-wide association study provides evidence for a breast cancer risk locus at 6q22.33.** *Proc Natl Acad Sci U S A* 2008, **105**:4340–4345.
54. Kim HC, Lee JY, Sung H, Choi JY, Park S, Lee KM, Kim Y, Go M, Li L, Cho Y, Park M, Kim DJ, Oh JH, Kim JW, Jeon JP, Jeon SY, Min H, Kim HM, Park J, Yoo KY, Noh DY, Ahn SH, Lee MH, Kim SW, Lee JW, Park BW, Park WY, Kim EH, Kim MK, Han W, et al: **A genome-wide association study identifies a breast cancer risk variant in ERBB4 at 2q34: results from the Seoul Breast Cancer Study.** *Breast Cancer Res* 2012, **14**:R56.
55. Long J, Cai Q, Shu XO, Qu S, Li C, Zheng Y, Gu K, Wang W, Xiang YB, Cheng J, Chen K, Zhang L, Zheng H, Shen CY, Huang CS, Hou MF, Shen H, Hu Z, Wang F, Deming SL, Kelley MC, Shrubsole MJ, Khoo US, Chan KY, Chan SY, Haiman CA, Henderson BE, Le Marchand L, Iwasaki M, Kasuga Y, et al: **Identification of a functional genetic variant at 16q12.1 for breast cancer risk: results from the Asia Breast Cancer Consortium.** *PLoS Genet* 2010, **6**:e1001002.

56. Long J, Cai Q, Sung H, Shi J, Zhang B, Choi J, Wen W, Delahanty RJ, Lu W, Gao YT, Shen H, Park SK, Chen K, Shen CY, Ren Z, Haiman CA, Matsuo K, Kim MK, Khoo US, Iwasaki M, Zheng Y, Xiang YB, Gu K, Rothman N, Wang W, Hu Z, Liu Y, Yoo KY, Noh DY, Han BG, *et al*: **Genome-wide association study in East Asians identifies novel susceptibility loci for breast cancer.** *PLoS Genet* 2012, **8**:e1002532.
57. Zheng W, Long J, Gao YT, Li C, Zheng Y, Xiang YB, Wen W, Levy S, Deming SL, Haines JL, Gu K, Fair AM, Cai Q, Lu W, Shu XO: **Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1.** *Nat Genet* 2009, **41**:324–328.
58. Zheng W, Zhang B, Cai Q, Sung H, Michailidou K, Shi J, Choi JY, Long J, Dennis J, Humphreys MK, Wang Q, Lu W, Gao YT, Li C, Cai H, Park SK, Yoo KY, Noh DY, Han W, Dunning AM, Benitez J, Vincent D, Bacot F, Tessier D, Kim SW, Lee MH, Lee JW, Lee JY, Xiang YB, Zheng Y, *et al*: **Common genetic determinants of breast-cancer risk in East Asian women: a collaborative study of 23 637 breast cancer cases and 25 579 controls.** *Hum Mol Genet* 2013, **22**:2539–2550.

doi:10.1186/bcr3607

Cite this article as: Sanders and Samuelson: Significant overlap between human genome-wide association-study nominated breast cancer risk alleles and rat mammary cancer susceptibility loci. *Breast Cancer Research* 2014 **16**:R14.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

