(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 September 2006 (21.09.2006)

PCT

(10) International Publication Number
WO 2006/099329 A2

(51) International Patent Classification: Not classified

(21) International Application Number:
PCT/US2006/008958

(22) International Filing Date: 14 March 2006 (14.03.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/661,420 14 March 2005 (14.03.2005) US
60/739,798 23 November 2005 (23.11.2005) US

(71) Applicant *(for all designated States except US)*: **WISCONSIN ALUMNI RESEARCH FOUNDATION** [US/US]; 614 Walnut Street, Madison, Wisconsin 53705 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **GOULD, Michael, N.** [US/US]; 13 South Blackhawk Avenue, Madison, Wisconsin 53705 (US). **HAAG, Jill, D.** [US/US]; 503 Ethelwyn Circle, Mt. Horeb, Wisconsin 53572 (US). **SAMUELSON, David, J.** [US/US]; 5713 South Hill Drive, Madison, Wisconsin 53705 (US). **NELSON, Stephanie, E.** [US/US]; 2302 University Avenue, Apt. 205, Madison, Wisconsin 53726 (US).

(74) Agent: **BAKER, Jean, C.**; QUARLES & BRADY LLP, 411 E. Wisconsin Avenue, Milwaukee, Wisconsin 53202 (US).

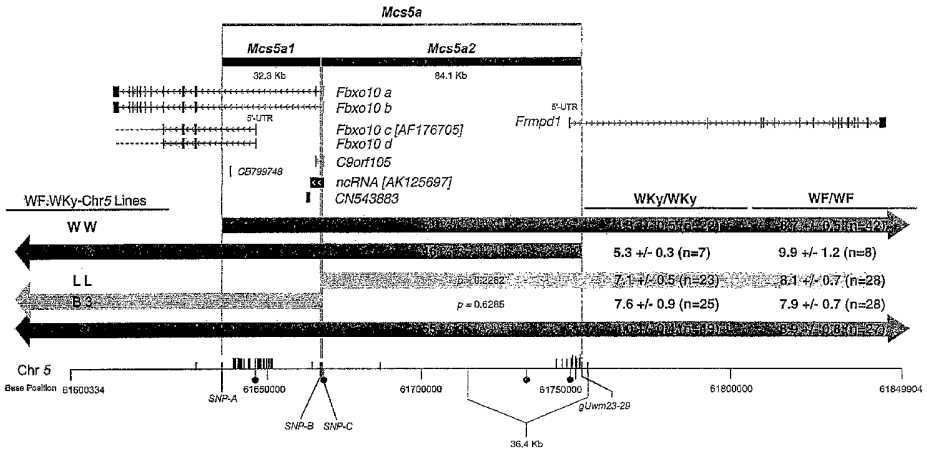
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) **Title:** METHOD OF ANALYZING BREAST CANCER SUSCEPTIBILITY AND RESISTANCE



(57) **Abstract:** A method of analyzing breast cancer susceptibility or resistance, comprising the steps of obtaining a DNA sample from a patient, genotyping the sample for the presence of the minor allele of single nucleotide polymorphisms rs6476643 or rs2182317, and correlating the identity of the allele with risk assessment data is disclosed.

WO 2006/099329 A2

METHOD OF ANALYZING BREAST CANCER SUSCEPTIBILITY AND RESISTANCE

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This application claims the benefit of U.S. 60/739,798, filed November 23, 2005 and 60/661,420, filed March 14, 2005. Both provisional applications are incorporated by reference herein as if fully set forth below.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

[0002] The invention was made with United States government support awarded by the following agency: NIH Grants: CA77494 and CA28954. The United States has certain rights in this invention.

BACKGROUND OF THE INVENTION

[0003] Despite immense efforts, the search for modifier genes underlying complex diseases has not been highly productive. Alleles of modifier genes that influence disease risk have a moderate to high population frequency with a low penetrance. It has been suggested that alleles acting in this manner comprise the majority of genetic risk for many common diseases, such as breast cancer (D. E. Reich and E. S. Lander, *Trends Genetics* 17, 502 (2001), B. A. Ponder, *Nature* 411, 336 (2001)). It is estimated that if most risk alleles are identified, it would be possible to assign ~90% of breast cancer risk to 50% of women (P. D. P. Pharoah *et al.*, *Nature Genet.* 31, 33 (2002)). In most studies, candidate modifier genes are selected based on their function, such as DNA repair or estrogen metabolism for breast cancer, and thus are classified as biased.

Over 100 biased candidate modifier genes have been tested in breast cancer case-control association studies (>400 SNPs); few, if any, show a consistent and significant association with risk in large sample populations (P. D. P. Pharoah, A. M. Dunning, B. A. J. Ponder, D. F. Easton, *Nature* 4, 850 (2004)). This suggests the need for an unbiased strategy to identify breast cancer modifier genes.

[0004] Previously, our laboratory has pursued the identification of unbiased candidate loci by using whole genome linkage studies in inbred-rat mammary cancer models followed by fine-mapping in congenic rats. Polymorphisms in human genomic regions orthologous to these rat loci can be subsequently tested for association with breast cancer risk in large case-control population-based studies. In this application we present an example that supports this unbiased approach.

[0005] Using a backcross of [Wistar-Kyoto (WKy) x Wistar-Furth (WF)]F_1 x WF rats, we previously identified four mammary carcinoma susceptibility QTL, *Mcs5*, *Mcs6*, *Mcs7*, and *Mcs8*, on rat chromosomes *5*, *7*, *10*, and *14*, respectively (H. Lan *et al.*, *Genetics* 157, 331 (2001)). The WKy allele of *Mcs5* acts to suppress mammary tumor multiplicity in a susceptible WF genetic background. To narrow the *Mcs5* QTL region, we collected multiple WF.WKy congenic recombinant lines within the *Mcs5* locus. Selected lines were phenotyped to determine their susceptibility to 7-12 dimethylbenz(a)anthracene (DMBA)-induced mammary carcinogenesis.

[0006] Data from these congenic rat lines containing various WKy genomic intervals of the *Mcs5* QTL region identified three *Mcs5* subloci. Congenic WKy-homozygous and heterozygous females from a line defining one sublocus, *Mcs5a*, had the phenotype of resistance to DMBA-induced mammary cancer (D. J. Samuelson, B. A. Aperavich, J. D. Haag, M. N. Gould, *Cancer Res.* 65, 9637 (2005)).

[0007] Needed in the art of breast cancer detection and therapy are specific minor alleles that correlate with breast cancer susceptibility or resistance.

BRIEF SUMMARY OF THE INVENTION

[0008] In one embodiment, the present invention is a method of analyzing breast cancer susceptibility or resistance, comprising the steps of obtaining a DNA sample from a patient, genotyping the sample for the presence of the minor allele of single nucleotide polymorphisms rs6476643 or rs2182317 or a polymorphism correlating at r^2 >0.7 with rs6476643 or rs2182317, and correlating the identity of the allele with risk assessment data.

[0009] Preferably, the minor allele is a minor allele selected from the group consisting of

rs2182317
rs6476643
138-9899
rs10758441
rs7042509
114-117
d3-169
rs12378421
r3-116
rs17505776
rs4878708
rs4878709
rs4878710
rs10973450
l4-70
m4-218
rs4490927
x4-77
z4-66
f5-152
rs4878713
y5-43
i6-31
i6-103

[0010] In another embodiment, the present invention is a method of finding minor alleles that correlate to human breast cancer susceptibility or resistance, comprising the steps of sequencing human genomic regions MCS5A1 or MCS5A2, identifying at least one polymorphism, genotyping individuals for the presence or absence of the polymorphism and, determining whether the polymorphism correlates to rs212317 (plus strand allele T) or rs6476643 (plus strand allele T) with the r^2 statistic at $r^2 > 0.7$.

[0011] In another embodiment, one determines whether the polymorphism correlates at $r^2 > 0.7$ to the group consisting of

rs2182317
rs6476643
138-9899
rs10758441
rs7042509
114-117
d3-169
rs12378421
r3-116
rs17505776
rs4878708
rs4878709
rs4878710
rs10973450
l4-70
m4-218
rs4490927
x4-77
z4-66
f5-152
rs4878713
y5-43
i6-31
i6-103

[0012] In another embodiment, the present invention is a method of treating or preventing breast cancer comprising mimicking the biological effect of *MCS5A1 and MCS5A2* alleles that increase/decrease susceptibility to breast cancer.

[0013] In another embodiment, the present invention is a method of assaying molecules for their ability to interact with the *FBXO10* and *FMRPD1* gene loci or the *FBXO10* and *FMRPD1* gene loci products comprising the steps of exposing a candidate molecule to the *FBXO10* and *FMRPD1* gene loci or gene loci products and determining whether the candidate molecule has interacted with the loci or the gene products.

[0014] Other objects, features and advantages of the present invention will be apparent to one of skill after review of the specification, claims and drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0015] **Fig. 1. Positional-mapping of synthetically interacting gene loci within the compound rat QTL, *Mcs5a*.** *Mcs5a* (~116 Kb) is a compound QTL located on rat chromosome *5*. Adjacent *Mcs5a1* (32.3 Kb) and *Mcs5a2* (84.1 Kb), are synthetically interacting loci within *Mcs5a* and were identified using recombinant congenic lines. Rat mammary gland transcripts are mapped as exons connected by intronic sequence. *Mcs* phenotypes of congenic lines (WF.WKy-Chr5), containing different segments of the *Mcs5a* WKy allele of congenic line O, were determined in female rats administered DMBA (65 mg/kg) at 50-55 days of age. Tumor multiplicity data presented are average number of mammary carcinomas ≥ 3 x 3 mm per rat ± S.E at 15 wks post-induction. Congenic homozygous (WKy/WKy) females were compared to WF/WF littermates using the nonparametric Mann-Whitney test. Percent reduction in mammary carcinomas/rat compared to *Mcs* – WF/WF littermates is shown at center for

lines determined to be significantly different ($P < 0.05$) than the *Mcs* – WF/WF rats. The *Mcs* resistant and *Mcs* susceptible congenic lines that define *Mcs5a* are shown by the WKy alleles they contain. The relevant end of the WKy sequence in each congenic line is indicated by a genetic marker and a dashed-vertical line. *SNP-A*, *SNP-B*, *SNP-C*, and *gUwm23-29*, are on chromosome *5* at base positions: 61,634,906; 61,667,232; 61,667,646; and 61,751,595-61,751,793; respectively. The x-axis represents rat chromosome *5* and the designated marks are SNPs, INDELs, microsatellites, CpG islands, and the UCSC Genome Browser June 2003 rat assembly base position. The 36.4 Kb region marked by brackets contains the human/rat conserved sequence segments (60% identity over 90 bp) that were resequenced.

[0016] Fig. 2. The human *MCS5A1* (30.1 Kb) and *MCS5A2* (63.4 Kb) loci contain breast cancer-associated polymorphisms. The genomic regions shown on human chromosome *9* are orthologous to the rat *Mcs5a1* and *Mcs5a2* regions. *MCS5A* transcripts present in human tissue are mapped as exons connected by intronic sequence. Haplotype blocks that overlap *MCS5A* contain the breast cancer risk-associated polymorphisms, *SNP-A1* (rs6476643) in *MCS5A1* and *SNP-3* (rs2182317) in *MCS5A2*. Association of these SNPs to breast cancer risk was determined using two population-based samplings that evaluated ~12,000 women. The minor allele frequency of *SNP-A1* (*MCS5A1)* was 0.25. *SNP-A1* was associated with an ~19% increased risk to breast cancer for women that are homozygous for the minor allele. The minor allele frequency of *SNP-3* (*MCS5A2)* was 0.13. *SNP-3* associated with an ~14% decreased risk to breast cancer for the 24% of women that carried at least one minor allele. The polymorphisms that correlated (high r^2) with *SNP-A1* are in the 5.7 Kb region marked by brackets within the *MCS5A1* region. The SNPs correlated (high r^2)

with *SNP-3* that lie within the human/rat *MCS5A2* orthologous region are within a 16.8 Kb bracketed region. The x-axis is the human chromosome *9* base position based on the May 2004 UCSC human genome assembly. The x-axis marks designate SNPs, INDELs, CpG islands, and chromosome *9* base positions.

DESCRIPTION OF THE INVENTION

A. In General

[0017] Our lab has collected DNA from ~1,500 women in Wisconsin diagnosed with breast cancer and ~1,500 controls. In order to locate genes relevant to human breast cancer risk, our lab has used QTLs (Quantitative Trait Locus) observed in inbred rat strains as a guide for candidate regions in the human genome. We have chosen a sub-locus of the *Msc5*-QTL that had the highest LOD (Log of the Odds Ratio) score of all QTLs found in our original experiment on which this study is based (Lan, et al., Genetics 157:331-339, 2001).

[0018] *Mcs5a* is a sub-locus of the *Mcs5-QTL* that confers resistance to the development of DMBA-induced mammary carcinomas in rats. The *Mcs5a* sub-locus has been confirmed in congenic strains. Currently, sufficient recombinant strains have been analyzed to narrow the candidate region to ~116,000 bp. The *Mcs5a* locus when homozygous, in addition to conferring mammary carcinoma resistance to rats after DMBA treatment, also reduces mammary tumor multiplicity by ~50% as compared to littermates without Mcs5a after NMU treatment and over expression of *neu*. The rat genes in this region have been assayed for different levels of gene expression between the susceptible and resistant parent strains and all gene exons have been sequenced.

[0019] In order to further fine-map the *Mcs5a* locus, recombinant rat lines were collected and phenotyped for resistance to DMBA-induced mammary carcinogenesis. The results are presented in Fig. 1. Carcinoma development in DMBA-treated WKy-homozygous rats from congenic lines O, WW, and XX was reduced ~50% for each line (Fig. 1).

[0020] The current boundaries of the *Mcs5a* locus are given by the overlapping WKy sequences of congenic recombinant lines WW and XX, which define a genomic interval of ~116 Kb containing *Mcs5a*. Further analysis of additional congenic recombinant lines within this interval demonstrated that at least two genetic elements exist within *Mcs5a*. They are defined by incorporating phenotype data from congenic lines LL and B3. These lines have a very short interval of overlap (415 bp) within the *Mcs5a* locus, and together cover the entire putative *Mcs5a* locus.

[0021] Interestingly, WKy-homozygous rats from either line LL or B3 do not have a phenotype of resistance (Fig. 1). The simplest hypothesis to explain these data is that *Mcs5a* is a compound locus containing two genetic elements, termed *Mcs5a1* and *Mcs5a2*, which interact synthetically to confer the *Mcs5a* mammary cancer resistance phenotype. *Mcs5a1* is defined by the overlapping genomic region of congenic lines WW and B3, which is ~32 Kb (~30 Kb in humans). *Mcs5a2* is defined by the overlapping genomic regions of congenic lines LL and XX and covers ~84 Kb (~63 Kb in humans).

[0022] We localized the human region that is orthologous to the rat *Mcs5a* locus (*Mcs5a1* and *5a2*) to ~93 Kb on human chromosome *9* and determined the human haplotype block structure of a 1 Mb region surrounding this area. Haplotypes from a single block (Block 4) in the *MCS5A* locus were tested for association with breast cancer risk in a Wisconsin case-control study (1,500 cases; 1,405 controls). Haplotype

block 4.2, allele 4, that spanned SNPs rs10758440, rs999988, rs2182317, rs2381718, was found to be marginally significant in the Wisconsin population.

[0023] These four SNPs were also tested in the larger UK population (4,364 cases; 4,547 control). Table 2 demonstrates a highly significant association of the minor allele of rs2182317 (SNP-3) with a reduction in risk of developing breast cancer. The Wisconsin and UK populations combined (N = 11,779 women) yielded a heterozygous odds ratio of 0.86 (0.79-0.94), and a homozygous odds ratio of 0.77 (0.57-1.04). The uncorrected Chi-square trends test, stratified by study, yielded a P value of 0.0003 (Table 2), and when corrected for multiple comparisons using the conservative Bonferroni method the P value was 0.001.

[0024] We tested the additional genetic variation that was revealed after resequencing the human regions to determine if a polymorphism within *MCS5A1* that was not correlated to SNP-3 in the *MCS5A2* region associated independently with breast cancer risk in women. A total of 48 human chromosomes were resequenced at the *MCS5A1* locus. Polymorphisms were "binned" into groups of polymorphisms that were highly correlated to each other. Bins were screened for an association with breast cancer risk in the Wisconsin case-control population.

[0025] The results identified one bin (rs6476643, rs10758441, rs7042509, and an indel, 138-9899, on chr9:37563886-37563887) that qualified for further testing in the second population. One SNP from this bin (rs6476643, SNP-A1) was evaluated further in the UK breast cancer case-control population. The combined results for SNP-A1 yielded an odds ratio of 1.05 (0.97-1.13) for heterozygotes and 1.19 (1.03-1.38) for minor allele homozygotes (Table 2). A trends test, stratified by study, was significant for the minor allele (P = 0.022).

[0026] The SNP bin that includes rs6476643 (SNP-A1) contains 4 polymorphisms with similar minor allele distributions but slightly different minor allele frequencies. The SNP bin including rs6476643 is close to an area of recombination and spans ~6 Kb. No polymorphisms correlated to this bin were found outside of this 6 Kb *MCS5A1* region. This result is based on the distribution of the minor alleles found in the resequencing effort.

[0027] SNP-3 (rs2182317) and other polymorphisms listed in the bin are candidates for the causative SNP(s) in *MCS5A2*. These candidates include many highly correlated SNPs. Any one, or combination of SNPs found exclusively, or mostly, in individuals carrying the minor allele fo rs2182317 may be the causative SNP. We hypothesize that the causative SNP is in or near a conserved region; thus, the genetic variation found in this study may effectively narrow the region containing the causative SNP.

[0028] Considering all SNPs that are more rare or more common than rs2182317 and found in carriers of the minor allele of this SNP, the possible candidate SNPs include: 114-117(MAF=~0.02-0.04), d3-169, rs12378421, r3-116, rs17505776, rs4878708, rs4878709, rs4878710, rs10973450, l4-70, m4-218, rs4490927, x4-77, z4-66, f5-152, rs4878713, y5-43, i6-31, and i6-103. Testing these SNP-3 correlated SNPs individually will not, by itself, distinguish one of these SNPs as the causative SNP in *MCS5A2*. This is because all but one of these SNPs, 114-117, has an identical or very similar minor allele frequency and distribution to rs2182317.

[0029] Table 1, below, summarizes our risk assessment data.

Table 1.

MCS5A1 –Minor Alleles that Increase Risk of Breast Cancer

Polymorphism	Minor Allele Associated with risk on the (+) strand
rs6476643	T
138-9899	* (deletion)
rs10758441	T
rs7042509	G

Sequences not found in NCBI database:

138-9899
Indel 37563886^37563887
AAAATCTTAATCAAATGTCTCTTTGTGTACATTCCTTTATGTATACAGTATCAAACATTTTAGACGAGTGGGAAACTAATAATCACT
AAACTAAAAGA[AAGA]GGGTAATGAGAGAAATTAGCAGACATTTCAAACAAAACCCATGACAGGTAACTCAAGAATAGGTTCATT
CATTAAGGCAAACTGAAAGAACAAAACACTTGTTCCATCTATTTCCTGACTGAAGCAGGTACAAGGAATTTGTTTACTTCACATCT
TCCTGTGAAAAG

MCS5A2 - Minor Alleles that Decrease Risk of Breast Cancer

Polymorphism	Minor Allele Associated with risk on the (+) strand
rs2182317	T
114-117	T
d3-169	A
rs12378421	G
r3-116	T
rs17505776	C
rs4878708	A
rs4878709	A
rs4878710	A
rs10973450	C
l4-70	T
m4-218	G
rs4490927	T
x4-77	A
z4-66	G
f5-152	G
rs4878713	G
y5-43	C
i6-31	T
i6-103	G

Sequences not in the NCBI database:

114-117
Chr9: 37610247

AAAAACACAAAAACCAAAGGCTTCCAGATGAGGACAGAATTCCATTTTACCCTTCACTTCTACACAAATAGGCAAAATTAGAAGT
GGAACACTCGTGTTTATCTGCCTCCAGGTCAYAGCATCACAGCAGAGTAGAAGGAGCAGTGGCTTGGCATTGCTGAGAACCAA
AAATAATGAGCAGTTTCGCCACTGACAAGGCA

d3-169
Chr9: 37611591
AAGAGGATTTCTTTGGAGGAAGCAGCTGGTGTGCTAAGTGCCGCTCATGGCCCCAGGGGATAAGGAAGGGGTGTGTGGGTGC
CTATCTCACCTCACTTCAGCAGGACCACTCAGAGCTTGAGCTGTGTCTCCTGCAGTTGGGGGCCCAGGGGACTGGGTCTGACT
CCCRCCCTGGAAATTTACAGGAAAAGAGCAGGGCTGGCTAATGCTTTGGCTG

r3-116
Chr9: 37618560
CAGAGAAAAAGAAAGGAGGAGGAGGAGGAAAGAGAAGTCAAGTTTTAAAAGCAAACAGAAATAATAGTTTTGATGGGATGGATT
TGTTTTCTTAATAAGCCTAATGATCCCAGGAKTCTCATTGGATTATTAGCAACTGTATTTAAAATTTAAATCTAAAACAAGTTTTGT
AGCTGCAAAGTGCTTATCAGAACCTAACCATGCCTCTCTGGCT

l4-70
Chr9: 37634719
AGAGCTCAGTTTCCCAAATAAACCTGAATCTGAATCCCATTTACCAGCTCTGTGACCTCATACCAGTCGYTTGAATTCTCTGAGC
TTGCCTCAGTTTCATCTGTGAAGTGGGGATTGTCATGTGTGTCCTGCCTAACTCAAGCAGCTGTTGAGAGAATAAGATGAGATG
ATTGCTCTGTTTGGTGTTTTACAAACTGAAA

m4-218
Chr9: 37635098
TAGTGCTAGAGAGGGTGGCATCTGCCCCCGGGGGCTGTGGGGAGCACGGTGGTAGTGGGGAGTTGCAGTCACCCAGCAGAT
GCAAAGCAGAATGAAGCCTATGGGAAGTGGCCAGATGTAGTTGTCAAGACCAGGGTAGGGCCACACTTGATGCTGGCCACAAC
CTCAAGACATGCCTTCTTCTGGGGAGCTGATCTCCCAGGAGCTGAGCCAGCAGRAAAAAAAAAAAAAGAGAG

x4-77
Chr9: 37639129
AGGCTGACACTGACGCAGGTAGCTAGGAAATGGAGGGGGACAGGGCAAAATTCCATGGGGAACATTTGTACTCACCRGACAA
GAGAATAGTTCTTATCTTACTTTCTGACAAGTGCTATGCTTTTTGGTGTTTAACTGCCTTGGAGAGAGTAATTTGATGATTAATCT
TTATCTACAAAATAATTTTTAAAATTAAACTTTATTTTGAGAAAATTGTAGATTCA

z4-66
Chr9: 37639690
AAGTGCTGCTAGTCTTCCCCCACCGCCCTTGGATGAAGGATGCACTTGTTAACATCTGCCTGTGCRCCATCCCCAGAATGATCT
AACATCCCAGTCTGATGGTGCCTCAACCCTACTCAGAACTCCTCCCATGAACCTGCCAGTGCACTGGGAACACGGTGCAAACT
CCTTGACCAGGAATTGGCCCCTGCTGACTCCCC

f5-152
Chr9: 37643111
AGGTGGTGAGGAATAGAAAGGCTATGTGTAGGTCAAGGAGCTGGTACGAGCAAGGAACTGGTGAGAGCTGGGAGAGCATGAG
ACAGATGAGAGCAGTGAGCCTGGTGGAAGCACAGAGAAGGCAGAGCACTTCAGAGACAGGGCAGCGGGARGAGGCTCTGGA
GGCTCTGAACCTGGGCTGGAATCCTCGCTGGCTGTGG

y5-43
Chr9: 37650583
ATTTCTCTTAAAGAACAGTGATTTTAAAGTAGGTTTAAACAAYGGGTTTAAAGAACAGTGATTTTCCATTTTGACAAGGCTTGTTT
GGTATAGCCACTTCAAAATATCCCCACATCAAAATATGATGCTTTATCTGGGATATCTTTTCCAGACTTTGGGAGCCTTCCCTCCT
GCACTTTCATTTCTGCAAGGCAATACTC

i6-31
Chr9: 37655211
CAACATATTTGTAATACAGACATCGAAAATAAACCAAATATCCAGCAATAGGGACTTAGCTTTAAAAATGGCACTCTTATCTATAG
CAAATAAACAATAATGTTGTAGAATAATAAAGACACAGGGAAATGTTTACAGTGTATATTTTTCAGTTTAGAAACCAGCATATATG
GTAAGTTCCCAATTATGTTGAAAATGTGTCTTTTCACACWAAAAGACTGGAAGAGTAATTAGCAACTTATAGCTTTTAGTGGACA
ATTTTTATTATCTTCTCTCCAAAAAAGAAAACTTTCAAAATTTTCTACCAGCCACATGTATTACTTTTATAAGAGGGAGGAAATAAG
TGACAATTAAGAAAAAGAGATAAGCTTTTGGAGAGCGTTGTTTGAGCGTCACTAGG

i6-103
Chr9: 37655573

TGGAAAACTAAATGAGCACATCATTTATCTTGCAAATTCCAGACAGGTGGTGTTTTCAGAAGGAAGAGTGGTCTTAGGTCCATTT
GTGTGAGTATATTTATAGAAGTGAAAGCTTTGGGGGAAGGAAATAGATTGATTTTTTTTCCCCTTGAACTTCTGAAATTATTTTTT
CCRCTCCATTTGTAATTGAGCCCAGGGAGCTATTCTTATTTCTTCCTTTCTTGGGCACTGCGTTAGACCTAAAAATGTTAACTGG
CTTAGGATGTGGGTTTTGCTAAAATGATTCCCCTTGAAGTCTTCACTGGGCTTTCTCATGCTTAAAAGTGGGGTCCCGCAGAAG
ATCACTTTCTACCTAATGCACTTTGCTCCTGAAGTCTTTGGCAACGTTGGGGTGGTCAGA

All chromosome positions are from the current NCBI build. These numbers will change with successive builds. However, the flanking sequence will allow others to identify these SNPs in the future.

B. Diagnostic Method

[0030] In one embodiment of the present invention, one would collect a DNA sample from a patient and genotype that DNA sample in order to estimate risk. For example, a patient's sample with at least one T allele of rs2182317 would indicate that the patient is at a lower risk for breast cancer than the general population.

[0031] We describe a preferable prophetic diagnostic method below. Of course, one of skill would be able to substitute similar method steps. The method is based on the same two SNPs, rs6476643 and rs2182317, that we used for our experiments described below.

Collect DNA from individual:

[0032] Individual rubs cheeks against teeth with ~15 ml of SCOPE mouthwash in mouth for 30 seconds and rinses into 40 ml Nalgene container, repeat.

Extract DNA:

[0033] Collect mouthwash from individuals, transfer to a 50 ml Falcon centrifuge tube, and centrifuge in a JS-4.2 rotor at 3000 RPM (2000xg) at room temperature for 15 minutes. Pour off SCOPE and then extract DNA from pellet according to PUREGENE cell and tissue kit from GENTRA systems according to manufacturer's protocol. Sample should be stored at -20°C after collection, cell lysis, or when sample is in isopropanol.

Genotyping:

[0034] Use 5 ng of individual's genomic DNA in ABI SNP genotyping assay according to manufacturer's protocol. Plate must be read on ABI 7700 or 7900 instruments with SDS software. One of skill in the art would understand that there are numerous methods to characterize SNPs. Any accurate method of genotyping would be suitable for the present invention.

[0035] One would examine the DNA for the presence or absence of the minor alleles listed in Table 1 or minor alleles correlated with these alleles. The rs numbers listed in Table 1 and Table 2 and throughout the application are listed in a public database (NCBI, National Center for Biotechnology Information) and refer to a specific location in the genome where two different bases (A,T,C, or G) are found in a particular position in the human population. When looking at the sequences (as opposed to the rs numbers) we supply with Table 1, the IUPAC letters indicate a polymorphism, i.e. R = A or G, Y = C or T, M = C or A, K = T or G, W = T or A, S = G or C. ~99.9% of all bases are identical in all people. While any individual can have a mutation that is extremely rare (for example one person has the base change in millions of people) at any given position in the human genome, there are some specific sites that vary in many people. A base variation is considered a SNP if the rare base is found at a frequency of 1 in 100 chromosomes. The less common base variation is called the "minor allele."

[0036] The genotype of the sample determines the risk prediction. Women who are heterozygous (G/T) or homozygous (T/T) for the minor allele at SNP rs2182317 have a 14% reduction in risk. Women who are homozygous for the minor allele of rs6476643 (T/T) have a 19% increase in risk. The polymorphisms listed in Table 2 can

also be used to predict risk. The alleles associated with risk are listed in the Table as well.

Risk Assessment:

[0037] Our data indicate that the minor allele of rs6476643 in the *MCS5A1* region acts in a mostly recessive manner to increase risk by ~19% in ~6% of women. Within the *MCS5A2* region the minor allele of rs2182317 acts in a dominant manner to decrease risk by 14% in ~24% of women.

Use of other SNPs

[0038] We have characterized susceptible and resistant individuals using SNP rs2182317 to estimate resistance and rs6476643 to estimate susceptibility. However, one could genotype other SNPs which are correlated to either of these SNPs (most preferably $r^2 > 0.9$, preferably $r^2 > 0.7$) to characterize the experimental sample. In one method, one would sequence regions and discover previously unidentified polymorphisms and then genotype a set of people for these new polymorphisms including our two and then haplotype all individuals. Individuals can be haplotyped for all genotyped SNPs including rs2182317 and rs6476643 by use of a statistical program such as PHASE (Stephens, M., and Donnelly, P., Am. J. Hum. Gene. 73:1162-1169, 2003; Stephens, M., et al., Am. J. Hum. Gene. 68:978-989, 2001). Other polymorphisms can be tested for correlation with rs2182317 or rs6476643 with the r^2 statistic (Weir, B.S., 1996, Genetic Data Analysis, Sinaur).

[0039] Polymorphisms with a high correlation to our two SNPs ($r^2 > 0.7$) should replicate the same association pattern. Also, any polymorphisms on Chromosome 9 between the SNPs rs10973432 and rs12554736 with allele frequencies smaller than

rs2182317 that are exclusively or mostly present in individuals who carry the minor allele of rs2182317 may replicate the pattern if they are in fact the causative polymorphism.

C. Therapeutic Target

[0040] The *FBXO10* and *FMRPD1* gene loci products (i.e. the entire genomic sequence including exons, introns, promoters, and 3'-end) would be useful as a target for chemoprevention or therapeutic drugs. Other *FBXO10* and *FMRPD1* gene loci instantiations may be non-coding RNAs (ncRNA), *cis*-regulatory elements, enhancers, silencers, splice modifiers, etc. Another embodiment of the present invention is a therapeutic intervention mimicking the biological effect of *FBXO10* and *FMRPD1* gene loci alleles that decrease/increase susceptibility to breast cancer, preferably using small molecules. Experimental evidence suggests that these drugs would be efficacious for all members of the population since there is a dosage effect seen with the *Mcs5a1* and *Mcs5a2* resistance alleles in animal studies. Therefore, the drug target would be *FBXO10* and *FMRPD1* gene loci products and other biological molecules that work in the same mechanistic pathway.

[0041] In one embodiment, this method would comprise a method for assaying for modulators of the *FBXO10* and *FMRPD1* gene loci products, consisting of the steps of exposing the *FBXO10* and *FMRPD1* gene loci products to test compounds and determining whether the test compound mimics the biological effect of *FBXO10* or *FMRPD1* alleles that decrease susceptibility to breast cancer. In one embodiment, this would be done by determining gene loci that are differentially expressed between primary mammary cell cultures derived from WKy and WF parental strains using gene expression array analysis. A subset of differentially expressed gene loci would become

indicators for drug screening assays. The promoters from the selected reporters would be placed in a reporter gene construct using standard molecular biology procedures and transfected into cultured mammary cells. The drug screening assays would involve exposing transfected mammary cells to a panel of small molecules and monitoring expression of the reporter gene contained within the construct. Those drugs that yield a similar pattern of reporter expression similar to the WKy strain would become lead candidates for further development and *in vivo* testing.

[0042] The following are proposed therapeutic targets:

- Target 1 - **F-Box Motif of** *FBXO10* : disrupt or enhance the *FBXO10* interaction with the ubiquitin-conjugating enzyme. This could be done with small molecules to inhibit the protein-protein interaction between *FBXO10* and the ubiquitin-conjugating enzyme. Antisense-RNA targeting *FBXO10* and/or the conjugating enzyme could be used to disrupt the interaction.
- Target 2 – **Target Specificity Domain of** *FBXO10* disrupt or enhance the *FBXO10* interaction with its specific biological target(s). F-Box proteins have evolved to function as cell regulatory ligases. The regulatory function is achieved by having a target specificity domain in addition to the conserved F-Box. The target specificity domain serves as a binding domain for specific biological molecules that the F-box protein regulates. Therefore, the interaction between *FBXO10* and the gene it regulates would be a target for drugs and antisense-RNA.

D. *FBXO10* and *FMRPD1* loci as a probe to find biological molecules acting in the same mechanistic pathway

[0043] In another embodiment of present invention, one would use the *FBXO10* and *FMRPD1* gene loci as probes to identify other biologically important molecules. One would seek to identify molecules that interact with and act in the same biological pathway as the *FBXO10* and *FMRPD1* gene loci using the *FBXO10* and *FMRPD1* gene loci as probes. In one embodiment of the present invention, one would assay molecules for their ability to interact with the *FBXO10* and *FMRPD1* gene loci or the *FBXO10* and *FMRPD1* gene loci products.

[0044] *FBXO10* is an F-Box protein, which indicates it functions biologically as a ligase to ubiquitinate, thus regulates specific targets within the cell. Therefore, there are at least two other gene products with which *FBXO10* has direct interaction. The function of the F-Box motif is to provide a link between the ligase and the ubiquitin-conjugating enzyme. F-Box proteins have evolved to function as cell regulatory ligases. The regulatory function is achieved by having a target specificity domain in addition to the conserved F-Box. The target specificity domain serves as a binding domain for specific biological molecules that the F-box protein regulates. Based on this evidence and because of the specificity *FBXO10* gene locus will provide, the locus is the best biochemical probe to identify the upstream and downstream interacting genes.

Proposed Methods

[0045] *Yeast Two Hybrid Screen for Mcs5a interacting proteins* - Yeast two hybrid screens in which different segments of the *FBXO10* and *FMRPD1* gene loci products are used as bait would be useful to screen for specific proteins that interact with various domains of *FBXO10* and *FMRPD1* loci products. For example, using the sequence that encodes the F-Box as bait would be expected to pull out the ubiquitin-conjugating enzyme. In a separate screen, the target specificity domain of *FBXO10* used as bait would be expected to pull out the regulatory target of the *FBXO10* locus.

[0046] *Pull-Down Assays* - In this assay the *FBXO10* and *FMRPD1* loci open reading frame sequence is placed in an expression vector that adds a detectable tag (i.e. GFP, FLAG, GST) forming a recombinant protein that can be captured on using an affinity ligand specific for the tag. This method allows the study of protein-protein interactions from their endogenous environment. It can be used to confirm interactions discovered in the yeast two hybrid screens and also as a discovery tool. As a discovery

tool the pull-down assay can be useful to identify directly and indirectly interacting proteins because the entire protein complex may be co-purified. For verification the interacting protein can be tagged and immunoblotted with an antibody specific to the tag after coimmunoprecipitation of the protein complex. Or, if an antibody specific to the interacting protein is available, this can be used to confirm the presence of the interacting protein. For discovery, mass spectrometry could be employed to identify the proteins present in the complex.

[0047] *Mammalian Two Hybrid Screen* – This screen is similar to the yeast two hybrid system except that mammalian cells are used. The advantage is the false-positive rate is reduced.

[0048] *Correlated Gene Expression* – One method to find gene loci that act in the same molecular pathway would be to identify mRNAs that exhibit a positive or negative correlation with the expression patterns of the gene locus of interest. To do this, conditions that alter the mRNA levels of *FBXO10* and *FMRPD1* would be identified. Microarray analysis of gene expression would then be performed under the same condition to detect other gene loci that change in expression under the same condition. The gene loci that change in expression would then be considered as candidates acting in the same molecular pathway.

[0049] If the important genetic elements in the *Mcs5a1* gene locus is an ncRNA, enhancer, splicing modifier, silencer, etc. then the approach would be to identify specific biological molecules being modulated by this element. This could be accomplished by comparing the gene expression profile between mammary glands of congenic rats with or without the modifier element.

EXAMPLES

Rat *Mcs5a* is a Compound QTL with Orthologous Human Loci that Associated with Breast Cancer Risk

Abstract

[0050] To identify human breast cancer risk genes, a rat mammary carcinogenesis model was used to positionally identify and characterize susceptibility loci. We identified a compound QTL, *Mcs5a*, which required a synthetic interaction of two elements, *Mcs5a1* located within *Fbox10*, and *Mcs5a2* within *Frmpd1*, to confer mammary cancer resistance. In two large case-control populations including ~12,000 women, the minor alleles of rs6476643 (SNP-A1) in *MCS5A1* (p=0.02) and rs2182317 (SNP-3) in *MCS5A2* (p=0.0003) associated with an altered risk of breast cancer. These results illustrate the complexity of breast cancer genetics while demonstrating the utility of rat models to identify unbiased candidates that associate with breast cancer risk.

Introduction

[0051] In order to further fine-map the *Mcs5a* locus, recombinant rat lines were collected and phenotyped for resistance to DMBA-induced carcinogenesis. Carcinoma development in DMBA-treated WKy-homozygous rats from congenic lines O, WW, and XX was reduced ~50% for each line (Fig. 1). The boundaries of the *Mcs5a* locus are given by the overlapping WKy sequences of congenic recombinant lines WW and XX, which define a genomic interval of ~116 Kb containing *Mcs5a*. Analysis of additional congenic recombinant lines within this interval demonstrated that at least two genetic elements exist within *Mcs5a*. They are defined by incorporating phenotype data from congenic lines LL and B3. These lines have a short interval of overlap (415 bp), and together cover the *Mcs5a* locus. Interestingly, WKy-homozygous rats from either line

LL or B3 do not have a phenotype of resistance (Fig. 1). The simplest hypothesis to explain these data is that *Mcs5a* is a compound locus containing two genetic elements, *Mcs5a1* and *Mcs5a2*, which interact synthetically to confer the *Mcs5a* mammary cancer resistance phenotype.

[0052] *Mcs5a1* is defined by the overlapping genomic region of congenic lines WW and B3, which is ~32 Kb (~30 Kb in humans). *Mcs5a1* contains mostly intronic sequence of the uncharacterized *Fbxo10* gene, which contains an F-Box domain. F-box proteins are components of SCF complexes (ubiquitin ligases). The *Mcs5a1* region contains a small section of intron 1, the small exon 2, and the majority of intron 2 of the *Fbxo10* gene. Resequencing the WF and WKy alleles of *Fbxo10* exon 2 revealed no nucleotide differences. However, resequencing of the entire *Mcs5a1* region, using gDNA from WF and WKy rats, identified 90 SNPs and 28 indels within the *Fbxo10* intronic sequence of the *Mcs5a1* locus. This observation supports a search for potential *Mcs5a1* candidates within the intronic sequence of the *Fbxo10* gene. Several known mammary expressed transcripts map to the *Mcs5a1* region within the *Fbxo10* intronic sequence (Fig. 1).

[0053] When assayed by quantitative RT-PCR with two different *Fbxo10* TaqMan™ probes, no significant differences in mRNA levels were found in mammary gland tissue from 12-week-old virgin rats homozygous for the *Mcs5a* WKy allele versus those homozygous for the *Mcs5a* WF allele. Similarly, no significant differences in *Fbxo10* mRNA levels were found in mammary glands of DMBA-exposed rats.

[0054] *Mcs5a2* is defined by the overlapping genomic regions of congenic lines LL and XX and covers ~84 Kb (~63 Kb in humans). This locus spans a region from within the first intron of *Fbxo10*, including its first exon and proximal promoter, to a

location between the 5' UTR and the first coding exon of the uncharacterized gene, *Frmpd1* (FERM and PDZ domain containing 1). No difference in the amino acid sequence encoded by the first exon of *Fbxo10* was identified between WKy and WF rats. Analysis of *Frmpd1* expression in mammary gland tissue identified no differences in mRNA levels in rats at 12 weeks or after DMBA exposure. In addition, the nucleotide sequence of the transcribed *Frmpd1* 5' UTR is identical between the WF and WKy rat strains.

[0055] We localized the human region that is orthologous to the rat *Mcs5a* locus (*Mcs5a1* and *5a2*) to ~93 Kb on human chromosome *9* and determined the human haplotype block structure of a 1 Mb region surrounding this area. Haplotypes from a single block in the *MCS5A* locus were tested for association with breast cancer risk in a Wisconsin case-control study (1,500 cases; 1,405 controls). Allele 4, a subregion of this block that spanned SNPs rs10758440, rs999988, rs2182317, rs2381718, was found to be marginally significant in the Wisconsin population. These four SNPs were then evaluated in the larger UK population (4,364 cases; 4,547 controls). Table 2 summarizes the results, which demonstrate a highly significant association of the minor allele of rs2182317 (SNP-3) with a reduction in risk of developing breast cancer. The Wisconsin and UK populations combined (N = 11,779 women) yielded a heterozygous odds ratio of 0.86 (0.79-0.94), and a homozygous odds ratio of 0.77 (0.57-1.04). The uncorrected Chi-square trends test, stratified by study, yielded a P value of 0.0003 (Table 1), and when corrected for multiple comparisons using the conservative Bonferroni method the P value was 0.001.

[0056] Polymorphisms highly correlated to SNP-3 could include the causative polymorphism(s) for the association to breast cancer risk. To locate correlated

polymorphisms, the human-rat conserved non-coding sequence (CNS) segments within *MCS5A2* (~27 Kb total sequence, ~30% CNS) and the entire *MCS5A1* region (~30 Kb) were resequenced in 24 women from the Wisconsin case-control population (12 cases; 12 controls). These samples were chosen to represent the haplotype allele frequencies observed in the population.

[0057] After resequencing, haplotype block 4 was separated into haplotype blocks A, B, and C (Fig. 2) based on r^2 estimates. We found no polymorphisms in the resequencing data of the *MCS5A1* region that correlated to SNP-3. We did find 12 SNPs, highly correlated to SNP-3, in the *MCS5A2* region that localized to a ~17 Kb interval, which is ~40 Kb distal to the *MCS5A1/5A2* boundary (Fig. 2). This ~17 Kb sequence covers a region upstream of *FRMPD1* to an interval that includes its 5' UTR exon (Fig. 2). It is likely that the *MCS5A2* causative genetic alteration resides either within, or in close proximity to this reduced interval.

[0058] (SNP-3 (rs2182317) and other polymorphisms listed in the bin are candidates for the causative SNP(s) in *MCS5A2*. These candidates include many highly correlated SNPs that are heterozygous in all individuals carrying, exclusively, allele 4 of haplotype block 4.2. Any SNP-3 correlated SNP, or SNP combination, found in individuals carrying haplotype block 4.2 allele 4 could be the causative SNP. We hypothesize that the causative SNP is in or near a conserved region; thus, the genetic variation in our database may effectively narrow the region containing the causative SNP to chr9:37,610,247-37,655,573.

[0059] Considering all SNPs that are more rare or more common than SNP-3 and found in carriers of haplotype block 4.2 allele 4, the possible candidate SNPs include: 114-117(MAF=~0.02-0.04), d3-169, rs12378421, r3-116, rs17505776, rs4878708, rs4878709, rs4878710, rs10973450, l4-70, m4-218, rs4490927, x4-77, z4-66, f5-152, rs4878713, y5-43, i6-31, and i6-103. Testing these SNP-3 correlated SNPs individually will not, by itself, distinguish

one of these SNPs as the causative SNP in *MCS5A2*. This is because all but one of these SNPs, 114-117, has an identical or very similar minor allele frequency and distribution to rs2182317. There were only two SNPs, 128-249 and a correlated SNP 105-3, in *MCS5A1* in which the minor alleles were found only in individuals with haplotype block 4.2 allele 4. The SNP 128-249 (MAF 2%) was not associated with a reduction of breast cancer risk; thus, the only causative SNP-3 correlated SNP candidates reside in *MCS5A2*.).

[0060] Like the rat model, no SNPs were identified within the *FRMPD1* 5' UTR in the human samples. Interestingly, resequencing of the distal ~36 Kb conserved regions of *Mcs5a2* in the rat strains revealed that polymorphisms between the WKy and WF sequence occur only in the region orthologous to the SNP-3 correlated SNP region in the human (Figs. 1 and 2).

[0061] We tested the additional genetic variation, revealed after resequencing the human regions, to determine if a polymorphism within *MCS5A1* that was not correlated to SNP-3 in the *MCS5A2* region associated independently with breast cancer risk in women. A total of 48 human chromosomes were resequenced at the *MCS5A1* locus. Seventy-two (19 not listed in dbSNP) *MCS5A1* SNPs with a minor allele observed in more than a single individual were documented. These SNPs were "binned" into groups of polymorphisms that were highly correlated to each other.

[0062] Bins were screened for an association with breast cancer risk in the Wisconsin case-control population. The results (Table 2) identified one bin (rs6476643, rs10758441, rs7042509, and an indel, 138-9899, on chr9:37563886-37563887) that qualified for further testing in the second population. One SNP from this bin (rs6476643, SNP-A1) was evaluated further in the UK breast cancer case-control population. The combined results for SNP-A1 yielded an odds ratio of 1.05 (0.97-1.13)

for heterozygotes and 1.19 (1.03-1.38) for minor allele homozygotes (Table 2). A trends test, stratified by study, was significant for the minor allele (P = 0.022). The polymorphisms in this bin are not highly correlated to any other variation; therefore, one of these polymorphisms is likely causative (The SNP bin that includes rs6476643 (SNP-A1) contains 4 polymorphisms with similar minor allele distributions but slightly different minor allele frequencies.

[0063] The minor allele of rs6476643 was found 84% of the time in women with haplotype block 4.2 allele 3, and 27% of the time in women with haplotype block 4.2 allele 1. The SNP bin including rs6476643 is close to an area of recombination and spans ~6 Kb. No SNPs correlated to this bin were found outside of this 6 Kb *MCS5A1* region. This result is based on the distribution of the minor alleles found in the resequencing effort. Only one other SNP bin had minor alleles found mostly in haplotype block 4.2, allele 3. This SNP bin marked by SNP 24-131 has a minor allele frequency of 3%, and was found in 16% of women who have haplotype block 4.2, allele 3. However, this SNP did not show an association with breast cancer risk. The indel 138-9899 is correlated to rs6476643 (r^2=0.9). This correlation was not observed in the resequencing data because heterozygous and homozygous individuals for the deletion could not be distinguished (*i.e.*, both appear as homozygous for the deletion). When tested using allelic discrimination, this polymorphism showed a trend towards resistance in the Wisconsin population with a P-value of 0.22).

[0064] The human sequence variation correlated to SNP-A1 (rs6476643) in *MCS5A1* is located at the beginning of haplotype Block B, and the variation correlated to SNP-3 (rs2182317) in *MCS5A2* is located in haplotype Block C (Fig. 2). The SNP-A1 cluster spans ~6 Kb of human chromosome *9* and contains a CpG island located at a

transcriptional start for a potential candidate gene identified in the rat and human (*AF176705*) (Figs. 1 and 2). *AF176705* encodes FBXO10, and two splice variants originate from the homologous CpG island in the rat. The SNP-3 cluster in *MCS5A2* contains the putative regulatory region of *FRMPD1*, which is expressed in both rats and humans.

[0065] This study demonstrates the utility of a comparative genomics approach to identify DNA polymorphisms that can serve as risk markers of multigenic disease and to reveal the genetic complexity underlying common diseases. Complexity is illustrated in the rat compound *Mcs5a* locus that consists of at least two interacting elements *Mcs5a1* and *Mcs5a2*. An analysis of the human homologous regions identified alleles from both *MCS5A1* and *MCS5A2* that associated with breast cancer risk. The minor allele of SNP-A1 in the *MCS5A1* region acts in a mostly recessive manner, similar to the rat WF allele, to increase risk by ~19% in ~6% of women. Within the *MCS5A2* region the minor allele of SNP-3 acts in a dominant manner, similar to the rat WKy allele, to decrease risk by ~14% in ~24% of women.

[0066] The *Mcs5a* elements responsible for altering cancer risk may not reside in protein coding sequence in either the rat or human. While the percentage of non-protein coding elements contributing to the risk of complex diseases is unknown, other such elements have been associated with disease risk (E. S. Emison *et al., Nature* 434, 857 (2005), G. G. Loots *et al., Genome Res.* 15, 928 (2005)). Thus, the identification of these two breast cancer modifier alleles by taking advantage of natural genetic variation in model organisms, reinforces the possibility that a sufficient number of modifier alleles can be identified using similar methods to impact breast cancer risk estimation in women.

Table 2. Breast Cancer Risk Associated SNPs in *MSC5A1* and *MCS5A2*

Genotype frequencies and genotype-specific risks in 5848 women with breast cancer and 5931 controls from Wisconsin and the UK.

MCS5A1 SNP-A1 rs6476643	P -values Trend (1df) Heterogeneity (2df)	Series	Major allele frequency	Minor allele frequency	Major Allele homozygote No. (%)	Heterozygote No. (%)	Minor Allele homozygote No. (%)	Number genotyped
WISC	0.058	Cases	0.72	0.28	778 (53)	590 (40)	113 (8)	1481
	0.143	Controls	0.75	0.25	765 (55)	532 (39)	84 (6)	1381
		OR (95%CI)	1 (ref)	1.12 (1.03-1.22)	1 (ref)	1.09 (0.93–1.27)	1.32 (0.98–1.78)	
UK	0.123	Cases	0. 74	0.26	2373 (55)	1657 (38)	316 (7)	4346
	0.230	Controls	0.75	0.25	2517(56)	1704 (38)	290 (6)	4511
		OR (95%CI)	1 (ref)	1.05 (1.01-1.11)	1 (ref)	1.03 (0.94–1.13)	1.16 (0.98–1.37)	
Combined	0.022	Cases	0. 73	0.27	3151 (54)	2247(39)	429 (7)	5827
	0.049	Controls	0. 75	0.25	3282 (56)	2236 (38)	374 (6)	5892
		OR (95%CI)	1 (ref)	1.07 (1.01-1.13)	1 (ref)	1.05 (0.97–1.13)	1.19 (1.03–1.38)	

MCS5A2 SNP-3 rs2182317	P -values Trend (1df) Heterogeneity (2df)	Series	Major allele frequency	Minor allele frequency	Major Allele homozygote No. (%)	Heterozygote No. (%)	Minor Allele homozygote No. (%)	Number genotyped
WISC	0.140	Cases	0.89	0.11	1166 (79)	291 (20)	15 (1)	1472
	0.126	Controls	0.88	0.12	1074 (78)	284 (21)	26 (2)	1384
		OR (95%CI)	1 (ref)	0.89 (0.75-1.04)	1 (ref)	0.94 (0.78–1.13)	0.53 (0.28–1.01)	
UK	0.001	Cases	0.89	0.11	3436 (79)	878 (20)	62 (1)	4376
	0.003	Controls	0.87	0.13	3430 (75)	1045 (23)	72 (2)	4547
		OR (95%CI)	1 (ref)	0.86 (0.81-0.92)	1 (ref)	0.84 (0.76–0.93)	0.86 (0.61–1.21)	
Combined	0.0003	Cases	0.89	0.11	4602 (79)	1169 (20)	77 (1)	5848
	0.0016	Controls	0. 87	0.13	4504 (76)	1329 (22)	98 (2)	5931
		OR (95%CI)	1 (ref)	0.86 (0.80-0.94)	1 (ref)	0.86 (0.79–0.94)	0.77 (0.57–1.04)	

Materials and Methods

Animal Experiments

[0067] Inbred WF and WKy rats were obtained from Harlan Sprague-Dawley Inc. (Indianapolis, IN) and fed Teklad lab blox chow and acidified water *ad libitum*. Rats were maintained in a 12-hour light/dark cycle in an AALACC-approved facility and all protocols were approved through the University of Wisconsin Medical School Animal Care and Use Committee. Congenic lines were established and maintained as previously published (D. J. Samuelson *et al.*, *Carcinogenesis* 24, 1455 (2003)). Congenics are defined as genetic lines developed on a WF genome and carrying the selected WKy alleles shown in Fig. 1. The congenic generations (number of backcrosses) used to

determine the mammary carcinoma multiplicity phenotypes of lines O, LL, WW, XX, and B3 were N12, N14, N14, N15, and N15, respectively.

[0068] Primer information of the genetic markers and the SNP bp (rat genome June 2003 assembly available at the UCSC Genome Browser, www.genome.ucsc.edu) defining the ends of the WKy allele carried by each line are: lines O, XX, and B3 proximal microsatellite marker *gUwm40-18* FWP: 5'-GACTTAATGTGGGGAGTGAA, RVP: 5'-AGCACATATGGAGGTTTGAC; lines O, LL, and WW distal microsatellite marker *gUwm45-5* FWP: 5'-CTAGAAAGGTGCTTTGGTTG, RVP: 5'-TCAGCTTCTCCTCCTTCC; line WW proximal SNP-A at chr*5*:61634906$^{C>T}$; line LL proximal SNP-B at chr*5*:61667232$^{A>G}$; line XX distal microsatellite marker *gUwm23-29* FWP: 5'-CCAGTCTGATGACCTGAGTT, RVP: 5'-CTTGCATGTGTGTAAGTGCT; and line B3 distal SNP-C at chr*5*:61667646$^{G>A}$.

[0069] To determine mammary carcinoma susceptibility phenotypes, WF.WKy female rats aged 50-55 days that were WKy-homozygous or WF-homozygous for the selected congenic intervals were administered a single dose of 7,12-dimethylbenz[a]anthracene (DMBA, 65 mg/kg body weight) (ACROS Organics; Fisher Scientific, Pittsburgh, PA) in sesame oil by gastric intubation, or were injected IP with *N*-nitroso-*N*-methylurea (NMU, 50 mg/kg body weight) (Ash Stevens, Detroit, MI). For Her2/*neu* experiments, female rats were infused with replication defective retrovirus containing the activated *neu* oncogene (pJR*neu*). The retrovirus was titered *in vitro* as previously described and 2.0 x 10^5 CFU/ml were injected into the central duct of each mammary gland. For detailed information on the pJR*neu* vector see (B. Wang, W. S. Kennan, M. N. Gould, *Cancer Res.* 51, 5649 (1991)). Mammary carcinomas ≥ 3x3 mm were counted at 15, 17, and 8 wks post-induction for DMBA, NMU, and Her2/*neu*

experiments, respectively. Mammary carcinoma multiplicity data were analyzed using the Mann-Whitney nonparametric test of StatView (SAS Institute Inc., Cary, NC).

Comparative Genomics

[0070] The *Mcs5a* rat chromosome *5* genomic region is based on the bp in the rat June 2003 genome assembly (UCSC Genome Browser) for the SNP at chr*5*:61634886$^{C>T}$ and the microsatellite marker (*gUwm23*-29) that mark the respective ends of the WKy *Mcs5a* allele delimited by the congenics. The *Mcs5a* proximal and distal rat chromosome *5* bp (chr*5*:61634886-61751793) and the VISTA browser (//pipeline.lbl.gov/) were used to find the human *Mcs5a* orthologous region on human chromosome *9* (chr*9*:37550841-37644624). Genes/genetic elements mapping to the region were found using the UCSC and VISTA Rat and Human Genome browsers.

[0071] Transcribed elements were verified using RT-PCR of total RNA from WF and WKy mammary gland and brain tissues. cDNA was synthesized from 2 µg total RNA for 2 hours at 42°C in a 20 µl final reaction volume consisting of 0.5x RNA Secure (Ambion, Austin, TX), 0.05 µg/µl oligo(dT)^{18}, 125 µm dNTP mix, 1x first strand buffer (Invitrogen, Carlsbad, CA), 10 mM DTT, and 200 U Superscript II reverse transcriptase (RT) (Invitrogen). For PCR reactions, RT reactions were diluted 1:2 or 1:4 and a microliter of this dilution (~25-50 ng RNA equivalent cDNA) was used in a 20 µl PCR reaction. The reaction components were 1x Herculase Buffer (Stratagene, LaJolla, CA), 200 µM each dNTP mix, 500 nM each primer, and 1 U Herculase DNA Pol (Stratagene). PCR cycling conditions in a Biometra T3 thermocycler were 95°C for 2 min, followed by 35 cycles of 92°C for 1 min, 59°C for 45 sec, and 72°C for 2 min. A 5 min extension at 72°C was added at the end. Human/Rat conserved non-coding sequences (CNS) in the

Mcs5a2 region were identified using the VISTA browser with settings of 90 bp and 60% identity.

TaqMan Quantitative RT-PCR

[0072] TaqMan Quantitative PCR (QPCR) primers and probes (ABI, Foster City, CA) were designed using Primer Express v2.0 (ABI) according to manufacturer's specifications. Primer and probe sequences, respectively, were: *Fbxo10* probe1: 5'-FAM-AGGCCTCATCACAGAAAACGTCATCCG-TAMARA, FWP: 5'-GCTGGCATAGCAGTGAACGA, RVP: 5'-CTACACCTCCCCACTGGTTCTC; probe 2: 5'-FAM-CTTTTCCCAGACCACATC-MGB, FWP: 5'-CCAGCCTCTATGATCGAATCG, RVP: 5'-CATTCTCAAAGTTGCAGTTGTCAA; and *Frmpd1* probe: 5'- FAM-CCTTTCCACCACCTGCTG-MGB, FWP: 5'-CGCCTACACCTTTTGCATGAG, RVP: 5'-GCGGGAGTCCTGTGATTCTTC. GAPDH probes from ABI were used as an internal control standard. Two-step TaqMan RT-QPCR was performed using mammary gland total RNA collected from virgin female WF and line O rats aged 12 wks. cDNA was synthesized as described above from 2 µg total RNA. For PCR reactions, RT reactions were diluted 1:8 and a microliter of this dilution (~12.5 ng RNA equivalent cDNA) was used in a 16 µl PCR reaction with ABI 7900 default cycling conditions. The reaction components were 1x TaqMan Buffer A (ABI), 3.5 mM $MgCl_2$, 200 µM each dATP, dCTP, dGTP, dTTP, 300 nM each primer, 200 nM TaqMan probe (ABI), and 0.4 U Taq Gold DNA Pol (ABI). PCR cycling conditions in the ABI 7900 were 50°C for 2 min, 95°C for 10 min, followed by 40 cycles of 95°C for 15 sec and 60°C for 1 min. FAM (*Mcs5a* region gene locus probe) and VIC (GAPDH probe) fluorescence values were measured using ABI SDS v 2.2 software when amplification reactions were in the linear phase;

quantities of transcript were measured by comparison of C_t values with those of serial dilutions used to calculate a standard curve. Measurements for each rat were an average of two replicates. Each quantity was standardized by dividing by the quantity of rodent GAPDH (ABI). Data were analyzed using the Mann-Whitney test of StatView (SAS Institute Inc., Cary, NC).

Rat *Mcs5a* Resequencing

[0073] To resequence gDNA regions, exons, 5'-UTR, and ORFs, primers spanning these elements were designed using Primer 3 (S. Rozen, H. J. Skaletsky, *Methods Mol. Biol.* 132, 365 (2000)). Spleen gDNA (~50 ng) or mammary gland cDNA (~25-50 ng) samples from WF and WKy rats were used as PCR templates. Spleen gDNA was extracted using previously published conditions. cDNA was synthesized as described above. The target gDNA or cDNA was amplified by the PCR conditions used for the verification of transcripts described above and gel-purified using Qiagen's (Valencia, CA) gel purification kit according to the manufacturer's protocol. One microliter of the gel purified suspension was used in each forward and reverse sequencing reaction (15 µl) that included 0.5 µl BigDye v3.1 (ABI), 1x reaction buffer (ABI), 0.67 µM primer. Conditions of the sequencing reaction were 95°C for 3 min followed by 32 cycles of 95°C for 30 sec and 58°C for 2.5 min. A 72°C, 7 min extension was added at the end. The sequence reactions were processed using CleanSEQ magnetic beads (Agencourt, Beverly, MA) according to the manufacturer's directions. The final volume after cleaning was 50 µl. The sequence was determined by the University of Wisconsin Biotechnology Center (Madison, WI) sequencing facility from a

1:1 dilution of the sequencing reaction. Sequence data were analyzed using Sequencher v 4.2 software (Gene Codes, Ann Arbor, MI).

Wisconsin Human DNA Collection and Extraction

[0074] As part of an on-going epidemiologic study, we recruited population-based cases of incident invasive breast cancer and community controls from Wisconsin according to a protocol approved by the University of Wisconsin Health Sciences Institutional Review Board. Cases aged 20-69 years were identified though the Wisconsin statewide tumor registry. Controls were randomly sampled from driver's license files (ages 20-64 years) and Medicare beneficiary lists (ages 65-69 years). The controls were frequency-matched in five-year intervals to have a similar age distribution as the cases. All participants were required to have an available telephone number and no previous diagnosis of breast cancer. Prior to April 2003, when changes in federal law affected the willingness of physicians to acknowledge their care of our eligible participants, physicians (identified on the tumor registry reports) were contacted prior to case enrollment to obtain information that might contraindicate study participation, such as senility.

[0075] All participants were contacted by mail prior to receiving an interviewer's call. At the conclusion of the telephone interview, individuals were asked to provide a mouthwash rinse. Those agreeing were mailed a kit and instructed to rub cheeks against teeth and swish with ~15 ml of Scope (Procter and Gamble, Cincinnati, OH) mouthwash for 30 seconds and rinse into a 40-ml Nalgene container. They were instructed to repeat this 1x. For reasons of confidentiality, participants returned consent forms in separate envelopes directly to staff at the University of Wisconsin. For the

Wisconsin sample population, 1,737 case and 1,790 control samples were collected. When samples yielded less than 5 µg of DNA, participants were asked to submit another sample.

[0076] Mouthwash from individuals was transferred to a 50-ml centrifuge tube and spun in a Beckman JS-4.2 rotor @ 3000RPM (2000xg) at room temperature for 15 minutes. Scope was poured off and DNA was extracted from the pellet using the PUREGENE cell and tissue kit from Gentra Systems (Minneapolis, MN) according to the manufacturer's protocol. DNA was hydrated in ddH_2O. The PicoGreen dsDNA Assay (Invitrogen) was used to determine DNA concentrations. Samples contained an average yield of 29.3 µg of DNA. UK samples are described (Anglian Breast Cancer Study Group, *Br. J. Cancer* 83, 1301 (2000)).

Construction of Haplotype Block Map for *MCS5A*

[0077] When data from congenic rat phenotyping first narrowed the *Mcs5a* region to 1 Mb, a haplotype block map of the human region was constructed. Ninety-one SNPs that spanned a 1 Mb region on chromosome 9 orthologous to the rat *Mcs5a* region were used to make a haplotype block map. The following SNP genotyping assays were available from ABI: rs6476611, rs1999142, rs7034763, rs2279556, rs7039784, rs1492713, rs10973239, rs13290794, rs2029646, rs12000309, rs2790063, rs3780334, rs186299, rs2087358, rs11793053, rs2235096, C__2013184_10, rs309444, rs1571234, rs1887455, rs3739576, rs2296775, rs10758435, rs17413120, rs7044153, rs4878697, rs13285217, rs10758440, rs2381718, rs1886909, rs12554736, rs13298495, rs7025444, rs10814604, rs2296553, rs1952125, rs10124071, rs2296552, rs12551499, rs7873508, rs1409145, rs1059059, rs7158, rs10973556, rs2005084, rs2025440,

rs776018, rs1976936, rs12553058, rs1105773, rs943940, rs10973637, rs2183130, rs2890783, rs11790106, rs1928246, rs1928249, rs731841, rs2585668, rs1867178, rs645259, rs716933, rs3043, rs4878806, rs11361, rs2038589, and rs3849928. Twenty-four additional SNPs were chosen from the NCBI SNP database to narrow the distance between the SNPs from ABI. Twenty-one of these had minor alleles represented in the Wisconsin population and primer sequences are as follows:

rs3747539 forward primer- AGAACACTCTTCTTCCTTTTAGGTAAATGG, reverse primer- TGTGCTTTCCGTGTCTGGAATAAA, VIC- AAGAGCTGGAGACCAG, FAM-AGAGCTGGAGCCCAG;
rs495304 forward primer-ACAGGATGACATAGTGAAGCTCTTATTAAAAAT, reverse primer- ACACAATTTAAAATAACTTGAGGCAGCAAG, VIC-CTCTAGATTTCCATTAGTATC, FAM- CTCTAGATTTCCAATAGTATC;
rs3789019 forward primer- TTCAGGAGCTTGCAGTCTAGTTG, reverse primer-TTTCATCCTGCCTTGGACAATCA, VIC- ACGTCCACCGTCCCT FAM-ACGTCCACCCTCCCT;
rs2013458 forward primer- TCTGCAACAGCTATCAAAGTTTCTGT, reverse primer-ATGCATTTTGGAACAGTGCTTTCAT, VIC- AAAGCCAGGTGTATCTA, FAM-AAAGCCAGGTATATCTA;
rs999988 forward primer- CCCCTCTGAACAGAGCCATTTTATA, reverse primer-CCACCTCTGTTTCCGCTAGAA, VIC-CATTAATGTTCAATTGAATTT, FAM-ATTAATGTTCAATCGAATTT;
rs308492 forward primer- CGTTTGAAATGTGAATGCAGTCTGA, reverse primer-GTGCTTTCCAACATAGGGCAAAA, VIC- TCTCCAACAAAATAC, FAM-CATCTCCAATAAAATAC;
rs1325916 forward primer- GGAGTATTGCCTTGCAGAAATGAAA, reverse primer-CCCCACATCAAAATATGATGCTTTATCTG, VIC- CCAGACTTTGGGAGCC, FAM-CAGACTTCGGGAGCC;
rs3780335 forward primer- CCTTGTATGGGTTTAGGATGCAGAT, reverse primer-CCCACAGAGAGTCTTTAGCTTCAC, VIC- AGTTGGTGCTTTGACCTA, FAM-TTGGTGCCTTGACCTA;
rs2182317 (SNP-3) forward primer- TCCTACTAAACAGAAGCCCCTTGTA, reverse primer- CCAACATCCCCCAGTTACTTTCATT, VIC-ATTTACTCTGCTTATTCCTGT, FAM- ATTTACTCTGCTTATGCCTGT;
rs309458 forward primer- CATTCGGTGTCCAGAGATTTCTGTA, reverse primer-CCTGTGGAAATCAAGGCTTCACTTA, VIC- CCTGGAATTCTGCTGCT, FAM-CTGGAATTCCGCTGCT;
rs1033790 forward primer- CAGCCATGTAGAGAGACCAGATT, reverse primer-GGTTTTTCCCTCCCATTGTGTAGAC, VIC- ACTGAGCTTCAGTTCC, FAM-CTGAGCTTCGGTTCC;
rs763936 forward primer- GGAGAAGCCATACTGAAGTGCAT, reverse primer-CCTTATTGCCCTAATGTTTTACTACAAATGC, VIC- CCTTGACATCTCCTTAAA, FAM-CTTGACATCTGCTTAAA;

rs10511947 forward primer- GCATGGCAAGTGTCCAAGGA, reverse primer-TGTCCAAATCCCACCCAATCTTT, VIC- CAGAGTCATAAAGCC, FAM-AGAGTCGTAAAGCC;
rs10511954 forward primer- CGGTTATCTCATGTCCAAAGCTCAT, reverse primer-CCACACTATCAATATGCCTGCTTCT, VIC- CAGCACATTAAAAGAA, FAM-CAGCACATTAGAAGAA;
rs1885491 forward primer- TCGAACATCCCATAAAGCTCATTTCTT, reverse primer-TGGAAAGCCCAGCGGAATT, VIC- CCACTTTGAAGTGTTCTGT, FAM-CACTTTGAAGTATTCTGT;
rs1138374 forward primer CTTACGAACCTGAAGGCCAAAG, reverse primer-GGGCCTGTCGTCATCCT, VIC- TGTTGACTCAGACTCGGA, FAM-TTGACTCGGACTCGGA;
rs1928233 forward primer- CGCAGTCAAACCAGACATCATC, reverse primer-CAAGATACTGCCCTTTGTGGGATAT, VIC- TGTAGAACCCTGACAATG, FAM-TAGAACCCCGACAATG;
rs920707 forward primer- GCAATGTCTTACCCCAAAGCAAGAT, reverse primer-TGAGGGTCCTGGAGGTATTCG, VIC- CCCCATCATCTCATC, FAM-CCCCATCGTCTCATC;
rs1004604 forward primer- AGACTCAGTACATTATTAGAAATGCCTTTCAC, reverse primer- ATGGAAATGTCAACTTCATTGTCCCTAT, VIC-AACTGTTTTATTTGTTAAATGTTA, FAM- CTGTTTTATTTGTTAGATGTTA;
rs885431 forward primer- CACAGCACAGGTGTGACTTG, reverse primer-CCTGCATCTTTTATGTGTCCTGGAA, VIC- CATGCCTTCCTTGGAGTA, FAM-ATGCCTTCCTTGAAGTA;
rs2073478 forward primer- ATGGCGCCGGATGGAT, reverse primer-CTCGATCCCGCTCCACTAG, VIC- CCAGGCGGTTCAG, FAM- CCAGGAGGTTCAG.
Three SNPs from the NCBI database had no variation in our population and are as follows: rs2296556 forward primer-CAGGGATGAGGGTCTGAGTTG, reverse primer-CTCTTCCAGGGCTAAAGTTGCT, VIC- CTGATGGGCCCGCCAG, FAM-CTGATGGGCCCACCAG;
rs717739 forward primer- CCACTTAAAAAACAAAGGGCCAAGTA, reverse primer-AGTTGTCTAGAGACTTGGGTTTCAGA, VIC- CCTAGAGCAACTTC, FAM-ACCTATAGCAACTTC;
rs913282 forward primer- CCACAGGCAAGAATTCCAAATGAC, reverse primer-GCTGCCTGACCATCAACACT, VIC- TTCTATCCATGGAAGCAA, FAM-ATCCGTGGAAGCAA.

[0078] To determine the linkage in the region, 39 CEPH (Centre d' etudes du Polymorphisme Humain) family grandparents (unrelated individuals with DNA available from the Coriell Institute http://ccr.coriell.org/nigms/) and a selection of 100 case-control high-DNA-yield samples from the Wisconsin population were used to estimate the haplotype phase of 88 SNPs. Ten CEPH families (each including four grandparents along with two parents and two children) were genotyped and then haplotyped by

inference from pedigree information. CEPH grandparents and 100 case-control samples were assigned haplotypes using the PHASE software package (M. Stephens, N. J. Smith, P. Donnelly, *Am. J. Hum. Genet.* 68, 978 (2001). PHASE, version 2.0 (http://www.stat.washington.edu/stephens/software.html)). A comparison of the results of the PHASE software program and inferred pedigree information of all the CEPH family members was made to determine the accuracy of the haplotype estimates. Of the 79 CEPH family members, only 6 had haplotypes with discrepancies between haplotypes inferred from pedigrees and the PHASE output with the highest posterior probability. Out of 11 total discrepancies, 10 were "switch errors" in areas of low linkage disequilibrium and one "switch error" occurred in an area of high linkage disequilibrium.

[0079] Block boundaries were defined by areas of low linkage between adjacent SNPs . Linkage disequilibrium between each pair of SNPs was tested by calculating the Lewtonin D' statistic in the CEPH grandparents and Wisconsin case-control samples. Adjacent pairs of SNPs with a D' of ≥0.89 were considered to be in the same block, while those with a D'< 0.89 were not included in the same haplotype block.

Analysis of Polymorphisms in Haplotype Block 4

[0080] The initial 1 Mb interval containing *MCS5A* was subsequently narrowed, using additional phenotyping data from congenic rats, to an orthologous human region of ~94 Kb that was completely contained in block 4. Tag SNPs listed below in haplotype block 4 were genotyped in our Wisconsin case-control population (~1,500 cases and ~1400 controls). Twelve common haplotype alleles were observed in block 4. In order to reduce the number of haplotypes for testing, block 4 was divided into part

4.1, which had 4 major variants, and part 4.2, which had 7 major variants. The division between the second and third SNPs of block 4 resulted in the fewest haplotype variants.

Genotyping Human DNA Samples

[0081] Samples from the Wisconsin and UK populations were genotyped in 5 µl reactions using 5 ng of genomic DNA from the Wisconsin samples and 10 ng of primer extension preamplification (PEP) DNA from the UK samples using ABI SNP-genotyping assays according to the manufacturer's protocol. The following Tag SNPs were used for this analysis: rs4878697 and rs13285217 (tested in only the Wisconsin samples), and rs10758440, rs999988, rs2381718, and rs2182317 (tested in both the Wisconsin and UK case-control populations). Fluorescence levels were determined using an ABI 7900 instrument. Samples were amplified on MJ Research/Bio-Rad Laboratories (Hercules, CA) thermocyclers with conditions recommended by ABI. Genotyping results are in Table 2.

Statistical Analysis of Block 4 Haplotypes:

[0082] All common haplotypes in block 4 (four in 4.1, seven in 4.2) were screened using COCAPHASE v2.403 (F. Dudbridge, *Genet. Epidemiol.* 25, 115 (2003). COCAPHASE v2.403), which calculates odds ratios and an overall p-value based on a likelihood ratio test. The haplotype block labeled 4.2 merited further investigation as a likely candidate, based on the odds ratio and 95% confidence interval. It is important to note that the minor allele of rs2182317 (risk reduction-associated allele SNP-3 in *MCS5A2*) was found only in allele 4 in block 4.2.

Statistical Analysis of Individual SNPs

[0083] The four tag SNPs that characterize the common haplotypes in block 4.2 (rs2381718, rs10758440, rs999988, rs2182317) were subjected to a battery of tests (i.e., Cochran Armitage test for trend and normal approximation of log odds ratios for heterozygotes and minor allele homozygotes) using a population of 4,376 breast cancer cases and 4,547 controls from the UK samples. The same tests were applied retroactively to the Wisconsin population, as well as the combined stratified sample.

Resequencing of *MCS5A*

[0084] The human gDNA region spanning *MCS5A1* (chr9: 37544050-37582460) and the human/rat conserved regions (60% sequence identity over 90 bases) in *MCS5A2* (chr9: 37586100-37658620) were resequenced in 24 women representative of the case-control population frequency of block 4.2 haplotype variants. DNA was submitted to Polymorphic DNA Technologies (Alameda, CA) for resequencing. The following polymorphisms had minor alleles that were observed in more than one individual, and were not listed on NCBI. Base pair positions in bold type are the SNP locations based on the human chromosome 9 May 2004 build at UCSC. The bolded IUPAC SNP symbol identifies each SNP.

MCS5A1 Polymorphisms

37545018
AATTCTCCTCGATGGAAACCTGAATTACTTTCAGTTTGGGGCTATTATGGATAATGCTGCTGTGAATATTCTTATATAAGTGTTTT
TGTGGACACGTTTTAATTTTTCTTGAGTCAATACTTAGGAGTGGAATTCCTAGGTTCTTTTCCCAGAGAGGCTATACTATTTTACA
CGCCTACCAAGAAAGAWTAAGAAACACAGTTGCTCCACATTTGTTGTGGTGTTTTTGTTTGTTTTTTGCAGGGTCACACTTTGTC
AC

37546702
CTGAGTTCAAGCCATTCTCCTACCTCAGCCTCCCAAGTAGCTGGGATTACAGGTGTGTGCCGCCACGCCCRGCTAATTTTTGTA
TTTTTAGTAGATATGGGGGTTTCACCATGTTGGCCAGGCTGGTTTTGAACTCCTGACATCAGGTGATCTGTCCGTCTTGTCCTCC
CAAAGTGCTGGGATTACAGGCGTGAGCCACCGCATCCAGCCTGTTCTGTATTCTTTAATGGAATATGTCA

37547001 24-131

AAACCACAATCTGGATATTAGACATGCTTTTTCTACTGGCTTGGTCATTGTTTCTAAGCCTTTTGCATCCTCCACTGGATTGTTTA
CGATAGGGACTGGGTACTTTTCTCTCTGAAGACTCAATACCTGGYGTTGACTCCTTTCTCTATTACAACTATTCAGGGCAACAGA
AAAGACAAGGAGAGCATTCTTAGCTACAGGGTGCCTACATTATTCCAATACTGGGTCTACGTGTCCATGCACTATGAGA

37551088, 37551094, 37551166
GGGCTTGCAGTGAGCCRAGATCRTGCCACTGCACTCTAGCCTGGGCGAAAGAGCAAGACTCCGTCTCAAAAAAACAAAACAAA
ACAAAACAAAAYCAACTTTTTTTTTTTTTTTTTTTTTAAAGAAGTGAGGTCTCTCTGTGTTGCCCAGGCTGGAGTACAGTGGCAGG
ATCATAGCTCACTGCAGCCTCGAACTCCTGGCCACAAGTGATCCTCCCACCTTAGCCTCTCAAAATACTGTGATTA

37555869
CACTCGTTAGAGAGGTGCAGAACCACTGAAGCCCAGCCCGTCCCCAGAGACTCTTGTTTTTAACCACTAAGCCAAGCTTCATGG
AGGAGCTGGTATCTGAACTGGACTCTGAAATTTGCATAGGACTGGGACATTCAGAGGAGGAAGAAAGGGGGCTATRGCAGAGG
AAACAGCATCAGCAAAAGCTCCTGAGGTAGAAAACCTTCAGCTGGGCTTAGGACATGTTGGGAGGTTCAGT

37558463, 37558666
CTCCTGACCTCAGGCAAACCGCCCACCTCAGCCTCCTAAAGTGCTSGGATTACAGGTGTGAGCCACTGGCACACATCCCCACC
CACCCAATTTACTTTGTTTTCAAGAAACCAATTTACTTTGTTTTCTCAGGTAAATTATTTACTCTCTTCCTTTTTCTCTCTGTAGGCT
AGTAAGACTCCAATCAAAGTTGATACATTGTATTTACATCTCCTCTACCCTAAGGTGGAAAAAGGATAAACGGAGTTY

37560924
CTGAGGCAGGTGAATCACGAGGTCAAGAGATGGAGACCATCCAGACCAACGTGGTGAAACTCCGTCTCTACTAAAAATACAAA
AAATTAGCTGGGCGTGGTGGCACGTGCCTGTAGTCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCGCTTGAAACCGGAAG
GCGSAGGTTGCAGTGAGCTGAGATTGCGCCACTGCAGTCCAGCCTGGCGACAGGGCAAGACTCCGT

37561238
CTCAAAAAAAAAAAAAAAAAATTATAAACAACTTTATATAAGTAAATATGAAGGGATAAACTCCTAGGAAAATATTACTTAGTTAAGA
AATTATAAAAACAAAATCACCCTAGAACCATTAAAGACATTGAGTCTGTGGTCAAAAATAAGATTGCAAAGAAAACACACCTGCG
AGCCCAGGTGGCTTCATCACCAAGTTCTATCATACATTCAAGGGACAAAGAAGGGGAAACTAAAATGGATGGTGGAAGAKGAT
GATATGGGTATT

Indel 37563886^37563887
AAAATCTTAATCAAATGTCTCTTTGTGTACATTCCTTTATGTATACAGTATCAAACATTTTAGACGAGTGGGAAACTAATAATCACT
AAACTAAAAGA[AAGA]GGGTAATGAGAGAAATTAGCAGACATTTCAAACAAAACCCATGACAGGTAACTCAAGAATAGGTTCATT
CATTAAGGCAAACTGAAAGAACAAAACACTTGTTCCATCTATTTCCTGACTGAAGCAGGTACAAGGAATTTGTTTACTTCACATCT
TCCTGTGAAAAG

37571387, 37571587
CCRATAACATAAAAAATAGATTAATATGTATTTTGTATCATTATCTATTATATACTGTATTCTTACAACACAGTAAGCTAGAGAAAA
TGTTAAGAAAATCACAGAGAGAAAAAGCATTGACAGTACTATACCGTATTTATTTAGATTCTAAGTTTAGGTTGTCTGTTTACAAG
ATTAGTCGTGTCTGAAATCAGAGCAACTAMAGCTGCAGACCTCAATCTATAGTACATATCAAGAATTCAACTTTTTCCTTTAATGT
CATGACTTTTCTC

37572952
GGAACTGCAGTATACACAATAGCCAGCAGATGTCAGAATGGAATCTCTTAAGACTCAAGTATGAAAACACAATAAATAACAGGA
GTTTCTGGAAGGACTGAGGCTTTCTTGCTCTTCTGTGCACAACACACTTCAGTCACACTGGTCTTTCTTCCTCAAATACATCTGC
TTTAACACTTTTTCAAAGATGCCCCGCTCTATCTGGCCAATTCCTATTCACCCAAACCTTACTYTATGTGTCTCTTCCTAAGG

Indel 37576899^37576900
ATATACACTTACATACCCATAAAAAATAATAAAATTATTTTAAAACCTACAAAATTTGAGAAATAAAGATACAAGTTTCAAGTTTAAA
TAGGTTCACCAAGTGCCCAC[AC]CACAGTGTATTAAAATAAACACACACTAAATGATGTCATCATGGAATATAAGCATACTAGGA
ACAAAGATCCTTGTTTCTCAGGAAGAAATACTTAGTCCCACAAAAAAATCAGGAATCAGAATGATTTCAGATTTCTCAGCAGCAA
CACTGAAAGCTAGAG

37577085
GCTTTGCTTTAAAATTCTGAAGGRAAATGATTGCTAATCTAGAACCCTATATTCATCAAACTATAGAATAAAGACATTTTTAAGGC
CAGGCACAGTGGCTCATGCCTATAATCCCAGCACTTTGGGAGGCTGAGGTGGGCAGATTACCTGAGGTCAGGAGTTCGAGAC
CAGCCTGGCCAACATGGCAAAACCCTGTCTCTACTAAAAATACAAAAATTAGCCAGGCGTGGTGGCACGTGCCTGTAAT

37579753
CAGCCTCCTGAATAGTTGGGACTATAGGCACATGCCACTATGCCTGGCTAATTTTTGTATTTTTTTAAAAGACGAGGTATCACCA
TGTTGCCAAGGCTGGTCTTGAACTCCTGAGCTCRAACAATCTGCCCGCCTCAGCCCCGCAAAGTGCTGGAATTACAGGCATGA
GCCACTGCGCCCAGCTCAATTTCTTAATATTAAACTGAAGACACTGAGATCTGTCAGAAGCTGATAAAGTCAGCA

37582143
AAAGTGCTGGGTGGGATTACAGGTGTGAGCCWACGCGCCCGGCAGCTCAAACATTCTTCAAGCTGGGAGAACCAGGATGATT
AAACACGCGCGCCGGCCACCACGTGCACTTCAGTGCCCGGACCCTGACCCGCAGACCTCAAACCGCGCATATGCCCGTCGCC
TTCAACGCGCACCCTCCTTGCTCCCCGCTACCGTTCAGCTCAGTTCGAAGGCCCCGATG

MCS5A2 Polymorphisms

37596576, 37596656
TTCTGGTGGCCCTGAGTTCTAAATTTTAAAGCCAAAGAGAGGTTGGTGTATAAAGCACCTCTTGCTAAATAGCGTTTTCTTCTGK
CACCACCACGCTGAATATGCACAAAACTCTTTTCTTTTCCCTCTTAGCTTGGCTGGAAATATAAAACTAAAATTTACTTYAAAATTA
ATAGAAAACAAAAAAGATTCCCTTTAAA

37597576
AGGTCTGTGATTGTATGTGAYCTTAGTCATCTATTTGTTACACAAGCATTTTTGGATTGTTTGAACCATTCAACACCTCTTTCCTTT
GCATATAAATGGAGATTAGGAGGATTCTTAATAGGTAGTTATTGAGCATCAAGCCTGTCATAGATATTAAGGAGAATTACAGGGT
CGGGGATTTATAGTCAAAATTCAAAGAGATGTGAAAATAATGAGAAACCTCATGGTTTAGGTTTTGATGTTAGAAGCTCGCTGCT
TCAAAGATCTA

37599480
AAGTTTAAAATGTCTAAAACTTGGCTCCTAATATTCCCTTCTCGTTCCTCTGCCCCCAAGAAACTTCCTCCCCTTCAGTTTTCCTC
ATCTCAGGTAGCGGCAAGTCCAGCCTTCAGGTTGCTGAGGCCCAAAACCTTGGAGTCATCTTTGACTCATCTCTTTCTCTCAMA
CCCTGCATCCAATCTGGTCAGCAAATCCTGTTGCTGCACCTTCATAAAGCATATCCAGAATCAGAGTACTTCTCACTCCTC

37601836, 37601857
TTTGGAAACATTTTTGACATACCTTAGTGCACAGAGCTTAATGCAGAGCTTGGCATACAGTAAGCCTTCAATGTCTATCTTCTGTA
TYCTTTTTTTCCCTTTACTCCTKATCTTAGAATGTGCAGCATTTCACAGATCTGGTTATAATGACATTGGACCTAAAGTGATTCTCT
CAGCCTAGGAGAGAGGCCAGTGGCGAT

37602870
AAGTATTTTAAAGGATTTGAGTTCACTGGAGAGATGTCCTGTTTTTAAGTGACTACGAGAACCTGGTCACTGTTYCCAATATCAG
AAGGACTGCCACGGGGCAGAGGATGAGGAGTTGTCCTTCATGGGTCCAGAGGGCAGAACTAGGACCAGTAGGGAGGTAGATG
TTAACCTGAGAATGGGAAGAACTTTAATGGCTA

37603261
AAGGCAAGGTGATTTTTCATTTTCAAAAATAATTTATTGATTAATTACCTACCATGTGCAGGACATTGTCCTAAGTGCTAGAGGAG
CCCAAATGTGAACTAGGCAAGTAGTGTGCTCCTAGAGAGCTCCTGGTGTAAGAGGGAAGATAAAAGTCATCTATCCATGCATCC
ATTCACTCAGTGAATATATACTGAGCACCT

37605662
CTGTTTCTTTAAGGTGACTTAGCAAATCTCTACTCTAATACATTATGTGTTGATGTCCTATTTTAACTCTCAAGGACAATCTCATTT
CTAATTCTTTCAGGATCACTCACTGTTGCCCTTAGTACTGACAACACTTTACAGCCTCRAGATTT

37610247
AAAAACACAAAAACCAAAGGCTTCCAGATGAGGACAGAATTCCATTTTACCCTTCACTTCTACACAAATAGGCAAAATTAGAAGT
GGAACACTCGTGTTTATCTGCCTCCAGGTCAYAGCATCACAGCAGAGTAGAAGGAGCAGTGGCTTGGCATTGCTGAGAACCAA
AAATAATGAGCAGTTTCGCCACTGACAAGGCA

37610427
AGGCTCATTTTCCTCATCTATAAAATAGGAATAAATACCATATTCTTCACAGGGCTGCTGTGAGAAYTAAATAACATGCACCATAA
ATCACTTAGTCCAGCACCGGGCATAGAGAAGGCCCTAAGTAAATGGTTGCTATTGATCATCATGATTTAAAAAAAAAAAAACT

37611591
AAGAGGATTTCTTTGGAGGAAGCAGCTGGTGTGCTAAGTGCCGCTCATGGCCCCAGGGGATAAGGAAGGGGTGTGTGGGTGC
CTATCTCACCTCACTTCAGCAGGACCACTCAGAGCTTGAGCTGTGTCTCCTGCAGTTGGGGGCCCAGGGGACTGGGTCTGACT
CCCRCCCTGGAAATTTACAGGAAAAGAGCAGGGCTGGCTAATGCTTTGGCTG

37616956, Indel 37617088^37617090
GAAGTTCGAGCAGTACTGGTTTAGAGTACACCTGCTTTGCAAGTGATWGTAAGTGTGTCTTCATTCCCTTTAATGTAACAGAGCT
CCACACATAATTGGACTTTTGTAGTCATTGCCAGTTACTCCACCTTCGAGGCGATCTTTGCTGACTCAGGTTTTCCTGCGTTTTC
CCAGAGCTG[TTT/*]CTTTCAGCTCCCATGGTC

37618560
CAGAGAAAAAGAAAGGAGGAGGAGGAGGAAAGAGAAGTCAAGTTTTAAAAGCAAACAGAAATAATAGTTTTGATGGGATGGATT
TGTTTTCTTAATAAGCCTAATGATCCCAGGAKTCTCATTGGATTATTAGCAACTGTATTTAAAATTTAAATCTAAAACAAGTTTTGT
AGCTGCAAAGTGCTTATCAGAACCTAACCATGCCTCTCTGGCT

37618765
CTTATCAGAACCTAACCATGCCTCTCTGGCTGGCGGCTAAGGGAACCTGCCTACCTGAGGGATGGCCACTTGATGTAGCATAG
TGAGGAGAAGGGACTTAGGCCTGACTTGACCAAAATCTTAGTGTGTGATCTTTCSCAAGCTCCTTCCTCCTTCCCTTCCTGGTG
CTGCAGCGTCTTCCCTATAAAATGAATAATGGA

37619515, 37619523
ACTAGATGACCGATTGTAGGCASGCCAATTRCATGCTCAGGGCAGTAGCAATGCAGGGCCACCCGCCAACATCATTTTGTCTG
GAAAACATTTGATACCTGATGCTTTCCTAAAAACATTAAAGTAATCCTCACAGGAAATATCAGAACTTTACTGGACCTCTTTATAC
TTAGTTTATAGTTCAGTGCTTTTTAAAAAAA

37634719

AGAGCTCAGTTTCCCAAATAAACCTGAATCTGAATCCCATTTACCAGCTCTGTGACCTCATACCAGTCGYTTGAATTCTCTGAGC
TTGCCTCAGTTTCATCTGTGAAGTGGGGATTGTCATGTGTGTCCTGCCTAACTCAAGCAGCTGTTGAGAGAATAAGATGAGATG
ATTGCTCTGTTTGGTGTTTTACAAACTGAAA

37635098
TAGTGCTAGAGAGGGTGGCATCTGCCCCCGGGGGCTGTGGGGAGCACGGTGGTAGTGGGGAGTTGCAGTCACCCAGCAGAT
GCAAAGCAGAATGAAGCCTATGGGAAGTGGCCAGATGTAGTTGTCAAGACCAGGGTAGGGCCACACTTGATGCTGGCCACAAC
CTCAAGACATGCCTTCTTCTGGGGAGCTGATCTCCCAGGAGCTGAGCCAGCAGRAAAAAAAAAAAAAGAGAG

37639129
AGGCTGACACTGACGCAGGTAGCTAGGAAATGGAGGGGGACAGGGCAAAATTCCATGGGGAACATTTGTACTCACCRGACAA
GAGAATAGTTCTTATCTTACTTTCTGACAAGTGCTATGCTTTTTGGTGTTTAACTGCCTTGGAGAGAGTAATTTGATGATTAATCT
TTATCTACAAAATAATTTTTAAAATTAAACTTTATTTTGAGAAAATTGTAGATTCA

37639690
AAGTGCTGCTAGTCTTCCCCCACCGCCCTTGGATGAAGGATGCACTTGTTAACATCTGCCTGTGCRCCATCCCCAGAATGATCT
AACATCCCAGTCTGATGGTGCCTCAACCCTACTCAGAACTCCTCCCATGAACCTGCCAGTGCACTGGGAACACGGTGCAAACT
CCTTGACCAGGAATTGGCCCCTGCTGACTCCCC

37641238
CTCGTGTCCCMAAGCCCGGCTCGGGGGCGCGTCCGTGGGCTAGCTCAGGACCTGCCATCGTGAACTGTTTGCACAGTAGCGA
TGTGTGCTTAAGTTTCCGAGATCTGCATGGTGTAGTGAGTGGAAACAGCGGCAACGCGGGAGGGGGAGAGACCTGGGTTCAG
GTGCGGCCGCGGCCAGCGCCAACTTTCTGAGTGACCTCAGACCAACCCTCTCTGTCCGTCCTCGATGGC

37642526
GATTTTATTTCCTTCCTCRACATCCTGGTTTTAAGTTTCTAGCATTATGAAATCAGAGAGCTGGTGAGGGTGTGCTTGGTGACCA
GAGCAGAGTTGTGGTACCTAAGTTTGTCTTTCCTGAATGAGACCCCATCCTTGAAGCTAATTTGTCCAGAGCCTTTTATCTCTCT
GACGCATTCAAATGCCA

37643111
AGGTGGTGAGGAATAGAAAGGCTATGTGTAGGTCAAGGAGCTGGTACGAGCAAGGAACTGGTGAGAGCTGGGAGAGCATGAG
ACAGATGAGAGCAGTGAGCCTGGTGGAAGCACAGAGAAGGCAGAGCACTTCAGAGACAGGGCAGCGGGARGAGGCTCTGGA
GGCTCTGAACCTGGGCTGGAATCCTCGCTGGCTGTGG

37648196
TTATTACAAAGTTCTTCATTATTCCAAGCCAAAAATCTTCCTCTCGGCAATGGTTGCCCCATTGGCCCGATCCTGTCCTTTGAAG
CCCCACACACAAAAAATCCATTTCTTCTTCCCTATGACAGCTAGTTAGGCCACTTGGACCTTCAAAGACAGGGATCAGGCCCCC
ACATCTTCACTTCTTTAGGCAGAATGCCCACTTTATCTGGCCCCCTGAAAAGCTAAGAATGGGACACCGCTGTGCAGGTGGGG
GTCTGATCAGCCCAGAYAGGAGCAAGCCCACTGCAGCCTCACTTCGGAATCTATATTTGCGTTAG

37650583
ATTTCTCTTAAAGAACAGTGATTTTAAAGTAGGTTTAAACAAYGGGTTTAAAGAACAGTGATTTTCCATTTTGACAAGGCTTGTTT
GGTATAGCCACTTCAAAATATCCCCACATCAAAATATGATGCTTTATCTGGGATATCTTTTCCAGACTTTGGGAGCCTTCCCTCCT
GCACTTTCATTTCTGCAAGGCAATACTC

37650831, 37650879
TGGGCAGGGACCGAGCCTTGCTGCTCATTYAGTACGGAGCCCAGAAGGAGGCCTGGGGGGTACAGTGTTGACCGATGRAATG
AATGAATGCATGGCTATGACTACTTCCTTAGCTGCATCTGGCTCCCAGCTGCCTTCATTGGTCTCTACAGTCATCAGTAGTATCC
TGGCTAAGTCAAGAGCATATGTGGGCCTCTGTG

37653206
GATGCGTACAACCTGTGCACACTCATCCCTCTCTTGTATTTAGTCTGCCCAATTTAATTTGTTTAGATACCGTAGGCTTTTATTGA
TTGCCCATTAGAAGTTTAGCCTAGTGTTGACAGCCTAGGGGTATAAGAGACATCATTAGCCAGAAATTTCTTTTGTAGCTGGACT
AGTAAGTCTTATACCCAAGAGTCAGTGAACAGAATKATCTGTTCACTGGGGATTGCTGAGTAAATTGGTGTGAGTGCCAGGATT
TCTGCCCCAGGCTGAGAGATCTGGATAGATGGATGCTGAAGATGGTTCTTCAGGAAGTCTTTTGGAGGAGTCTGCATTGGAGA
TGGGCCAGGTGGGGATTGCATGCCAGAGTGGAGATGCAGGGGGAGGGGGCTGATCAGCAGGT

37655211
CAACATATTTGTAATACAGACATCGAAAATAAACCAAATATCCAGCAATAGGGACTTAGCTTTAAAAATGGCACTCTTATCTATAG
CAAATAAACAATAATGTTGTAGAATAATAAAGACACAGGGAAATGTTTACAGTGTATATTTTTCAGTTTAGAAACCAGCATATATG
GTAAGTTCCCAATTATGTTGAAAATGTGTCTTTTCACACWAAAAGACTGGAAGAGTAATTAGCAACTTATAGCTTTTAGTGGACA
ATTTTTATTATCTTCTCTCCAAAAAAGAAAACTTTCAAAATTTTCTACCAGCCACATGTATTACTTTTATAAGAGGGAGGAAATAAG
TGACAATTAAGAAAAAGAGATAAGCTTTTGGAGAGCGTTGTTTGAGCGTCACTAGG

37655573
TGGAAAACTAAATGAGCACATCATTTATCTTGCAAATTCCAGACAGGTGGTGTTTTCAGAAGGAAGAGTGGTCTTAGGTCCATTT
GTGTGAGTATATTTATAGAAGTGAAAGCTTTGGGGGAAGGAAATAGATTGATTTTTTTTCCCCTTGAACTTCTGAAATTATTTTTT
CCRCTCCATTTGTAATTGAGCCCAGGGAGCTATTCTTATTTCTTCCTTTCTTGGGCACTGCGTTAGACCTAAAAATGTTAACTGG
CTTAGGATGTGGGTTTTGCTAAAATGATTCCCCTTGAAGTCTTCACTGGGCTTTCTCATGCTTAAAAGTGGGGTCCCGCAGAAG
ATCACTTTCTACCTAATGCACTTTGCTCCTGAAGTCTTTGGCAACGTTGGGGTGGTCAGA

Determining SNP bins

[0085] SNPs identified as having variation in the Wisconsin population were compiled after resequencing. The distribution of minor alleles for each SNP in the 24-member group was compared to the distribution of the haplotype block 4.2 alleles. SNPs with minor alleles that were only found in individuals carrying a copy of block 4.2 allele 4 were determined to be correlated to rs2182317 (the minor allele of SNP rs2182317 is only found in block 4.2 allele 4). SNPs at different locations were put in the same "bin" if the minor alleles were distributed similarly among the 24 women resequenced. (All SNPs highly correlated to rs2182317 were found only in the orthologous region to the rat *Mcs5a2* locus.)

Definition of Blocks 4A, 4B, and 4C

[0086] New blocks A, B, and C, based on the distribution of minor allele patterns observed after resequencing, were defined in the regions that were previously called blocks 4 and 5. Because r^2 does not relate to distance, polymorphism bins are of different sizes and multiple bins overlap each other. Blocks were defined as regions that exclusively contained sets of minor allele distribution patterns. Even though D' values across the entire region are 1 for many pairs of SNPs, r^2 values are very high for groups of SNPs only within the blocks. These blocks are based on the distribution of the SNP minor alleles instead of the D' values used in earlier versions of the map.

Genotyping New SNP Bins Identified after Resequencing

[0087] Seven polymorphisms from five additional SNP bins, located in *MCS5A1*, were genotyped in the Wisconsin samples. The primers are as follows:

indel 138-9899 (chr9: between 37,563,887 and 37,563,888) forward primer-TCAAACATTTTAGACGAGTGGGAAACT, reverse primer-

TGAGTTACCTGTCATGGGTTTTGTT, VIC- CTCATTACCCTCTTTCTTTT, FAM-TCTCTCATTACCCTCTTTT;
rs6476640 forward primer- TCCCATAGTATCCAGCACTGACA, reverse primer-AGCCTAGGGTATTCATTATTGATGAGCTA, VIC-ATCTGGGAGAGAAAT, FAM-ATCTGGGACAGAAAT;
rs10973418 forward primer- GGGAAACAAACCAATAATAATGGAAGACAA, reverse primer- GGGAGCACATTTTGAGTTGTAAGG, VIC-AGCAGGTAGCTTTT, FAM-CAGCAGATAGCTTTT;
24-131(chr9: 37548802) forward primer-GGACTGGGTACTTTTCTCTCTGAAG, reverse primer-CTTGTCTTTTCTGTTGCCCTGAAT, VIC-AGTCAACACCAGGTATT, FAM-CAACGCCAGGTATT;
rs7021977 forward primer- GGGCCACGTAGATGAGAAAGTT, reverse primer-GGAAGATGCGCGAGGATGT, VIC-CCGTATGGAGGAGATC, FAM-CGTATGGAAGAGATC;
rs6476643 (SNP-A1) forward primer-TGGGCTTCCCGACCAC, reverse primer-CCGGTGCCCCTCCTG, VIC-CAGATACAGCTTTTATGCGC, FAM-AGATACAGCTTTTCTGCGC;
128-249 Forward primer –CATCTCCTCTACCCTAAGGTGGAAA, Reverse Primer-CGACCTTAGACCACGTGAACTC, VIC- ACCACATGAACTCC, FAM-ACCACATAAACTCC;
rs10117312 Forward Primer- CCTTCGGAGGCCTTGGAT, Reverse Primer-CTTTGTGAGGAGAAAAGCAACATTCA,
VIC- CTGCCTCTGATTATT, FAM- TCTGCCTCTAATTATT.
SNPs were run using ABI Assays-by-Design.

[0088] SNP 10117312 was not used in an association test because it has an r^2 of 1 with rs6476640. These data are uncorrected for multiple comparisons, because only one SNP (rs6476640) was tested in the combined Wisconsin and UK populations. SNP rs6476643 (SNP-A1) was analyzed in the Wisconsin population along with the SNPs, where it is subjected to the same battery of statistical tests described above.

Quality Control for Human Sample Genotyping

[0089] Quality control of the Wisconsin samples was conducted with DNA from 85 subjects who had submitted two independent samples. These duplicate samples were genotyped for all polymorphisms tested in the entire Wisconsin population; 95% (1648/1729) of the genotypes were identical for the two samples, 5% (81/1729) of the samples had a call for one, but a no call for the other sample. There were no sample

sets that resulted in mismatched calls. Wisconsin samples that were genotyped twice (initially when the original haplotype block map was made and again when a subset of these SNPs was genotyped in the entire population) yielded the same genotype 98.4% (1794/1823) of the time. Only, 1.5% (27/1823) had a no call in one of the replicates. Mismatches occurred in 0.1% (2/1823) of the replicates. Due to a small number of women of non-European descent in the Wisconsin set (73 cases, 90 controls), and an effort to avoid population stratification, these samples were not included in the study. For the UK samples each 384-well plate had 12 samples that were duplicated on a separate plate. All genotyping calls for the first four SNPs tested in the UK population were in agreement for each duplicate, except for one genotype of an individual with no call. The call rate for SNP rs6476643 was 0.981 in the UK study set.

CLAIMS

We claim:

1. A method of analyzing breast cancer susceptibility or resistance, comprising the steps of

a) obtaining a DNA sample from a patient,

b) genotyping the sample for the presence of the minor allele of single nucleotide polymorphisms rs6476643 or rs2182317 or a minor allele correlating at r^2 >0.7 with rs6476643 or rs2182317, and

c) correlating the identity of the allele with risk assessment data.

2. The method of claim 1 wherein the minor allele is a minor allele of rs6476643 or rs2182317.

3. The method of claim 1 wherein the minor allele is a minor allele selected from the group consisting of

rs2182317
rs6476643
138-9899
rs10758441
rs7042509
114-117
d3-169
rs12378421
r3-116
rs17505776
rs4878708
rs4878709
rs4878710
rs10973450
l4-70
m4-218
rs4490927
x4-77
z4-66
f5-152
rs4878713
y5-43
i6-31
i6-103

4. The method of claim 3 wherein the minor allele is a minor allele correlating at r^2 >0.7 with the minor alleles of claim 3.

5. The method of claim 1 wherein the DNA sample is collected from mouthwash after the patient has rinsed his or her mouth with mouthwash.

6. The method of claim 1 wherein the patient is found heterozygous or homozygous for the minor allele at rs2182317.

7. The method of claim 6 wherein the correlation indicates that the patient has a 14% reduction in risk.

8. The method of claim 1 wherein the patient is homozygous for the minor allele of rs6476643.

9. The method of claim 8 wherein the correlation indicates that the patient has a 19% increase in risk.

10. The method of claim 1 wherein the DNA sample is genotyped for the presence of a minor allele that correlates to either rs6476643 or rs2182317 at $r^2 > 0.9$.

11. A method of finding minor alleles that correlate to human breast cancer susceptibility or resistance, comprising the steps of

a) sequencing human genomic regions MCS5A1 or MCS5A2,

b) identifying at least one polymorphism,

c) genotyping individuals for the presence or absence of the polymorphism and,

d) determining whether the polymorphism correlates to rs212317 (plus strand allele T) or rs6476643 (plus strand allele T) with the r^2 statistic at $r^2 > 0.7$.

12. The method of claim 11 wherein step (d) comprises determining whether the polymorphism correlates at $r^2 > 0.7$ to the group consisting of

rs2182317
rs6476643
138-9899
rs10758441
rs7042509
114-117
d3-169
rs12378421
r3-116
rs17505776
rs4878708
rs4878709
rs4878710
rs10973450
l4-70
m4-218
rs4490927
x4-77
z4-66
f5-152
rs4878713
y5-43
i6-31
i6-103

13. The method of claim 11 wherein the polymorphism correlates to rs212317 (plus strand allele T) or rs6476643 (plus strand allele T) with r^2 >0.9.

14. A method of treating or preventing breast cancer comprising mimicking the biological effect of *MCS5A1 and MCS5A2* alleles that increase/decrease susceptibility to breast cancer.

15. A method of assaying molecules for their ability to interact with the *FBXO10* and *FMRPD1* gene loci or the *FBXO10* and *FMRPD1* gene loci products comprising the steps of exposing a candidate molecule to the *FBXO10* and *FMRPD1* gene loci or gene loci products and determining whether the candidate molecule has interacted with the loci or the gene products.

FIG. 1

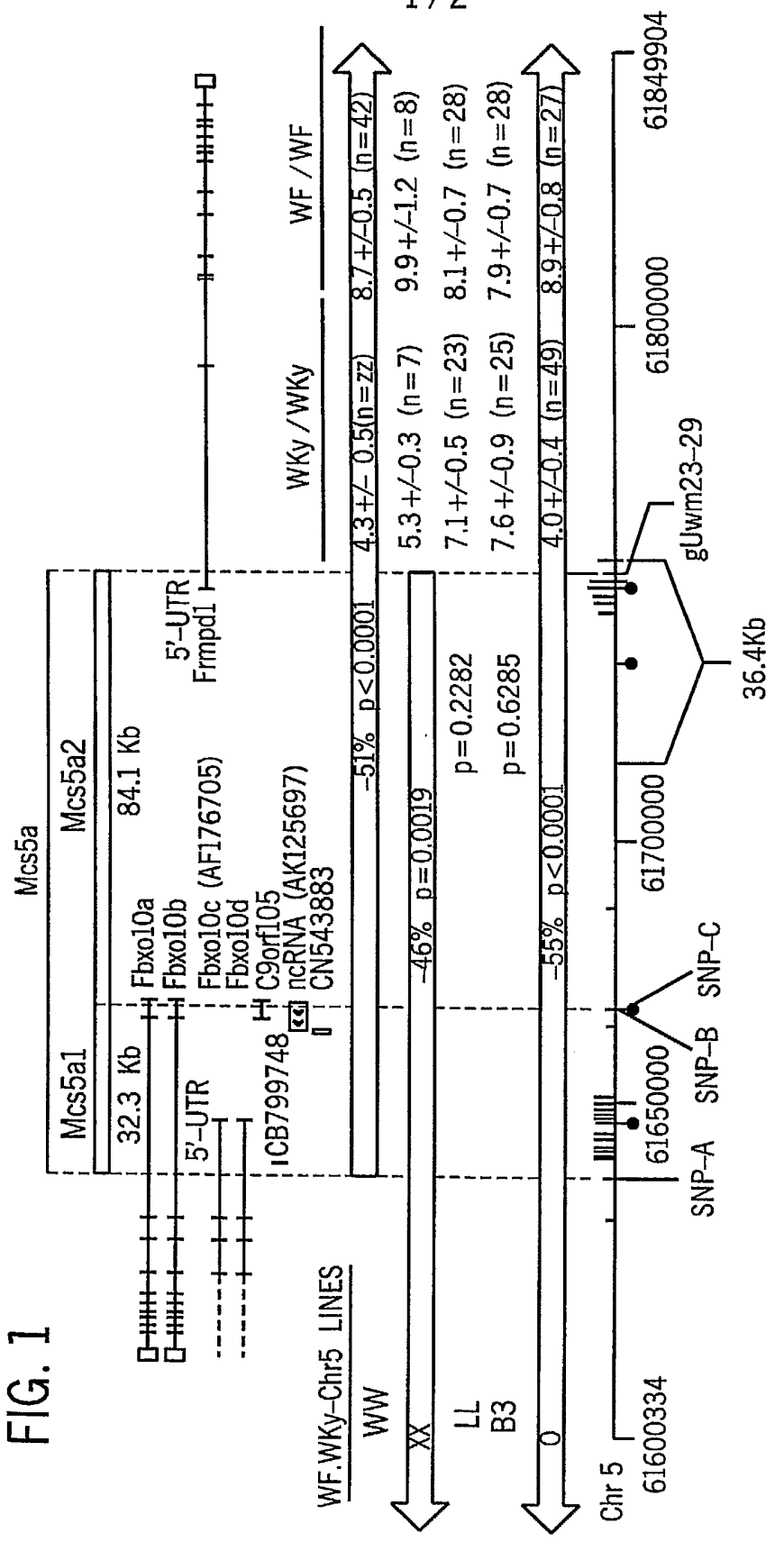
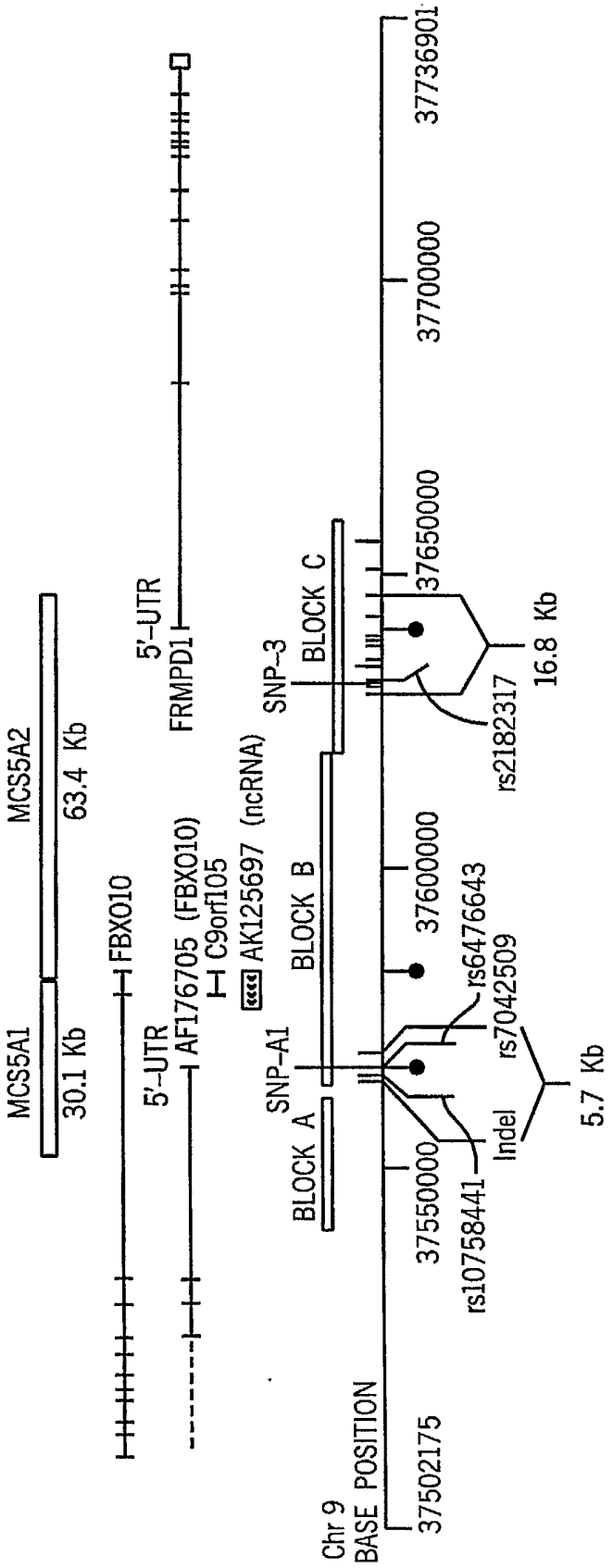

Mcs5a
Mcs5a1
Mcs5a2
32.3 Kb
84.1 Kb
Fbxo10a
Fbxo10b
Fbxo10c (AF176705)
Fbxo10d
C9orf105
ncRNA (AK125697)
CN543883
CB799748
5'-UTR
5'-UTR
Frmpd1
WKy / WKy
WF / WF
4.3 +/- 0.5(n=zz)
8.7 +/-0.5 (n=42)
-51% p<0.0001
5.3 +/-0.3 (n=7)
9.9 +/-1.2 (n=8)
p=0.2282
7.1 +/-0.5 (n=23)
8.1 +/-0.7 (n=28)
-46% p=0.0019
p=0.6285
7.6 +/-0.9 (n=25)
7.9 +/-0.7 (n=28)
-55% p<0.0001
4.0 +/-0.4 (n=49)
8.9 +/-0.8 (n=27)
WF.WKy-Chr5 LINES
WW
XX
LL
B3
0
Chr 5
61600334
61650000
61700000
61800000
61849904
gUwm23-29
36.4Kb
SNP-A
SNP-B
SNP-C

FIG. 2


MCS5A1
30.1 Kb
MCS5A2
63.4 Kb
FBXO10
5'-UTR
AF176705 (FBXO10)
C9orf105
AK125697 (ncRNA)
5'-UTR
FRMPD1
SNP-A1
BLOCK A
BLOCK B
SNP-3
BLOCK C
rs10758441
rs6476643
rs7042509
Indel
rs2182317
16.8 Kb
5.7 Kb
37502175
37550000
37600000
37650000
3700000
3736901
Chr 9
BASE POSITION