Genetic Loci Controlling Breast Cancer Susceptibility in the Wistar-Kyoto Rat

Hong Lan,* Christina M. Kendziorski,† Jill D. Haag,* Laurie A. Shepel,*
Michael A. Newton† and Michael N. Gould*

*McArdle Laboratory for Cancer Research and †Department of Biostatistics and Medical Informatics,
University of Wisconsin, Madison, Wisconsin 53792

ABSTRACT

In this study, the Wistar-Kyoto (WKy) rat was genetically characterized for loci that modify susceptibility to mammary carcinogenesis. We used a genetic backcross between resistant WKy and susceptible Wistar-Furth (WF) rats as a panel for linkage mapping to genetically identify mammary carcinoma susceptibility (*Mcs*) loci underlying the resistance of the WKy rat. Rats were phenotyped for DMBA-induced mammary carcinomas and genotyped using microsatellite markers. To detect quantitative trait loci (QTL), we analyzed the genome scan data under both parametric and nonparametric distributional assumptions and used permutation tests to calculate significance thresholds. A generalized linear model analysis was also performed to test for interactions between significant QTL. This methodology was extended to identify interactions between the significant QTL and other genome locations. Chromosomes 5, 7, 10, and 14 were found to contain significant QTL, termed *Mcs5*, *Mcs6*, *Mcs7*, and *Mcs8*, respectively. The WKy alleles of *Mcs5*, *-6*, and *-8* are associated with mammary carcinoma resistance; the WKy allele of *Mcs7* is associated with an increased incidence of mammary cancer. In addition, we identified an interaction between *Mcs8* and a region on chromosome 6 termed *Mcsm1* (modifier of *Mcs*), which had no significant main effect on mammary cancer susceptibility in this genetic analysis.

A N individual woman's susceptibility to developing breast cancer is influenced both by inherited genes or alleles and environmental factors. Inherited alleles that influence breast cancer include those that are rare in a population but have a high genetic penetrance, such as the suppressor genes *BRCA1* and *BRCA2* (WARNER *et al.* 1999). In addition, other inherited alleles also influence breast cancer susceptibility but have a lower penetrance and the potential to have a higher population frequency. This can be illustrated by examining the effect of the family history of breast cancer on individual susceptibility. Women with a first degree relative affected with breast cancer that carries a functional mutation in *BRCA1* or *BRCA2* bear an increased risk of developing breast cancer with an average relative risk (RR) of 5.2. Importantly and occurring more commonly, women with first degree relatives with breast cancers that do not harbor *BRCA1* or *BRCA2* mutations also have a higher relative risk (RR = 1.7) of developing breast cancer (WARNER *et al.* 1999).

We hypothesize that alleles that confer increased susceptibility to breast cancer make up only one side of a distribution of susceptibility-gene potency. The other half of this distribution contains alleles that confer resistance to breast cancer. Identifying such alleles would

be of value both for risk estimation as well as providing targets for the development of chemoprevention drugs. It has been possible to genetically identify genes such as *BRCA1* and *BRCA2*, which increase susceptibility to breast cancer, on the basis of family linkage studies. Family linkage studies also have the potential to identify additional alleles that increase breast cancer risk by either having a high penetrance or function in a well-defined population of high risk women. In contrast, it would be much more difficult to genetically identify alleles that decrease risk on the basis of human population linkage studies. This is in part due to the fact that it would be difficult to label families that have a low incidence of breast cancer as having an inherited allele yielding this low risk as opposed to having good fortune.

On this basis, we have chosen an alternative approach to genetically identify alleles that confer a resistance phenotype to breast cancer. We chose to first identify such resistance alleles in the rat. The rat was chosen over the mouse in that rat mammary cancer better models human breast cancer in terms of hormonal responsiveness, as well as histopathological stages of development and carcinoma morphology. We have focused on two rat strains that show resistance to chemically induced mammary carcinogenesis. These are the Copenhagen (Cop) and Wistar-Kyoto (WKy) rat strains. We have previously reported the genetic identification of the loci that control resistance in the Cop rat (SHEPEL *et al.* 1998a). Four loci were identified: mammary carcinoma susceptibility (*Mcs*) *1*, *-2*, *-3*, and *-4*, of which three

Corresponding author: Michael N. Gould, McArdle Laboratory for Cancer Research, University of Wisconsin Medical School, 1400 University Ave., Madison, WI 53706-1599.
E-mail: gould@oncology.wisc.edu

contribute to the tumor resistance phenotype and one of which has the potential to increase susceptibility. It is unknown if these four loci are part of a very large set of mammary cancer susceptibility loci or whether, on the other extreme, they represent the complete set of modifier loci in rats. We have also shown that the WKy rat strain has a resistance to 7,12-dimethylbenz-[*a*] anthracene (DMBA)-induced mammary carcinogenesis (HAAG *et al.* 1992) similar to that of the Cop rat. Of all the rat strains we tested for susceptibility to DMBA-induced mammary carcinogenesis, the WKy was the most resistant (SHEPEL and GOULD 1999; J. D. HAAG and M. N. GOULD, unpublished results). To begin to address questions regarding the population frequency of rat resistance loci and to genetically identify additional loci that confer lower susceptibility, we characterized the genetic basis for inherited mammary cancer resistance in the WKy rat strain.

MATERIALS AND METHODS

Genetic cross and genotyping: Inbred WKy and Wistar-Furth (WF) rats were purchased from Harlan Sprague-Dawley Inc. (Indianapolis). All rats were maintained in our animal care facility on a 12-hr light/dark cycle and were provided Teklad lab blox chow and acidified water *ad libitum*. Female WKy rats were bred with male WF rats to produce (WKy × WF)F_1 rats. F_1 male and female rats were then bred to WF female and male rats to produce (WKy × WF)F_1 × WF or WF × (WKy × WF)F_1 backcross rats. At weaning, a tail tip of each female backcross rat was removed for genomic DNA extraction. At 50–55 days of age, the female rats were given DMBA (Acros Organics; Fisher Scientific, Pittsburgh) by gastric intubation in a single dose of 65 mg/kg body weight. At 17 wk after DMBA administration, the rats were removed from the study and the number of mammary carcinomas ($\geq 3 \times 3$ mm in diameter) was scored for each rat. A total of 363 rats were scored for tumor development. All the mammary tumors observed at necropsy were identified as carcinomas. For the genome-wide linkage analysis, selective genotyping (LANDER and BOTSTEIN 1989) was performed for a subset of 194 backcross rats with extreme phenotypes, consisting of 94 rats with ≤ 1 tumor and 100 rats with ≥ 6 tumors. Parental and F_1 samples were also genotyped as controls.

Published rat microsatellite markers, including *Mgh*, *Mit*, *Rat*, and *Wox* markers, were obtained from Research Genetics (Huntsville, AL). *Uwm* markers newly generated in our laboratory (SHEPEL *et al.* 1998b; LAN *et al.* 1999) were also used. It should be pointed out that the WKy rats used in this study, which were purchased from Harlan Sprague-Dawley and therefore should be designated substrain WKy/HSD, are genetically distinct from the WKy/N substrain in the rat genome database (http://www.informatics.jax.org/rat). Polymorphisms between the WKy/HSD and the WF rats were thus determined in our laboratory. The initial genome scan was carried out by the National Heart, Lung, and Blood Institute (NHLBI), National Institutes of Health Mammalian Genotyping Service (http://www.marshmed.org/genetics/Genotyping_Service) at the Center for Medical Genetics, Marshfield Medical Research Foundation, Marshfield, Wisconsin. Complementary genotyping to complete this scan was performed in our laboratory using 5-μl PCR reactions and agarose gel electrophoresis (LAN *et al.* 1999). Briefly, the PCR reaction contained 50 ng of genomic DNA, 10 mM Tris-HCl, pH 8.3, 1.5 mM MgCl$_2$, 50

mM KCl, 0.001% w/v gelatin, 200 μM dNTPs, 132 nM of each primer, and 0.2 units AmpliTaq polymerase (Perkin-Elmer, Foster City, CA). The following cycling conditions were used: denaturation at 94° for 3 min, followed by 30 cycles of 94° for 1 min, 50°–60° (depending on specific primers) for 1 min, and 72° for 30 sec, and then an extension at 72° for 5 min. The PCR products were resolved on 2.5–3.5% MetaPhor agarose (FMC BioProducts, Rockland, ME) gels in 1× TBE buffer. The gels were stained with SYBR Gold (Molecular Probes, Eugene, OR) and the genotypes were visualized using a Fluor-Imager SI (Molecular Dynamics, Sunnyvale, CA).

Linkage analysis: Unless otherwise specified, all statistical analyses were carried out in *S-PLUS* (MATHSOFT 1997). Genome scan data were analyzed using interval mapping (IM) methods under both parametric (P; LANDER and BOTSTEIN 1989) and nonparametric (NP; KRUGLYAK and LANDER 1995) assumptions. Since the parametric method assumes a Gaussian-distributed phenotype, a square root transformation was applied to stabilize the variance (MCCULLAGH and NELDER 1989). The nonparametric method described in POOLE and DRINKWATER (1996) and implemented in Q-link (DRINKWATER 1997) was also used. In addition, IM using a negative binomial (IM-NB)-distributed phenotype was considered. The negative binomial distribution is often used to model count data, particularly when the data are overdispersed. Such is the case in many studies of tumor count data (DRINKWATER and KLOTZ 1981).

To determine an exact threshold value for each mapping method, we used permutation tests as described by CHURCHILL and DOERGE (1994). Specifically, tumor counts were randomly reassigned to the marker genotype vectors by permutation and LOD profiles were recomputed for each mapping method. This was repeated 1000 times. For presentation, LOD profiles for the observed data were divided by the 95th percentile of the permutation distribution of maximum LOD scores to produce a relative significance profile. Regions in which the relative significance profile exceeds unity are referred to as significant QTL, or QTL with significant main effects.

Interaction identification: To identify interactions between quantitative trait loci (QTL), generalized linear models were used (MCCULLAGH and NELDER 1989). Expected tumor number was a function of both additive QTL effects and possible pairwise interactions. Markers closest to the LOD peaks from interval mapping were used as surrogates for the QTL, and an optimal model was selected using the Bayes information criteria (BIC; SCHWARZ 1978; KASS and RAFTERY 1995). The BIC provides a balance between goodness of fit of a model to data and the number of model parameters. Let z_j represent the number of WKy alleles at the selected marker on chromosome *j*. Then, the expected tumor count, μ, is modeled by

$$\mu = \exp\left(\alpha_0 + \sum_j \alpha_j z_j + \sum_{j,k} \delta_{j,k} z_j z_k\right). \qquad (1)$$

Here, the α_0 term represents the baseline tumor rate for animals with no WKy alleles. The main effect term, α_j, represents the effect of having a WKy allele at the chosen marker on chromosome *j*, while the interaction term, $\delta_{j,k}$, quantifies the effect of having a WKy allele at each chosen marker on chromosomes *j* and *k*. The first sum in (1) is over chromosomes deemed to harbor significant QTL on the basis of interval mapping. The second sum considers interactions between any of these significant QTL and other genome regions. In particular, a genetic locus may have little or no main effect, but may interact with one of the significant QTL. If there is an interaction between a significant QTL and another QTL (say Q2) that was not identified as significant in the initial genome scan, then the LOD profile on the chromosome containing Q2 is expected to be highest in the Q2 region, provided no

other QTL are on that chromosome (see APPENDIX). Therefore, one marker at the highest point of the linkage profile on each chromosome (excluding the chromosomes harboring the significant QTL) was chosen as a candidate to test for interactions with the peak main effect markers. The model was first fit with only main effect terms for *Mcs5-8*. Terms were then added if they lowered the BIC and removed if they raised the BIC. By this stepwise method, the model with the lowest BIC out of all models considered was identified as the best model. Missing genotypes were imputed 50 times by sampling from their distributions, conditional on flanking marker genotypes. We report only significant effects that arise in the majority of these imputations.

RESULTS

Tumor multiplicity: The tumor multiplicity in the parental WKy, WF, (WKy × WF)F_1 and the (WKy × WF)F_1 × WF or WF × (WKy × WF)F_1 backcross rats is shown in Figure 1. In a previous study, 30 WKy rats developed 1 tumor after 30 wk (HAAG *et al.* 1992); the 17 WKy rats in this experiment developed no tumors at 17 wk after DMBA administration. In contrast, the WF rats developed an average of 7.3 tumors per rat. The F_1 rats ($n = 33$) developed 1.1 carcinomas on average, and the backcross rats ($n = 363$) developed an average 3.9 tumors per rat.

Genotyping and QTL analysis: The NHLBI Mammalian Genotyping Service provided genotypes for 115 microsatellite markers throughout the genome (excluding the Y chromosome). In addition, we genotyped 130 additional markers to increase the resolution of the genome scan, especially around potential QTL. After removing markers with incomplete genotypes (*i.e.*, those for which >10% of the rat samples tested yielded no PCR product), the current data set contained 228 markers. The average level of completeness for the 228 markers was 96%. Multipoint linkage analysis (LANDER *et al.* 1987) indicated that all the markers fell into one of the 21 chromosomal linkage groups, excluding the Y chromosome. The average space between adjacent markers over the entire genome is 8.0 cM (Kosambi mapping function; KOSAMBI 1944). The average marker density for chromosomes with QTL is 4.0 cM, while that for the rest of the genome is 11.0 cM. There is not a single gap over 20 cM. The markers are roughly evenly distributed over the genome as shown by the distribution of the vertical ticks on the x-axis in Figure 2.

Our data analysis identified four significant QTL on chromosomes 5, 7, 10, and 14, respectively (see Figures 2 and 3 and Table 1). In Figures 2 and 3, the LOD profiles were normalized by threshold values to give a relative significance profile. This was done to facilitate comparison across profiles since different assumptions concerning the phenotype distribution give rise to different threshold values. Specifically, the threshold values determined by permutations for each of the four methods are 2.962 (IM-P), 2.695 (IM-NP), 2.663 (Q-link), and 2.083 (IM-NB). Thus, a relative significance value
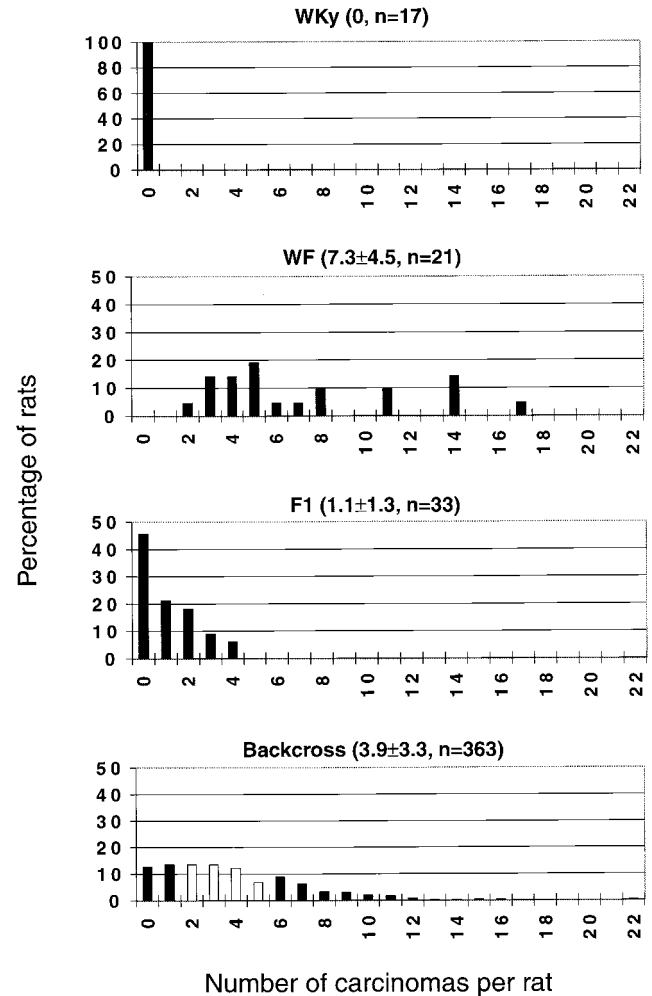

FIGURE 1.—Tumor distribution of WKy, WF, F_1, and backcross rats. The rats were killed at 17 weeks after DMBA administration and the number of mammary carcinomas (3 × 3-mm diameter or larger) was scored for each rat. The average tumor number, the standard deviation, and the number of animals for each group are shown on the top of each graph. The open bars in the backcross graph represent rats that were not included in linkage analysis ($n = 169$ rats, 363 total—194 genotyped).

of 1 indicates a raw LOD score of 2.962, 2.695, 2.663, and 2.083 for IM-P, IM-NP, Q-link, and IM-NB, respectively. Table 1 shows the raw LOD scores in the four peak regions for each mapping method and the mean tumor number of the two genotypic classes. The four QTL identified as significant are termed *Mcs5, -6, -7,* and *-8*, respectively (Figures 2 and 3), following *Mcs1–4* defined in the Copenhagen (Cop) rat (SHEPEL *et al.* 1998a). The minimum resolutions, *i.e.*, the largest mapping distances between adjacent markers, are 5 cM around *Mcs5* and 10 cM around *Mcs6, -7,* and *-8*.

The WKy alleles of *Mcs5, -6,* and *-8* act to suppress tumor multiplicity. As seen in Table 1, rats with the genotype of WKy/WF at each of the *Mcs5, -6,* and *-8* loci have a lower tumor multiplicity than rats with the WF/WF genotypes. For *Mcs7*, the direction of gene func-
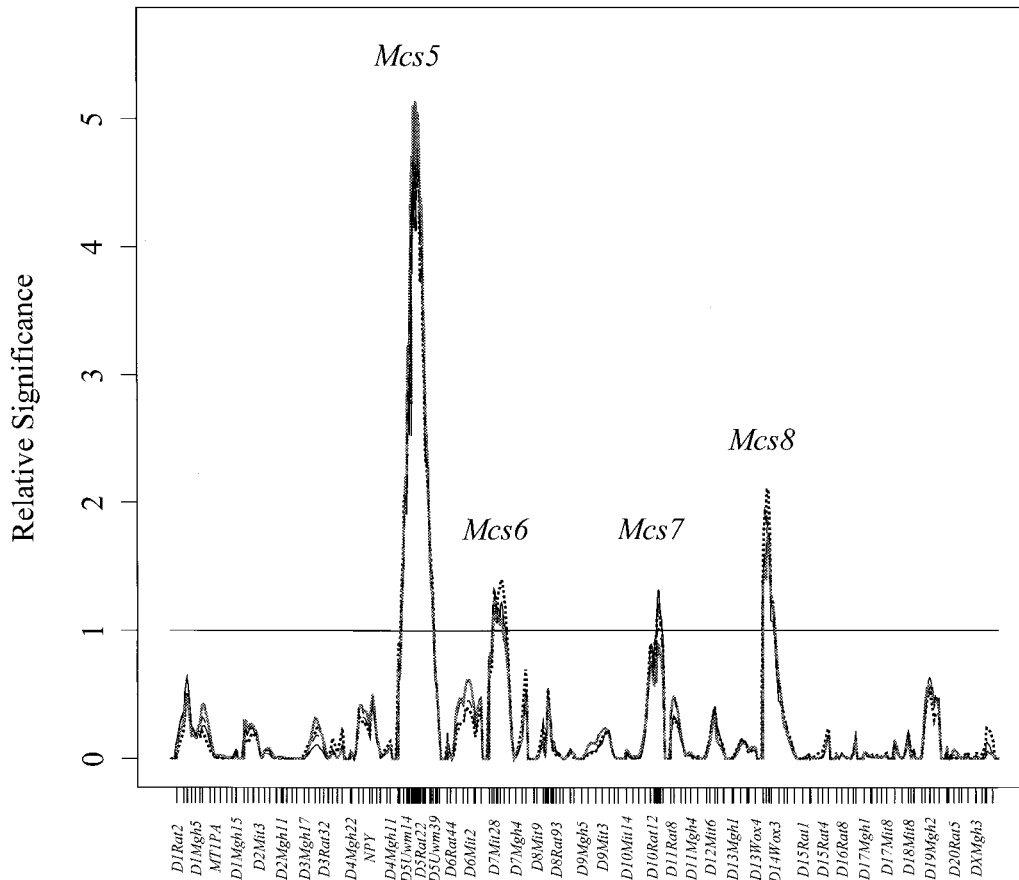
Figure 2.—Genome scan profiles from interval mapping assuming Gaussian (IM-P, —) and Negative Binomial (IM-NB, ···) distributed phenotypes. Nonparametric methods (IM-NP, - - - and Q-link, – – –) were also used. Each profile is divided by the 95th percentile of the permutation distribution of maximum LOD scores to produce a relative significance profile. A relative significance value of 1 equals a LOD score of 2.962 (IM-P), 2.695 (IM-NP), 2.663 (Q-link) and 2.083 (IM-NB). The 228 markers used in the genome scan are represented with vertical ticks on the *x*-axis. The names of only a selected few markers are shown.

tion is the opposite. The data for all four loci were analyzed using a generalized linear model as described in the *Interaction identification* section in materials and methods.

Gene interaction: The best model as assessed by the BIC criterion includes the four main effects from *Mcs5*, *-6*, *-7*, and *-8* and a significant first-order interaction between a region on chromosome 6 and *Mcs8*. The resulting parameter estimates are listed in Table 2 and are the only parameters remaining in the final model determined by the model selection procedure. As shown in Table 2, the estimated interaction coefficient is positive. This indicates that an animal with the WKy allele at the two interacting peak markers, *D6Mit2* and *D14Wox3* (for *Mcs8*), will have (on average) a higher tumor number than that expected from the combined additive effects of the WKy allele at each marker individually. This is in fact the case as shown in Table 3, which gives average tumor numbers within genotype classes. Consider the WF/WF groups at markers *D6Mit2* and *D14Wox3*. Substitution of the WKy allele at marker *D14Wox3* results in an average decrease (from 5.5 to 1.9) of 3.6 tumors, whereas substitution of the WKy allele at marker *D6Mit2* increases (from 5.5 to 6.7) tumor number an average of 1.2. If purely additive, one would expect an average decrease of ~2.4 tumors following substitution of WKy alleles at both markers (dif-

ference between 3.6 and 1.2); instead, substitution of the WKy allele at both markers results in an average decrease of only 1.6 tumors (from 5.5 to 3.9; Table 3). In this article, we have defined the *D6Mit2* region as *Mcsm1*, a modifier of an *Mcs* gene.

An additional potential interaction was also suggested by this analysis. A model in which interactions between *D5Rat22* (*Mcs5*) and *D10Rat12* (*Mcs7*) was included. While not fulfilling the criterion of having the lowest BIC, the model did consistently have a lower BIC than did a model having only main effects for *Mcs5–8* (Table 3).

DISCUSSION

We have identified four mammary carcinoma susceptibility loci, *Mcs5*, *Mcs6*, *Mcs7*, and *Mcs8*, on rat chromosomes 5, 7, 10, and 14, respectively. The WKy alleles of *Mcs5*, *-6*, and *-8* act to suppress tumor multiplicity, while *Mcs7* increases susceptibility to mammary carcinogenesis. In addition, an *Mcs* modifier, *Mcsm1*, was detected on chromosome 6. The WKy allele of *Mcsm1* partially counteracts the tumor resistance conferred by the WKy allele of *Mcs8*.

Mcs5 is located on rat chromosome 5. Significant linkage was observed in a very large region of this chromosome. This suggested that this locus may contain more
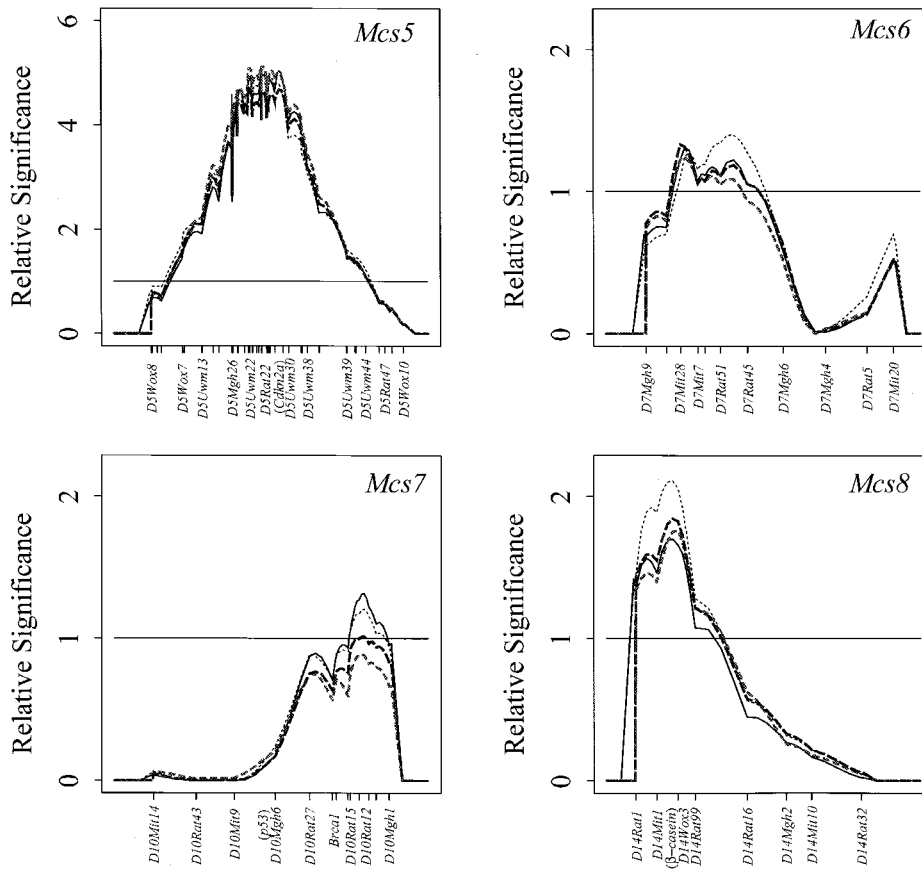
FIGURE 3.—Profiles of chromosomes 5, 7, 10, and 14, containing *Mcs5*, *-6*, *-7*, and *-8*, respectively. The key is the same as that in Figure 2. The gene markers included in parentheses were not mapped in our backcross; rather, approximate locations of these markers were deduced from other genetic maps.

than a single gene that modifies susceptibility to mammary carcinogenesis. An integrated linkage/cytogenetic map (http://ratmap.gen.gu.se) shows that the *Mcs5* region covers rat chromosome 5q21–5q35. This region is conserved in the mouse but split into two pieces in the human genome (GAUGUIER *et al.* 1999). By using the comparative mapping tool in the Mouse Genome Database (http://www.informatics.jax.org), the *Mcs5* region was found to be homologous to the mouse chromosome 4 linkage map from 10 to 50 cM. Homologous regions

of the human genome include chromosomal regions 9p13–9p34 and 1p31–1p35.

Mcs6 is located on rat chromosome 7, with peak linkage at marker *D7Mit28*. Significant linkage was found in an ~30-cM area from marker *D7Mit28* to *D7Rat45*. This region is homologous to mouse chromosome 10 (50–70 cM) and human chromosome 12q (GAUGUIER *et al.* 1999).

Mcs7 shows a sharp peak linkage at *D10Rat12* on the distal end of chromosome 10. The subtelomeric region

TABLE 1

Raw LOD scores in peak regions of *Mcs5–8* and their phenotypic effects

QTL	Peak marker	LOD score				Ave$_{(WKy/WF)}$	Ave$_{(WF/WF)}$
		IM-P	IM-NP	Q-link	IM-NB		
Mcs5	*D5Rat22*	14.53	13.73	12.28	10.51	2.16	6.68
Mcs6	*D7Mit28*	3.63	3.31	3.54	2.29	3.34	5.65
Mcs7	*D10Rat12*	3.78	2.38	2.65	2.45	6.00	3.31
Mcs8	*D14Wox3*	4.91	4.74	4.85	4.28	2.93	6.00

IM-P and IM-NB denoted interval mapping under Gaussian and negative binomial assumptions, respectively; IM-NP denotes nonparametric interval mapping. Q-link is the nonparametric mapping method introduced by DRINKWATER (1997). Threshold values at the 95th and 99th percentiles are, respectively, 2.962 and 3.557 for IM-P, 2.695 and 3.482 for IM-NP, 2.663 and 3.409 for Q-link, and 2.083 and 2.657 for IM-NB. Ave$_{WKy/WF}$ is the mean tumor number for rats with genotype WKy/WF at the given peak marker, while Ave$_{WF/WF}$ is the mean tumor number for rats with genotype WF/WF at the given peak marker. Note that the means were calculated only on the basis of genotypes of the markers listed—genotypes in other regions were not considered.

TABLE 2

Parameter estimates for the negative binomial model

Parameter	exp(estimated parameter) (% change)	exp(est ± 2 × SE)
α_5-*Mcs5*	0.364 (−64)	(0.270, 0.492)
α_7-*Mcs6*	0.664 (−34)	(0.492, 0.896)
α_{10}-*Mcs7*	1.896 (+90)	(1.405, 2.560)
α_{14}-*Mcs8*	0.375 (−63)	(0.247, 0.571)
$\delta_{14,6}$	2.034 (+103)	(1.116, 3.706)

The exp(estimated parameter) value represents the multiplicative effect of having the WKy allele at the *Mcs* location represented by the corresponding parameter; the numbers in parentheses indicate the effective percentage decrease or increase in expected tumor number. The last column represents the confidence interval for exp(estimate). The percentage decrease corresponding to the interaction coefficient should be considered in the presence of the WKy allele at *Mcs8*. In particular, $\exp(\alpha_{14} + \delta_{14,6}) = 0.76$, which corresponds to an effective 24% decrease in tumor number.

of chromosome 10 contains the rat homologue of the human breast cancer predisposing gene *Brca1* (CHEN *et al.* 1996); however, *Brca1* is unlikely to be a candidate for *Mcs7*. Although it is only 10 cM proximal to the peak marker *D10Rat12*, the linkage profile at *Brca1* drops below a significant level at this point (Figure 3). The tumor suppressor gene *p53* is also located on chromosome 10 (CANZIAN *et al.* 1996). It is not a candidate because it is ~40 cM proximal to *D10Rat12*.

Mcs8 is located on the subcentromeric region of chromosome 14, with peak linkage at *D14Wox3* and significant linkage from *D14Rat1* to *D14Rat99* (20 cM). This region is homologous to human chromosome 4q11–21 (GAUGUIER *et al.* 1999). Interestingly, the casein gene family genomic region (GEORGE *et al.* 1997), which is mapped using the α-casein marker *Casag* or *D14Wox14* (GAUGUIER *et al.* 1999; http://ratmap.gen.gu.se), showed no recombination in our backcross with the *Mcs8* peak marker *D14Wox3*. The mapping of a breast cancer susceptibility locus, *Mcs8*, in close proximity to the casein gene is of specific interest. Virgin WKy rats have been shown to overexpress α-, β-, and γ-casein genes in comparison to most rat strains including the WF strain (HSU and GOULD 1993). This precocious differentiation of WKy mammary gland was further documented by BENTON *et al.* (1999). It was observed that the onset of

β-casein overexpression occurs at 6 wk of age, with morphological differentiation of the mammary gland detectable at 7 wk. Furthermore, the induction and maintenance of overexpressed β-casein in virgin WKy rat mammary gland was found to be independent of epidermal growth factor (EGF), which is normally required for *in vitro* lactational differentiation of the rat mammary gland. The potential association of this differentiation phenotype and that of reduced susceptibility to mammary carcinogenesis requires further experimental definition.

The *Mcs* loci that have been identified so far in our laboratory are summarized in Table 4. The *Mcs1*, *-3*, and *-4* loci identified in the Cop rat (SHEPEL *et al.* 1998a) did not contribute significantly toward the resistance to mammary carcinoma formation in the WKy rat. Likewise, the *Mcs5*, *-7*, and *-8* loci identified in the WKy rat in this study did not contribute to resistance observed in the Cop rat. However, the QTL peaks of the *Mcs2* locus in the Cop and the *Mcs6* locus in the WKy rat had large overlapping regions on chromosome 7 from *D7Mgh15* to *D7Mgh10*. The presence of either a Cop or WKy allele at these loci acts to increase resistance to mammary carcinogenesis; however, it remains to be determined whether *Mcs2* and *Mcs6* contain the same allele that modulates mammary cancer susceptibility. These data are compatible with a hypothesis suggesting that susceptibility to breast cancer development is influenced by a wide variety of genes and gene interactions.

A statistical model was used to test for interactions between the identified significant main effects (*Mcs5–8*) alone and between these main effects and other loci. A stepwise procedure was employed to identify the statistical model that best described our genome scan data. Optimality was measured by the BIC, an information criterion that provides a balance between goodness of fit of the model to the data and number of model parameters. The methodology implemented here enhances that performed in SHEPEL *et al.* (1998a) in three ways. First, the more flexible negative binomial distribution was used in place of the Poisson phenotype model; second, cases having missing marker data were multiply imputed instead of being omitted during the model selection phase; and third, interactions were assessed not only between QTL that are identified as significant in the initial genome scan, but also between these QTL and other genome regions.

TABLE 3

Average tumor number within marker genotype class

Markers	WF/WF, WF/WF	WF/WF, WF/WKy	WF/WKy, WF/WF	WF/WKy, WF/WKy
D6Mit2, D14Wox3	5.54	1.93	6.73	3.92
D5Rat22, D10Rat12	5.98	7.68	1.21	3.85

The genotypes for each marker pair are shown in the column headings.

TABLE 4

Summary of rat mammary carcinoma susceptibility loci

Locus	Chromosome	Allele, effect on tumor no.	Reference
Mcs1	2	Cop, decrease	SHEPEL *et al.* (1998a)
Mcs2	7	Cop, decrease	SHEPEL *et al.* (1998a)
Mcs3	1	Cop, decrease	SHEPEL *et al.* (1998a)
Mcs4	8	Cop, increase	SHEPEL *et al.* (1998a)
Mcs5	5	WKy, decrease	This article
Mcs6	7	WKy, decrease	This article
Mcs7	10	WKy, increase	This article
Mcs8	14	WKy, decrease	This article
Mcsm1	6	WKy, increase	This article

To test for interactions between the identified significant main effects (*Mcs5–8*) alone and between these main effects and other loci, a stepwise procedure was used to identify the statistical model that best described our genome scan data. A search over all possible models was not computationally feasible. However, we have shown here for the first time that the set of loci over which to search can be reduced to a manageable size by using information from a standard mapping analysis assuming a one-gene model. In particular, we have shown that if the standard mapping method (assuming a one-gene model) is applied to data in which there are two genes (Q1, which has a significant main effect, and Q2, with no main effect but a first-order interaction with Q1), then the LOD profile will be highest at the marker nearest the interacting gene (see APPENDIX). Thus, it is not necessary to test each marker ($n = 224$), but only those markers that are nearest the highest point of the LOD profile on each chromosome ($n = 17$).

The optimal model, as measured by the BIC, included the four main effects and an interaction between *Mcs8* and *Mcsm1* (on chromosome 6). *Mcsm1* had no significant main effect in this genetic analysis, but resulted in a decreased ability of *Mcs8* to reduce susceptibility to carcinogenesis. We also found that a model that included an interaction between *D5Rat22* (*Mcs5*) and *D10Rat12* (*Mcs7*), although not optimal in terms of the lowest BIC, had a BIC lower than that obtained with only main effects for *Mcs5–8*. These findings further suggest a high degree of genetic complexity underlying susceptibility to mammary carcinogenesis. It is likely that these findings of genetic complexity in rats will be also be found to exist for breast cancer in human populations.

On the basis of our findings that the genetics of mammary cancer susceptibility are complex, we suggest a model for inherited risk for breast cancer in families that do not carry strongly penetrant susceptibility alleles such as *BRCA1* and *BRCA2*. In this model, populations carry a relatively large number of genetic loci with alleles that act either to increase or decrease the risk to breast cancer. The relative risk of an individual woman to breast cancer development is in part governed by the total effect of her inherited alleles at such loci, as well as inherited modifiers of these alleles and the environmental factors to which these individuals are exposed. It is likely that women with first degree relatives with breast cancer have a higher ratio of susceptibility to resistance alleles. Conversely, we hypothesize that other families have increased resistance to breast cancer and carry an inverse ratio of these alleles. Identifying genes within these alleles by first identifying their rat homologues will allow this hypothesis to be tested. In addition, once identified, these genes may serve as drug discovery targets for novel breast cancer prevention compounds.

LITERATURE CITED

BENTON, M. E., K. S. CHEN, J. D. HAAG, C. A. SATTLER and M. N. GOULD, 1999 Precocious differentiation of the virgin Wistar-Kyoto rat mammary gland. Endocrinology **140:** 2659–2671.

CANZIAN, F., T. USHIJIMA, M. TOYOTA, T. SUGIMURA and M. NAGAO, 1996 Linkage mapping of the rat *Tp53* gene on chromosome 10. Mamm. Genome **7:** 630.

CHEN, K. S., L. A. SHEPEL, J. D. HAAG, G. M. HEIL and M. N. GOULD, 1996 Cloning, genetic mapping and expression studies of the rat *Brca1* gene. Carcinogenesis **17:** 1561–1566.

CHURCHILL, G. A., and R. W. DOERGE, 1994 Empirical threshold values for quantitative trait mapping. Genetics **138:** 963–971.

DRINKWATER, N. R., 1997 *Qlink Documentation.* MacArdle Laboratory for Cancer Research, University of Wisconsin, Madison, WI.

DRINKWATER, N. R., and J. H. KLOTZ, 1981 Statistical methods for the analysis of tumor multiplicity data. Cancer Res. **41:** 113–119.

GAUGUIER, D., P. J. KAISAKI, M. ROUARD, R. H. WALLIS, J. BROWNE *et al.*, 1999 A gene map of the rat derived from linkage analysis and related regions in the mouse and human genomes. Mamm. Genome **10:** 675–686.

GEORGE, S., A. J. CLARK and A. L. ARCHIBALD, 1997 Physical mapping of the murine casein locus reveals the gene order as alpha-beta-gamma-epsilon-kappa. DNA Cell Biol. **16:** 477–484.

HAAG, J. D., M. A. NEWTON and M. N. GOULD, 1992 Mammary carcinoma suppressor and susceptibility genes in the Wistar-Kyoto rat. Carcinogenesis **13:** 1933–1935.

Hsu, L.-C., and M. N. Gould, 1993   Cloning and characterization of overexpressed genes in the mammary gland of rat strains carrying the mammary carcinoma suppressor (*Mcs*) gene. Cancer Res. **53:** 5766–5774.

Kass, R. E., and A. E. Raftery, 1995   Bayes factors. J. Am. Stat. Assoc. **90:** 773–795.

Kosambi, D. D., 1944   The estimation of map distances from recombination values. Ann. Eugen. **12:** 172–175.

Kruglyak, L., and E. S. Lander, 1995   A nonparametric approach for mapping quantitative trait loci. Genetics **139:** 1421–1428.

Lan, H., L. A. Shepel, J. D. Haag and M. N. Gould, 1999   Linkage mapping of rat chromosome 5 markers generated from chromosome-specific libraries. Mamm. Genome **10:** 687–691.

Lander, E. S., and D. Botstein, 1989   Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics **121:** 185–199.

Lander, E. S., P. Green, J. Abrahamson, A. Barlow, M. J. Daly *et al.*, 1987   MAPMAKER: an interactive computer program package for constructing primary genetic linkage maps of experimental and natural populations. Genomics **1:** 174–181.

MathSoft, 1997   *S-PLUS 4: Guide to Statistics.* Data Analysis Product Division, Seattle, WA (http://www.splus.mathsoft.com).

McCullagh, P., and J. A. Nelder, 1989   *Generalized Linear Models.* Chapman & Hall, New York.

Mood, A. M., F. A. Graybill and D. C. Boes, 1974   *Introduction to the Theory of Statistics*, Ed. 3. McGraw-Hill, New York.

Poole, T. M., and N. R. Drinkwater, 1996   Two genes abrogate the inhibition of murine hepatocarcinogenesis by ovarian hormones. Proc. Natl. Acad. Sci. USA **93:** 5848–5853.

Schwarz, G., 1978   Estimating the dimension of a model. Ann. Stat. **6:** 461–464.

Shepel, L. A., and M. N. Gould, 1999   The genetic components of susceptibility to breast cancer in the rat. Prog. Exp. Tumor Res. **35:** 158–169.

Shepel, L. A., H. Lan, J. D. Haag, G. M. Brasic, M. E. Gheen *et al.*, 1998a   Genetic identification of multiple loci that control breast cancer susceptibility in the rat. Genetics **149:** 289–299.

Shepel, L. A., H. Lan, G. M. Brasic, M. E. Gheen, L.-C. Hsu *et al.*, 1998b   Mapping of 55 new rat microsatellite markers from chromosome-specific libraries. Mamm. Genome **9:** 622–628.

Warner, E., W. Foukes, P. Goodwin, W. Meschino, J. Blondal *et al.*, 1999   Prevalence and penetrance of *BRCA1* and *BRCA2* mutations in unselected Ashkenazi Jewish women with breast cancer. J. Natl. Cancer Inst. **91:** 1241–1247.

Communicating editor: N. A. Jenkins

APPENDIX

Identifying gene interactions is important in the analysis of congenic line data as well as when attempting to address questions of gene function. We wanted to test for first-order interactions between the significant QTL and other QTL that may not have been identified as significant in the original genome scan. To consider all possible pairs of interactions in a data set of this size (228 markers) was not feasible. However, we identified a way in which we could use the output from a standard mapping analysis to provide information about which QTL to test. In particular, we have shown below that if the standard mapping method (assuming a one-gene model) is applied to data in which there are two genes (Q1, which has a significant main effect, and Q2, with no main effect but a first-order interaction with Q1), then the LOD profile will be highest at the marker nearest the interacting gene (Q2). Given this information, one marker on each chromosome could be chosen

to test for significant interactions between the QTL with significant main effects and other genome regions.

Consider parental strains differing for a quantitative trait of interest, denoted P1 and P2 with genotypes qq and QQ respectively. Suppose that a backcross is performed with P1 as the recurrent parent. Let $\{y_i : i = 1, 2, \ldots, n\}$ denote the n offspring. For fixed individual i, assume that y_i is affected by a single gene G with genotype represented by g_i, where g_i is an indicator variable equal to the number of Q alleles. A standard model for y_i is

$$y_i = \alpha_0 + \alpha_1 g_i + \varepsilon_i, \qquad (A1)$$

where $\varepsilon_i \sim N(0, \sigma^2)$. At a marker M_j on chromosome j, where the marker genotype of individual i, $z_{i,j}$, can be measured, it follows that

$$g_{i,j} = \begin{cases} 0 & \text{w.p. } 1 - r_j \text{ if } z_{i,j} = 0 \\ 1 & \text{w.p. } r_j \text{ if } z_{i,j} = 0 \\ 0 & \text{w.p. } r_j \text{ if } z_{i,j} = 1 \\ 1 & \text{w.p. } 1 - r_j \text{ if } z_{i,j} = 1. \end{cases}$$

r_j is the recombination frequency between M_j and G. The generalized likelihood-ratio test statistic for testing $\alpha_1 = 0$ against the alternative that $\alpha_1 \neq 0$ is a simple monotone function of the LOD score,

$$\frac{-2 \ln(L(\hat{\hat{\alpha}}_0, 0, \hat{\tilde{\sigma}}^2)/L(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\sigma}^2))}{2 \ln 10},$$

where $L(\cdot)$ denotes the likelihood function for model (A1), $(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\sigma}^2)$ are unconstrained maximum-likelihood estimates (MLEs), and $(\hat{\hat{\alpha}}_0, 0, \hat{\tilde{\sigma}}^2)$ are MLEs under the assumption that $\alpha_1 = 0$. For the test at any marker M_j, it can be shown (Mood *et al.* 1974, p. 435) that

$$\text{LOD} = \frac{1}{2 \ln 10} \left\{ n \ln\left[1 + \frac{(n_0 n_1 / n^2)(\bar{y}_0 - \bar{y}_1)^2}{(1/n)(n_0 s_0^2 + n_1 s_1^2)} \right] \right\},$$

$$(A2)$$

where, for $k = 0, 1$, n_k represents the number of animals for which $z_{i,j} = k$, $\bar{y}_0 = (1/n_0)\sum_{i=1}^{n} y_i(1 - z_{i,j})$, and $\bar{y}_1 = (1/n_1)\sum_{i=1}^{n} y_i z_{i,j}$; s_0^2 and s_1^2 denote the within-group sample variances. Replacing $(1/n)(n_0 s_0^2 + n_1 s_1^2)$ by its limit in probability, expanding the logarithm, and invoking the central limit theorem to approximate the distribution of $n(\bar{y}_0 - \bar{y}_1)^2$ gives $E[\text{LOD}] \approx (1 + n\gamma)/2 \ln 10$, where $\gamma = (E[y_i|z_{i,j} = 1] - E[y_i|z_{i,j} = 0])^2/4\tilde{\sigma}^2$ and $\tilde{\sigma}^2 = \frac{1}{2}(\text{var}[y|z_{i,j} = 0] + \text{var}[y|z_{i,j} = 1])$. Suppose the phenotype y_i is determined by two genes *G1* and *G2* on distinct chromosomes j and k. Let

$$y_i = \alpha_0 + \alpha_1 g_{i,j} + \delta_1 g_{i,j} g_{i,k} + \varepsilon_i, \qquad (A3)$$

where $g_{i,j}$ and $g_{i,k}$ are determined as above by marker genotypes $z_{i,j}$ and $z_{i,k}$ of markers M_j and M_k, respectively; r_j and r_k represent the recombination frequencies between M_j and *G1* and M_k and *G2*, respectively. The test statistic (A2), derived under the assumption of the one-

gene model (A1), is evaluated at M_k as if the two-gene model (A3) governs the data. This gives $E[\text{LOD}] = (1 + n\tilde{\gamma})/2 \ln 10$, where

$$\tilde{\gamma} = \frac{(E[y_i|z_{i,k} = 1] - E[y_i|z_{i,k} = 0])^2}{4\{\frac{1}{2}(\text{var}[y|z_{i,k} = 0] + \text{var}[y|z_{i,k} = 1])\}}$$

$$= \frac{[(\delta_1/2)(1 - 2r_k)]^2}{2[((\alpha_1 + \delta_1)^2 + \alpha_1^2 + 2\delta_1^2 r_k(1 - r_k))/4 + 2\sigma^2]}.$$

Since $0 \le r_k \le 0.5$, for a fixed $\delta_1 \ne 0$, $E[\text{LOD}]$ is maximized when $r_k = 0$.