Inheritance of human breast cancer: Evidence for autosomal dominant transmission in high-risk families

(breast neoplasms/epidemiology/gene mapping/medical genetics/pedigree analysis)

BETH NEWMAN, MELISSA A. AUSTIN, MING LEE, AND MARY-CLaire KING

School of Public Health, University of California, Berkeley, CA 94720

Communicated by L. L. Cavalli-Sforza, December 29, 1987

ABSTRACT Segregation analysis of breast cancer in families can provide the logical basis and the specific genetic models for mapping and identifying genes responsible for human breast cancer. Patterns of breast cancer occurrence in families were investigated by complex segregation analysis. In a sample of 1579 nuclear families ascertained through a population-based series of probands, an autosomal dominant model with a highly penetrant susceptibility allele fully explained disease clustering. From the maximum-likelihood Mendelian model, the frequency of the susceptibility allele was 0.0006 in the general population, and lifetime risk of breast cancer was 0.82 among susceptible women and 0.08 among women without the susceptibility allele. Inherited susceptibility affected only 4% of families in the sample: multiple cases of this relatively common disease occurred in other families by chance. The same genetic models, with higher gene frequency, explained disease clustering in an extended kindred at high risk of breast cancer. Evidence for a highly penetrant, autosomal dominant susceptibility allele for breast cancer in a high-risk family and the general population suggests that high-risk families can serve as models for understanding breast cancer in the population as a whole.

Molecular approaches to human gene mapping have revolutionized the study of human disease. It is now possible to locate and characterize genes responsible for inherited susceptibility to disease as well as genes altered by somatic mutation in target tissues. Among purely genetic disorders, Huntington disease (1), cystic fibrosis (2-4), and Duchenne muscular dystrophy (5) have all been mapped, and isolation of the disease genes is underway. For cardiovascular disease, gene mapping strategies have helped enormously in the elucidation of disease etiology, genetic heterogeneity, and interaction of genetic susceptibility with environmental factors (6, 7).

Genetic analysis is equally important to the understanding of neoplastic diseases. Cancer is genetic, in the sense that tumor development requires the alteration of DNA sequences. Susceptibility to specific cancers may also be inherited in some families. Genetic analysis of cancer susceptibility in such families has led to the mapping of genes implicated in retinoblastoma (8), Wilms tumor (9), multiple endocrine neoplasia (10, 11), Hodgkin disease (12), and at least some forms of colon cancer (13, 14).

As the most common cancer among American women, breast cancer has become the subject of molecular gene mapping studies (15), motivated also by the significance of family history as a risk factor for breast cancer (16). Particularly dramatic increases in risk are associated with a family history of breast cancer in a young sister or mother or with a family history of bilateral disease, whereas multiple cases of postmenopausal breast cancer in families often occur due to chance (17). Among women with a severe family history of breast cancer, empiric risks of breast cancer by age 65 approach 50% (18).

Clustering of breast cancer in high-risk families is consistent with shared environmental effects, culturally transmitted risk factors, polygenic effects, the influence of individual major genes, or a combination of these (19). The present genetic analysis of 1579 families was undertaken to test whether individual genes with high penetrance influence breast cancer risk and, if so, the nature and extent of that influence. Genetic analysis of an extended, high-risk kindred from this sample was then undertaken to determine whether the pattern of breast cancer occurrence in this family was consistent with major gene effects. If a major gene strongly influencing breast cancer susceptibility is segregating in some families, then such high-risk families might serve as models for gene mapping studies.

MATERIALS AND METHODS

Families. Segregation analysis was undertaken for 1579 nuclear families of breast cancer probands diagnosed before age 55. The probands were consecutive, Caucasian women diagnosed with microscopically confirmed primary breast cancer between December 1, 1980, and December 31, 1982, in the San Francisco Bay Area and metropolitan Detroit regions of the National Cancer Institute's Surveillance, Epidemiology, and End Results Program (20). Cases were not selected for family history of disease. Within 6 months of diagnosis, cases were interviewed in their homes regarding family history of breast and other cancers. Details of the questionnaire and epidemiologic characteristics of the sample have been published elsewhere (21). Probandes were asked to recall cancer history among their grandmothers, aunts, mothers, and sisters, and any history of breast cancer in male relatives. Subsequent review of the family histories indicated that information was complete and reliable for mothers and sisters but was inconsistent for more distant relatives (18, 22); consequently, only nuclear families were included in segregation analysis of the entire sample of families. In these analyses, we defined affected as any breast cancer in a mother or a sister, regardless of age at diagnosis. Analyses were also carried out with only the families of the youngest probands (diagnosed before age 40). Sample size was reduced to 326 families, but parameter estimates for each model did not differ from those for the entire sample reported here.

From among those probands whose interviews reflected multiple cases of breast cancer in their families, we selected for further follow-up a 36-year-old case who reported a sister, paternal aunt, and paternal uncle with breast cancer. Interviews with 77 members of her extended kindred and review of medical records and death certificates revealed 14 confirmed cases of breast cancer, 11 in women and 3 in men.

Abbreviation: RR, relative risk.
among 252 adult relatives in five generations (Fig. 1). Cancer at sites other than the breast occurred in 11 of the 179 biological relatives without breast cancer, not in excess of the expected incidence for a sample of this age and sex distribution over the same time period. Of the 59 persons marrying into the family, one developed breast cancer and four developed other cancers, also within expectation. For segregation analysis of this family, only breast cancer cases were considered affected.

Segregation Analysis. Patterns of breast cancer occurrence in the pedigrees were investigated by using complex segregation analysis, with the goal of resolving influences of major genes, multifactorial (polygenic and cultural) heritability, and, in the extended kindred, cohort effects (23). All analyses were carried out by using the computer program POINTER, which calculates maximum-likelihood parameter estimates for segregation of a trait in the presence of major genes, environmental effects, polygenic heritability, and age- and sex-dependent onset (23, 24). The parameters used to define alternate models are described in Table 1. POINTER analyzes nuclear families, which may be either unrelated to one another or may be components of an extended kindred. For the present analysis, there were 1579 unrelated families from the population-based series and one extended kindred in which 58 nuclear families were linked by specifying the relationship of the proband to her closest relative in each component family.

To estimate meaningful parameter values, it was necessary to adjust for the mode of ascertainment of the families. In the population-based series, each family contained a single proband. We replicated analyses with ascertainment probabilities ranging from \( p = 0.01 \) (given the narrow time period for selection of probands) to \( p = 0.27 \) (the proportion of cases in the families who were probands). Results of parameter estimates and comparisons of hypotheses were virtually unchanged over this range of \( p \) values for the series of nuclear families and for the extended kindred. Values corresponding to \( p = 0.01 \) are reported here.

Parameter estimates also depended on disease risk in the general population. We estimated disease risk from population-based, age-specific, cumulative incidence rates for breast cancer for the geographic regions from which the families were drawn (20). Given the age distribution of subjects in the families, we defined four liability classes for the analyses: the probability of developing breast cancer was estimated to be 0.0010 for females by age 15 or at any age for males; 0.0045 for females by age 40; 0.0283 for females by age 55; and 0.0819 for females over the entire lifetime. The liability class for each woman in the analysis was defined by her age at diagnosis of breast cancer, at death if deceased without breast cancer, or at present if living and unaffected. All men were assigned to the lowest liability class.

Results

Population-Based Series of Families. A wide range of models was examined, postulating genetic transmission by various modes of inheritance, familial clustering without the influence of a major gene, both of the above, or no familial clustering beyond that occurring by chance for this common disease. The parameter estimates corresponding to the maximum-likelihood models under each set of constraints are presented in Table 1 for the population-based series of nuclear families. In comparison to the unrestricted model (model 1), both models omitting a single gene effect could be excluded. Specifically, the hypothesis that breast cancer susceptibility was not transmitted in these families (model 2) was rejected \( (P < 0.0001) \). Similarly, the hypothesis of familial clustering due to multifactorial inheritance without the influence of a major gene (model 3) was rejected \( (P < 0.0001) \).

The maximum-likelihood Mendelian, or single-locus, model was determined by fixing multifactorial heritability at zero and the probabilities that a susceptibility allele would be transmitted by homozygous susceptible, heterozygous, and homozygous normal parents at 1.0, 0.5, and 0.0, respectively (model 4). The log likelihoods of the general single-locus model and the unrestricted model were almost identical. To determine the mode of inheritance, additional models were specified and tested against the general single-locus model and the unrestricted model.

The hypothesis of a fully dominant susceptibility allele (model 5) was consistent with the distribution of breast cancer in these families. The likelihood of model 5 was virtually identical to the likelihoods of models 4 and 1. The equivalence of models 5 and 4, which differed primarily in dominance values, was due to the low gene frequency of the susceptibility allele \( (q = 0.0006) \); because the vast majority of susceptible individuals would be expected to be heterozygotes, dominant and codominant alleles would be indistinguishable. To evaluate the possibility that multifactorial heritability might be significant in the presence of a major gene, we attempted to estimate heritability given the Mendelian restrictions of models 4 and 5. In each case, multifactorial heritability reached a zero bound from various initial values.

The hypothesis of recessive inheritance of susceptibility could be rejected. A recessive model was first defined by setting dominance at zero and iterating on gene frequency and difference in liability. The results were not biologically meaningful: the allele frequency was so low that no susceptible individuals would be expected, and the difference in liability was so large that any susceptible individuals who did occur would have had breast cancer by age 15. Consequently, we further restricted the difference in liability to 2 standard deviations (its initial value for all analyses) and estimated the corresponding frequency for a recessive allele (model 6). This

Fig. 1. An extended family at high risk of breast cancer. Affected women and men are represented by filled circles and squares, respectively, with the age at diagnosis of breast cancer indicated for each case. The maximum-likelihood autosomal dominant (or codominant) model for this family predicts lifetime risks of breast cancer of 86% for genetically susceptible women, 9% for genetically susceptible men, and 6% for women without genetic susceptibility.
The hypothesis that multifactorial heritability was equivalent across generations could not be rejected. The hypothesis that the pattern of breast cancer in this family could be explained without the influence of a major gene was rejected by comparing model 4 to model 1 (P < 0.05). The maximum-likelihood dominant single-locus model (model 5) provided an adequate fit to the data and was virtually identical to the dominant single-locus model (model 6). Analyses including only a recessive single locus without heritability did not converge, but the less constrained recessive model, which also incorporated multifactorial heritability and a cohort effect (model 7), was significantly less likely than the unrestricted model (P < 0.02). Finally, the dominant single-locus model from the population-based series of families was fit to this extended pedigree, allowing the frequency of the susceptibility allele to vary (model 8). The hypothesis that the same model fit the population-based series of nuclear families and the extended family could not be rejected. Based on the maximum-likelihood single-locus model (model 5) from Table 2, the lifetime penetrance of the susceptibility allele in this family was 0.86 for females and 0.09 for males, whereas women in this family without the susceptibility allele had a lifetime disease risk of 0.06, as in the general population.

DISCUSSION

Complex segregation analysis of this population-based series of families indicates that an autosomal dominant allele with high penetrance could fully explain clustering of breast cancer in families but that genetic susceptibility was present in only 4% of the families. More than 12% of the probands in this sample have affected mothers or sisters: this disease

Table 1. Maximum-likelihood estimates and tests of hypotheses for complex segregation analyses of 1579 nuclear families of breast cancer patients diagnosed before age 55

<table>
<thead>
<tr>
<th>Model</th>
<th>Hypothesis</th>
<th>$H$</th>
<th>$T_2$</th>
<th>$d$</th>
<th>$t$</th>
<th>$q$</th>
<th>$-2 \ln L + C$</th>
<th>Comparison</th>
<th>$\chi^2$ (df)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Unrestricted model</td>
<td>0.00</td>
<td>0.70</td>
<td>0.72</td>
<td>2.74</td>
<td>0.0006</td>
<td>-11692.9</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>No inheritance of susceptibility</td>
<td>(0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-11515.7</td>
<td>2 vs. 1</td>
<td>177.2 (2)</td>
</tr>
<tr>
<td>3</td>
<td>Multifactorial heritability without major locus</td>
<td>0.26</td>
<td>(0.5)</td>
<td>0.77</td>
<td>3.01</td>
<td>0.0006</td>
<td>-11660.3</td>
<td>3 vs. 1</td>
<td>32.6 (4)</td>
</tr>
<tr>
<td>4</td>
<td>General single locus</td>
<td>(0)</td>
<td>(0.5)</td>
<td>1.0</td>
<td>2.32</td>
<td>0.0006</td>
<td>-11691.9</td>
<td>4 vs. 1</td>
<td>1.0 (2)</td>
</tr>
<tr>
<td>5</td>
<td>Dominant single locus</td>
<td>(0)</td>
<td>(0.5)</td>
<td>(1)</td>
<td>2.32</td>
<td>0.0006</td>
<td>-11691.9</td>
<td>5 vs. 4</td>
<td>0.0 (1)</td>
</tr>
<tr>
<td>6</td>
<td>Recessive single locus</td>
<td>(0)</td>
<td>(0.5)</td>
<td>(0)</td>
<td>(2)</td>
<td>0.065</td>
<td>-11683.7</td>
<td>6 vs. 4</td>
<td>8.2 (2)</td>
</tr>
</tbody>
</table>

$H$, heritability of susceptibility due to both polygenic and cultural effects; $q$, gene frequency of the hypothetical susceptibility allele ($q = 0$ if no major gene influences susceptibility); $d$, dominance of the susceptibility allele ($d = 1$ if susceptibility is dominant; $d = 0$ if susceptibility is recessive); $t$, difference in liability, measured in standard deviations, between homozygotes (larger $t$ values reflect higher penetrance of the disease allele); $T_2$, probability that a heterozygous parent transmits susceptibility; $L$, likelihood of the data under this model; $C$, a constant. Each model was defined by fixing the appropriate parameters (values in parentheses) and then iterating on the remaining parameters to obtain maximum-likelihood estimates. Hypotheses were tested by contrasting each model to a less restricted, alternative model by calculating twice the difference of their log likelihoods.

Table 2. Maximum-likelihood estimates and tests of hypotheses for 252 members of an extended family at high risk of breast cancer

<table>
<thead>
<tr>
<th>Model</th>
<th>Hypothesis</th>
<th>$H$</th>
<th>$Z$</th>
<th>$T_2$</th>
<th>$d$</th>
<th>$t$</th>
<th>$q$</th>
<th>$-2 \ln L + C$</th>
<th>Comparison</th>
<th>$\chi^2$ (df)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Unrestricted model</td>
<td>0.04</td>
<td>2.39</td>
<td>0.52</td>
<td>1.0</td>
<td>2.38</td>
<td>0.11</td>
<td>70.69</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>No transmission of susceptibility</td>
<td>(0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>112.29</td>
<td>2 vs. 1</td>
<td>41.60 (2)</td>
</tr>
<tr>
<td>3</td>
<td>No cohort effect, otherwise unrestricted</td>
<td>0.21</td>
<td>(1)</td>
<td>0.43</td>
<td>1.0</td>
<td>2.52</td>
<td>0.08</td>
<td>71.81</td>
<td>3 vs. 1</td>
<td>1.12 (1)</td>
</tr>
<tr>
<td>4</td>
<td>Multifactorial heritability without major locus</td>
<td>0.83</td>
<td>1.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>80.61</td>
<td>4 vs. 1</td>
<td>9.92 (4)</td>
</tr>
<tr>
<td>5</td>
<td>Dominant single locus</td>
<td>(0)</td>
<td>(1)</td>
<td>(0.5)</td>
<td>(1.0)</td>
<td>3.92</td>
<td>0.003</td>
<td>73.78</td>
<td>5 vs. 1</td>
<td>3.09 (4)</td>
</tr>
<tr>
<td>6</td>
<td>Codominant single locus</td>
<td>(0)</td>
<td>(1)</td>
<td>(0.5)</td>
<td>(0.5)</td>
<td>7.83</td>
<td>0.003</td>
<td>73.94</td>
<td>6 vs. 1</td>
<td>3.25 (4)</td>
</tr>
<tr>
<td>7</td>
<td>Recessive with heritability</td>
<td>0.22</td>
<td>4.34</td>
<td>0.54</td>
<td>(0)</td>
<td>2.63</td>
<td>0.08</td>
<td>76.91</td>
<td>7 vs. 1</td>
<td>6.22 (1)</td>
</tr>
<tr>
<td>8</td>
<td>Dominant single-locus model from Table 1</td>
<td>(0)</td>
<td>(1)</td>
<td>(0.5)</td>
<td>(1.0)</td>
<td>(2.32)</td>
<td>0.0042</td>
<td>76.08</td>
<td>8 vs. 1</td>
<td>5.39 (5)</td>
</tr>
</tbody>
</table>

See Fig. 1. Parameters are those of Table 1, with the addition of the cohort effect $Z$, the ratio of heritability in the parental generation to heritability in the children's generation ($Z = 1$ if there is no generation effect).
occurs more than once in some families by chance. The suggestion that autosomal dominant transmission could fully explain inheritance of breast cancer susceptibility in high-risk families is significant, not because such inherited susceptibility is common, but because high-risk families can serve as models for breast cancer as a whole.

In this sense, the etiology of human breast cancer may be analogous to retinoblastoma and Wilms tumor. Each of these tumors appears to be due to loss of gene function from both chromosomes of the cells of the target tissue (25–28). In the inherited form of each disease, one altered gene is inherited by all cells of a child from a parent, and the second alteration, at the same locus, occurs somatically only in the target tissue of the affected child (29). Thus, familial retinoblastoma and Wilms tumor are inherited as dominant diseases in families and are recessive at a single locus at the level of the malignant cell. Breast tumorigenesis might also be due to two alterations at a single locus. Alternatively, breast tumorigenesis might be due to alterations at two (or more) different loci. According to this model, rare families in which cancer susceptibility was inherited would carry a germ-line alteration at one locus. A second alteration would occur somatically at an unlinked locus only in the target tissue of the affected person. Most importantly, whether one or more loci were involved, alterations responsible for rare inherited cancer and the more common, entirely somatic cancer could be identified as dominantly inherited susceptibility genes in high-risk families. Segregation analysis of breast cancer in families is intended to provide the logical basis and the specific genetic models for mapping and identifying these genes by linkage analysis.

The present study applied complex segregation analysis of breast cancer to a sufficiently large series of families that statistical power was adequate to test competing hypotheses. Interpretation of these results must be made in light of the limitations of this approach. An underlying assumption of complex segregation analysis is that multifactorial heritability is adequately modeled by a liability that has a multivariate normal distribution. It is assumed that polygenic and cultural traits in spouses are uncorrelated and that influences of mothers and fathers on their children do not differ. Extension of complex segregation analysis to incorporate environmental and cultural traits in assessing multifactorial heritability is necessary (30).

The known ascertainment scheme for the families and the fact that the probands were a consecutive, population-based series of cases enabled results to be generalized to the Caucasian American population as a whole. These results are specific to breast cancer. We did not include cancer at other sites in the analysis because other cancers did not occur with increased frequency. Compared to control families (18, 21), mothers and sisters of the breast cancer cases were at significantly increased risk of breast cancer [relative risk (RR) = 2.0 for mothers and RR = 2.2 for sisters]. RR values for mothers and sisters of cases vs. mothers and sisters of controls were not significant for ovarian cancer (RR = 1.8), endometrial cancer (RR = 1.0), cervical cancer (RR = 1.1), colorectal cancer (RR = 1.1), or lung cancer (RR = 0.8). All histologic types of breast cancer were pooled in the analysis. Nearly 90% of the probands’ tumors were adenocarcinoma. Previous work has indicated that the distributions of histologic types do not differ between breast cancer patients with and without affected mothers or sisters (31).

Segregation analysis of breast cancer has been previously addressed with quite different samples of families (32). Segregation analysis is important because such families selected for high risk suggested that an autosomal dominant gene influenced risk of premenopausal breast and ovarian cancer, another gene (with different penetrance) influenced postmenopausal breast cancer, and a third (combined with environmental and/or polygenic effects) influenced childhood tumors and early-onset breast cancer (33). Analyses of three-generation Danish families ascertained through single probands diagnosed with breast cancer in the 1930s and 1940s (34) also suggested a role for a major autosomal dominant gene or genes for breast cancer (35). Complex segregation analysis of 201 families, ascertained through probands diagnosed with bilateral breast cancer before age 50, suggested that a major gene and heritability (either cultural or polygenic) contributed to familial clustering of disease in this higher-risk sample (36). In our analyses, the influence of polygenic and/or cultural effects is most likely to appear as the cause of incomplete penetrance. Although the lifetime disease risk to susceptible women is high (>0.8), it is not complete. Variation in age-onset and the occurrence of unaffected, but susceptible, women could be due to the action of other genes, environmental influences, or chance. Although the obvious purpose in mapping the major genes suggested by these analyses is to identify the alterations responsible for breast tumorigenesis, the complex etiology of breast cancer also involves understanding how these genes interact with background genotype and cultural and environmental factors that might influence the rate of proliferation of altered cells.

Etiologic heterogeneity poses complexities for the identification of genes responsible for human breast cancer. The disease is likely to be heterogeneous at several levels that affect genetic analysis. First, most breast cancer (96% in this sample) is not influenced by inherited susceptibility, and multiple cases of purely somatic disease may appear in the same family due to common exposure to environmental carcinogens or by chance. The present analysis was designed to address heterogeneity at this level, by determining the proportion of families with inherited susceptibility and the characteristics of susceptibility in those high-risk families. At the second level, the possibility of genetic locus heterogeneity of breast cancer cannot be completely resolved by segregation analyses. The present results suggest that inherited susceptibility to breast cancer may be attributable to the effects of autosomal dominant genes. However, these results are consistent with the existence of only one susceptibility gene or with different loci influencing breast cancer susceptibility in different families. It is not clear which epidemiologic or disease characteristics, if any, would be associated with different susceptibility genes (37). Genetic heterogeneity will probably be fully resolved only with the mapping and identification of susceptibility genes by linkage analyses of extended, high-risk kindreds with sufficient information to test for heterogeneity (38, 39). The final level of possible heterogeneity of breast cancer is allelic heterogeneity: the possibility that more than one allele at the same locus may influence susceptibility. If allelic heterogeneity exists, it would be detectable by comparing the sequences of susceptibility alleles in different families with linkage to the same locus. Direct comparison of such sequences would be possible only after susceptibility genes were identified.

In summary, these analyses reveal the importance of a highly penetrant dominant allele to breast cancer susceptibility in some families. Families with these susceptibility alleles are at high risk of the disease. Such families can serve as models for breast tumorigenesis in the general population. Only after breast cancer susceptibility genes are identified will the full extent of disease heterogeneity and the interaction of environmental and cultural risk factors with genetic susceptibility be understood.

We gratefully acknowledge G. Marie Swanson, William A. Satariano, and Ann Grossbart Schwartz of the Michigan Cancer Foundation and Donald F. Austin of the San Francisco Bay Area Resource for Cancer Epidemiology for assistance with family history data and
Stephanie Sherman for help with POINTER. These analyses were supported in part by National Institutes of Health Grant CA27632.


