# Segregation Analysis of 231 Ashkenazi Jewish Families for Evidence of Additional Breast Cancer Susceptibility Genes[1]

**David J. Kaufman, Terri H. Beaty, and Jeffery P. Struewing[2]**

Laboratory of Population Genetics, National Cancer Institute, NIH, Bethesda, Maryland 20892 [D. J. K., J. P. S.], and Department of Epidemiology, Johns Hopkins University School of Public Health, Baltimore, Maryland [T. H. B.]

## Abstract

**Between 5 and 10% of breast cancer is attributable to inherited cancer susceptibility genes. Mutations in the genes *BRCA1* and *BRCA2* account for two-thirds of hereditary breast cancer cases. Using segregation analysis, families of cases without *BRCA1/2* mutations were studied for statistical evidence of another major breast cancer gene in a community-based sample of Jewish probands tested previously for the presence of three *BRCA* founder mutations. A total of 231 probands with breast cancer, who do not carry a founder mutation, reported complete data on 602 female first-degree relatives of probands over age 20; 78 of these relatives had breast cancer.**

**Segregation analysis was used to evaluate the likelihood of various genetic and nongenetic models. Sporadic, environmental, and general Mendelian genetic models fit the family data poorly and were rejected. A Mendelian recessive model fit better than dominant and codominant models, although none of these could be rejected. Cumulative incidence curves predicted by the recessive and codominant models fit observed incidence among first-degree relatives well. The assumption of Mendelian transmission of a major recessive gene(s) is compatible with the data.**

**The recessive model predicts that 4% of women would carry the high-risk genotype, with 85% of them developing breast cancer by age 70. There was significant heterogeneity between these families and the 114 *BRCA1/2* mutation-positive families from the same study population, implying that this apparent recessive effect is not because of undetected *BRCA1/2* mutations. The study adds support for a major autosomal recessive component to breast cancer susceptibility.**

## Introduction

Breast cancer has the highest incidence and is the second leading cause of cancer mortality among women in the United States (1). If current rates persist, roughly 1 in 10 women in the United States would be diagnosed with breast cancer in their lifetime (2). Family history of breast cancer is a strong risk factor for developing the disease, with >12% of newly diagnosed cases reporting at least one affected first degree relative (3). Only a subset of familial breast cancer is clearly hereditary, owing primarily to mutations in a single, relatively high-penetrance gene (4–6). A sustained effort, encompassing several studies of multiple populations of breast-ovarian families, led to the discovery of two susceptibility genes, *BRCA1* and *BRCA2*, which were originally estimated to be responsible for 80–85% of families (7, 8). Summarizing studies in 17 countries, Szabo and King (9) concluded that roughly one-third of families with ≥3 cases do not carry a *BRCA1* or *BRCA2* mutation. Three United States studies of high-risk families support these findings (10–12); between 19% and 39% of these families were unlinked to either of these two genes. The authors of one study (11) concluded that "at least one more major gene for inherited (breast) cancer remains to be found." Some linkage studies looking for such a gene (13, 14) were based on a rare dominant gene model (15) that was also used as the basis for linkage studies for *BRCA2*. Whereas this model may be applicable to additional breast cancer genes, it may be possible to refine estimates of the penetrance and allele frequency of such genes by performing a new segregation analysis for breast cancer that excludes *BRCA1* and *BRCA2* families. One segregation analysis of families of known *BRCA1* and *BRCA2* noncarriers published recently (16) tested Mendelian and polygenic models, and found a Mendelian recessive gene best described breast cancer patterns among 858 study families with no detected *BRCA1/2* mutation. We performed segregation analysis on families without *BRCA1/2* mutations to look for statistical evidence of an additional major gene(s) that influences age-of-onset of breast cancer. A total of 231 Jewish families were studied that contain at least one affected family member who tested negative for mutations in *BRCA1* and *BRCA2*. We also tested whether age-specific risks, penetrance, and allele frequency of the hypothetical gene(s) appear distinct from *BRCA1* and *BRCA2* mutation carrier families from the same study population. This analysis of observational data cannot prove the existence of another causal genetic factor but can provide statistical evidence to support or reject additional genetic control of breast cancer and serve as the basis for linkage studies to help locate such genes.

| Proband type | Relatives of probands with complete data in Washington Ashkenazi Study | | | |
| --- | --- | --- | --- | --- |
| | Mothers | Sisters | Daughters >age 20 | Half-sisters |
| Breast cancer case without *BRCA1/2* founder mutations (*n* = 231) | | | | |
| Number of relatives | 227 | 173 | 186 | 16 |
| Percentage with breast cancer (*n*) | 21 (48) | 13 (23) | 2 (3) | 25 (4) |
| *BRCA1/2* founder mutation carrier (*n* = 114) | | | | |
| Number of relatives | 114 | 92 | 51 | 1 |
| Percentage with breast cancer (*n*) | 36 (41) | 18 (17) | 2 (1) | 0 (0) |

## Materials and Methods

### Study Population

Families studied were originally recruited in 1996 for the WAS[3] at the National Cancer Institute (Bethesda, MD; Ref. 17) in which 5318 Jewish volunteers from the Washington, DC area provided a family history of breast and other cancers, and a blood sample that was tested for three *BRCA1* and *BRCA2* founder mutations (185delAG and 5382insC in *BRCA1* and 6174delT in *BRCA2*), which account for ~80% of mutations in the Ashkenazim (18). Neither personal nor family history of breast cancer was used as criteria for entry into the study.

Only the 5108 (96%) volunteers who gave consent for future use of their information at the initial ascertainment were eligible for this study. After using personal information to determine that members of 4873 distinct families had participated, all of the personal identifiers were removed from the blood samples and questionnaires rendering them anonymous and preventing identification or recontact of volunteers and their relatives. Volunteers from 255 families had a history of breast cancer but did not carry a founder mutation.

Approximately two-thirds of Jewish breast-ovarian cancer families are predicted to carry a *BRCA1* mutation (19). Additionally, in both the CASH and WAS populations, after excluding *BRCA1/2* mutation carriers, a family history of ovarian cancer does not increase the odds of breast cancer (20, 21). Thus, families of WAS breast cancer cases (*n* = 24) with any reported ovarian cancer were excluded, to reduce the likelihood of including families segregating undetected *BRCA1* mutations. None of the families studied reported a case of male breast cancer.

Segregation analysis was applied to the remaining families of 231 unrelated volunteers with breast cancer who do not carry the three *BRCA1/2* founder mutations. The probands in these families are referred to as noncarriers. Family histories of cancer obtained from volunteers included information on parents, sisters, and daughters over age 20. Of 637 female first-degree relatives, 602 (95%) had complete data available (including current age or age at death, cancer history, type of cancer, and age at diagnosis). These included 78 of 82 relatives who were reported as having had breast cancer (Table 1).

In 42 of the families, more than one person volunteered for the study and provided data that was used to extend, where possible, the family histories used in the analysis. These families also provided an opportunity to estimate the validity of self-reported family histories, by comparing related volunteer reports of shared first-degree relatives.

A formal test of heterogeneity compared the families of

231 noncarriers to 114 eligible families of WAS probands who do carry a *BRCA1/2* founder mutation. Mutation carrier pedigrees were constructed and checked using the method described above for noncarriers, although families with ovarian cases were included to retain as much information on carriers as possible. Of 268 eligible first-degree female carrier relatives over 20 years of age, complete information was available on 258 (96%), including 59 of 60 breast cancer cases (98%; Table 1).

In addition, risk factor data from women who volunteered for WAS, and had neither breast cancer, a founder mutation, nor family history of ovarian cancer (*n* = 3,193) were used for comparison with information from the 231 affected probands.

This study was approved by the Johns Hopkins School of Public Health Institutional Review Board and the Office of Human Subjects Research at the NIH.

### Statistical Methods

**Preliminary Analysis.** Initially all of the female WAS volunteers with no founder mutation and no ovarian cancer in their family (*n* = 3193) were analyzed to examine whether known risk factors for breast cancer were associated with the disease. Age-adjusted multivariate logistic models predicted odds of breast cancer associated with oral contraceptive use, hormone replacement therapy, early age at first birth, parity, age at menarche, and menopausal status. Increased odds of disease associated with family history were also calculated in noncarriers, adjusting for the same covariates.

**Segregation Analysis.** Patterns of disease distribution within families (dominant Mendelian, recessive Mendelian, and so forth) were modeled using modified logistic regression (22). In addition to Mendelian models, sporadic, environmental, and polygenic models were tested. The models that best fit the study families are compared with a general, unrestricted model using the LRT. Models can include one, two, or three "types" of subjects, represented as type AA, AB, and BB. In the genetic models, these correspond to the genotypes of a hypothetical biallelic gene. Transmission parameters in the model ($\tau_{AA}$, $\tau_{AB}$, and $\tau_{BB}$) represent the probability that a parent of a given type transmits a factor "A" to a child. In Mendelian models these $\tau$s are fixed at values of 1.0, 0.5, and 0.0, respectively. The parameter $\gamma$ (penetrance) estimates the proportion of the population developing disease if they lived indefinitely (23). The parameters $\beta$ and $\alpha$ are used to describe the mean age of onset and its variance.

The 231 noncarrier families were selected because the volunteer (proband) is a breast cancer survivor. To correct for this ascertainment criterion, the ln-likelihood of the models given the observed phenotype and age of onset in the volunteer was subtracted from the ln-likelihood of the model fit to the entire pedigree.

---

[3] The abbreviations used are: WAS, Washington Ashkenazi Study; CASH, Cancer and Steroid Hormone; LRT, likelihood ratio test; *df,* degrees of freedom; p(AA), putative recessive genotype; AIC, Akaike Information Criterion.

*Table 2*  Comparison of Washington Ashkenazi Study volunteers with breast cancer and no founder mutation to noncase volunteers with no founder mutation

Excludes all women with a family history of ovarian cancer.

| Characteristic | Noncases ($n = 2962$) N (%) | Breast cancer cases ($n = 231$) N (%) |
|---|---|---|
| Age (yrs) | | |
| 20–29 | 208 (7) | 0 (0) |
| 30–39 | 427 (14) | 4 (3) |
| 40–49 | 1,006 (34) | 49 (20) |
| 50–59 | 636 (22) | 65 (28) |
| 60–69 | 363 (12) | 47 (20) |
| 70–79 | 249 (9) | 53 (23) |
| 80+ | 67 (2) | 13 (6) |
| Not reported | 6 (0.2) | 0 (0) |
| Mean age | 50 (21–93, SD = 14) | 61 (37–88, SD = 12) |
| Mean age at diagnosis | — | 53 (32–84, SD = 11) |
| Mean survival | — | 8 (0–38, SD = 8) |
| Married | 2,181 (74) | 165 (71) |
| Education | | |
| High school or less | 159 (5) | 17 (7) |
| Some post-HS | 366 (12) | 43 (19) |
| College degree | 855 (29) | 68 (29) |
| Postgraduate degree | 1579 (53) | 103 (45) |
| Not reported | 3 (0.1) | 0 (0) |

*Table 3*  Adjusted effect of family history on odds of breast cancer among women with no founder mutation and no family or personal history of ovarian cancer

Results of multiple logistic regression.

| Risk factor | Odds ratio | $P$ | 95% Confidence Interval |
|---|---|---|---|
| Number of affected first-degree relatives | 1.5 | 0.0009 | 1.1–2.0 |
| Increase of 1 year in age at censoring | 1.01 | 0.038 | 1.001–1.020 |
| Ever used hormone replacement therapy | 0.7 | 0.029 | 0.5–1.0 |
| Menarche <age 13 | 1.3 | 0.068 | 0.9–1.7 |
| Menstruation continued after age 45 | 1.6 | 0.010 | 1.1–2.2 |

Each model was tested independently with the REGTL subroutine of the S.A.G.E. software package (24). A second software tool, REGTLHUNT (25), was used to check whether the best fitting were at global and not local maxima.

Hypothesis testing was performed comparing the -2*ln* (likelihood) and Akaike Information Criterion (AIC) scores of nested and non-nested models, respectively. The best-fitting models were used to calculate genotype-specific cumulative incidence curves (26), which were compared with observed Kaplan-Meier risks of breast cancer in the first-degree relatives of probands.

Three tests of heterogeneity were performed. In these tests, the study families are divided into subgroups, and a separate set of models is fit for each strata. If there are significant differences between the subgroups, the additional parameters used to fit multiple sets of models should improve the fit. Significant improvement of fit is measured using a $\chi^2$ statistic, calculated by subtracting the sum of the $-2\ln$ (likelihoods) of the subgroup models from the $-2\ln$ (likelihood) of the entire-group model, with degrees of freedom equal to the additional number of parameters used to estimate the subgroup models. A significant result ($P < 0.05$) indicates rejection of the hypothesis that the fit of the subgroup models are equivalent to the fit of the model for the entire group. We may not have power to detect heterogeneity. Power estimates for segregation analyses are not readily calculated.

The first test of heterogeneity looked for differences between noncarrier families stratified by the menopausal status of the proband (pre-, peri-, or postmenopausal) at the time of diagnosis. The second test of heterogeneity compared the 231 noncarrier families with the families of the 114 *BRCA1/2* mutation carriers, to determine whether findings in the noncarrier pedigrees were distinct from *BRCA1/2* families in the WAS population. The third test of heterogeneity attempted to test for a cohort effect by stratifying families based on whether they contained any women born before 1910.

To check whether inference of a Mendelian genetic model is correct, we estimated the parameter $\tau_{AB}$ under an otherwise

Mendelian model. If $\tau_{AB}$ approaches the Mendelian value of 0.50, the original genetic model is less likely to be spurious (27).

## Results

**Characteristics of the Study Population.**  The three *BRCA1/2* founder mutations were present in 118 (2.4%) of the eligible volunteers. Half of the volunteers were between the ages of 40 and 60 in 1996, and the majority (70%) were women. Fifteen percent reported having had some type of cancer, breast cancer (5.5%) being the most common type. One in 5 volunteers reported at least one first-degree relative with breast cancer.

Among women with no founder mutation (noncarriers) and no personal or family history of ovarian cancer ($n = 3193$), 231 had survived breast cancer for an average of 8 years (Table 2). The mean age of onset for cases was 53 (SD = 11.1). The 231 case parents, siblings, half-siblings, and children over age 20 were eligible for analysis. Because 73% of the probands had incomplete data for at least one grandmother, grandmothers were excluded. Overlapping histories of 42 families that contained multiple WAS volunteers provided an opportunity to estimate the reliability of noncarrier case reported family history. Among relatives described by multiple volunteers, 57 of 58 breast cancer cases (98%) were also reported by another relative. Of 167 female first-degree relatives reported by multiple volunteers, 74% of the reports of current age or age at death matched exactly, 18% were within 1 year of each other, and 4% were within 5 years of one other. Similarly, 54% of the reported age of breast cancer diagnoses matched exactly, 27% of overlapping reports were within 1 year of each other, and 10% were between 1 and 5 years apart.

WAS probands born after 1925 have significantly higher cumulative risks of breast cancer by age 60 than those born between 1900 and 1925 (11% *versus* 5%; log-rank test $P = 0.0002$). This is probably because older cases are less likely to have survived and volunteered. Examination of first-degree relatives of the 231 volunteers, unselected with respect to survival, suggests a cohort effect may exist. Mothers of probands born after 1910 have significantly higher cumulative breast cancer risk at age 60 than those born before 1910 (12% *versus* 5%; log-rank test $P = 0.04$). For this reason, we attempted to account for a generational (cohort) effect in the segregation analyses.

Using multiple logistic regression to adjust for the nongenetic covariates in Table 3, a positive family history of breast cancer remained a risk factor for disease in noncarrier women. Odds of disease increased 50–60% (95% CI, 1.2–2.1) with each affected first-degree relative a volunteer had. Stratifying noncarriers by these covariates provided no evidence of interaction between any of these variables and the observed effect of

*Table 4*  Segregation analysis of 231 families of noncarrier cases: α-specific models

| Parameter | Non-mendelian | | Mendelian | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Sporadic | Environmental | Arbitrary | Codominant | Dominant | Recessive | General model | Recessive free tAB |
| qA | 1.0[a] | 0.939 | 0.794 | 0.245 | 0.0106 | 0.199 | — | 0.201 |
| τAA | — | 0.939[a] | 1.0[a] | 1.0[a] | 1.0[a] | 1.0[a] | 0.92 | 1.0[a] |
| τAB | — | 0.939[a] | 0.5[a] | 0.5[a] | 0.5[a] | 0.5[a] | 0.09 | 0.51 |
| τBB | — | 0.939[a] | 0.0[a] | 0.0[a] | 0.0[a] | 0.0[a] | 1.0[b] | 0.0[a] |
| βAA | | | | −8.2 | | | −9.0 | |
| βAB | | | | −12.7[a] | | | −14.5 | |
| βBB | | | | −17.2 | | | −11.12 | |
| β men | −50.0[a] | −50.0[a] | −50.0[a] | | −50.0[a] | −50.0[a] | −50.0[a] | −50.0[a] |
| β women | −6.87 | −9.62 | −8.88 | | −8.34 | −8.87 | | −8.87 |
| αAA[c] | 0.0858 | 0.103 | 0.0933 | | 0.1517 | 0.160 | 0.167 | 0.161 |
| αAB[c] | — | 0.183 | 0.0932 | | 0.1517[a] | 0.093[a] | 0.099 | 0.093[a] |
| αBB[c] | — | 0.150 | 0.16 | | 0.0902 | 0.093 | 0.021 | 0.093 |
| α men | 0.02[a] | 0.02[a] | 0.02[a] | 0.02[a] | 0.02[a] | 0.02[a] | 0.02[a] | 0.02[a] |
| α women | | | | 0.137 | | | | |
| γ[c] | 0.54 | 0.885 | 0.91 | 0.78 | 0.58 | 0.94 | 0.49 | 0.92 |
| −2lnL | 983.15 | 970.93 | 971.18 | 970.89 | 972.12 | 968.68 | 958.28 | 968.8 |
| χ² | 28.57 | 12.65 | 12.9 | 12.61 | 13.84 | 10.4 | — | 10.52 |
| AIC[d] | 986.15 | 984.93 | 983.68 | 980.89 | 982.12 | 978.68 | | |
| Degrees of freedom | 3 | 7 | 6 | 5 | 5 | 5 | 12 | 6 |
| p | 0.001 | 0.03 | 0.044 | 0.08 | 0.054 | 0.17 | — | 0.10 |
| Type frequency; mean age of onset | | | | | | | | |
| Type AA | 1; 80 | 0.34; 89 | 0.63; 95 | 0.06; 59 | 0.092; 61 | 0.04; 55 | — | 0.04; 55 |
| Type AB | 0 | 0.49; 100 | 0.32; 95 | 0.37; 91 | | 0.96; 95 | — | 0.96; 95 |
| Type BB | 0 | 0.17; 57 | 0.04; 55 | 0.56; 124 | 0.90; 87 | | — | |

[a] Parameter value fixed in analysis.
[b] Parameter hit bound.
[c] Female-specific.
[d] AIC, Akaike information criterion.

family history. ORs for the nongenetic risk factors were generally similar in magnitude to findings from other large epidemiological studies. As in the CASH (20), having a first- or second-degree relative with ovarian cancer was not related to noncarrier breast cancer risk (OR, 1.2; 95% CI, 0.8–1.9; $P$ = 0.36). The overall comparability of WAS to CASH (20) with respect to familial aggregation outside *BRCA1/2* families suggests that the WAS noncarrier families may be representative of a more general population for the segregation analysis.

**Segregation Analyses.** Patterns of age of onset of breast cancer were initially modeled using genotype-specific $\beta$ parameters. Sex-specific parameters for $\alpha$ and $\gamma$ were fit, fixing male values to indicate a very low penetrance and high mean age of onset among men. The sporadic model was rejected (LRT $\chi^2$ = 17.93; 7 *df*; $P$ = 0.012), but this analysis was unable to discriminate between environmental and genetic models, none of which were rejected. The most parsimonious $\beta$-specific model described a codominant Mendelian gene(s) decreasing the mean age of cancer onset with each risk allele inherited. A polygenic model could not be fit to the data. Limitations concerning the polygenic models are discussed below.

In many of the $\beta$-specific models, estimates of the $\alpha$ parameter were not stable; the same model, fit several times, would converge with equal likelihood at different values of $\alpha$. This suggested that variations in age-of-onset of breast cancer might be better described through differences in the $\alpha$ parameter. A second set of genotype-specific $\alpha$ models was fit. Sex specific $\beta$s and $\gamma$s were estimated, fixing male values of $\beta$ and $\gamma$. The serial analysis of gene expression software does not permit estimation of a codominant model with genotype-specific $\alpha$. Because the codominant Mendelian model was the most parsimonious of the $\beta$ models, it was included for comparison.

This necessitated fitting a general model that included type-specific values for both $\alpha$ and $\beta$.

With the exception of the models for arbitrary and dominant genes, $\alpha$-dependent models fit the observed families slightly better than $\beta$-models (Table 4). The sporadic (LRT $\chi^2$ = 28.57; 9 *df*; $P$ = 0.001) and environmental (LRT $\chi^2$ = 12.65; 5 *df*; $P$ = 0.03) models were both rejected. Although these differences were not dramatic, the analysis did discriminate between Mendelian models. A Mendelian gene with arbitrary effects assigned to each genotype was narrowly rejected (LRT $\chi^2$ = 12.90; 6 *df*; $P$ = 0.044). Mendelian transmission of major codominant ($\beta$), dominant ($\alpha$), and recessive ($\alpha$) genes could not be rejected. Of these, the recessive model had the best fit and was the most parsimonious (AIC = 978.68) among all of the models tested.

Although initial analysis of $\alpha$ models supported the existence of an additional major gene(s), the LRT statistics alone do not favor a single model. Fig. 1 compares the observed Kaplan-Meier incidence of breast cancer in first-degree relatives of probands with weighted averages of type-specific risks predicted by three models. The recessive and codominant model predictions match the observed data at both early and later ages, whereas the dominant model underestimates observed risks from age 45 onwards.

The assumption of Mendelian transmission was tested in the recessive and codominant models by estimating the transmission parameter $\tau_{AB}$ (with $\tau_{AA}$ and $\tau_{BB}$ still fixed at Mendelian values). In the recessive model, $\tau_{AB}$ was estimated at 0.51, quite close to the expected Mendelian value of 0.5. In the codominant model, $\tau_{AB}$ was estimated at 0.75, suggesting that parents with one risk allele were more likely to have children with higher cancer risks than expected under strict Mendelian

*Fig. 1.* Comparison of observed Kaplan-Meier cumulative incidence of breast cancer among first-degree relatives to incidence predicted by recessive, dominant, and co-dominant models. Kaplan-Meier figures are based on the 258 probands' reports of breast cancer among their first-degree female relatives. For each model, cumulative incidence was calculated as weighted average of type-specific penetrances, where the weights are equal to the genotypic frequencies predicted among first-degree relatives, and penetrance for a given age $t$ is calculated by $\gamma[e^{(\beta+t\alpha)}/1+ e^{(\beta+t\alpha)}]$. $\gamma$, $\beta$, and $\alpha$ are parameters from the models: $\gamma$ is lifetime penetrance, $\beta$ defines the mean age of onset, and $\alpha$ determines the slope of the cumulative incidence curve.
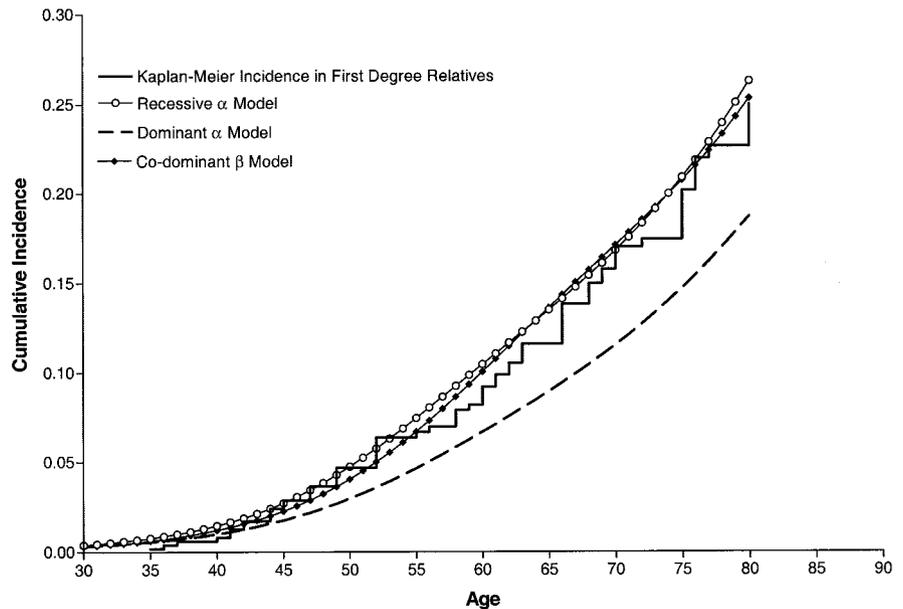
inheritance. When all three of the transmission parameters ($\tau_{AA}$, $\tau_{AB}$, and $\tau_{BB}$) were estimated, values were 1.00, 0.51, and 0.09, respectively, in the recessive model, and 0.75, 0.55, and 0.16 in the codominant model. The close fit of the recessive model to the observed data, and its adherence of estimated $\tau$s to Mendelian transmission values suggest that a major recessive gene is the best explanation of familial aggregation in these families. Additional statistical support for this conclusion comes from the arbitrary $\alpha$ model, which converged at the same values as the recessive model, while estimating an additional parameter.

It should be noted that we attempted to fit polygenic models, but were unable to achieve convergence when modeling only residual familial factors. It was determined that the study families contained insufficient numbers of affected mother-daughter pairs (excluding the probands, $n = 10$) to calculate the residual familial correlation needed to model a polygenic effect.

The recessive model predicts a common mutated allele ($q_A = 0.20$; 95% CI, 0.04–0.36), resulting in 4% of the population being high-risk ("AA") homozygotes (95% CI, 0.2–13%). Fig. 2 compares the predicted cumulative incidence of breast cancer among high-risk homozygotes and those with at least one wild-type allele. By age 50, 28% of high-risk AA carriers would develop breast cancer, compared with 1.3% of women carrying a wild-type allele; at age 70, the risks are 85% and 8%, respectively. Among homozygous carriers who develop breast cancer, the mean age-of-onset would be 55, compared with 72 among those with a wild-type allele. Among homozygous carriers, 70% of breast cancer would occur between the ages of 45 and 65, whereas 75% of cases in low-risk women would occur after age 65. Under this model, 36% of cases diagnosed under age 40, and 16% of those diagnosed by age 80 would carry the high-risk genotype.

A total of 37 first-degree female relatives (6%) in 32 families (14%) had a probability of carrying the p(AA) >0.50. All 37 were breast cancer cases. The mean age of diagnosis in these women was 51, compared with a mean age of onset of 71 in affected relatives with p(AA) ≤ 0.5 ($P = 0.0001$). Probands related to women with p(AA) > 0.50 are similar to probands without a high-risk relative. The two proband groups had the same mean age (61) and similar mean ages of breast cancer onset (probands who are relatives high risk women = 52; probands who are relatives of low risk women = 49; $P = 0.11$).

**Tests of Heterogeneity and Analysis of Birth Year.** In the 114 WAS families where the probands carried BRCA1 or *BRCA2* founder mutation, the dominant model was the most parsimonious. We rejected the hypothesis that the pattern of familial aggregation in *BRCA1/2* carrier families is similar to that in families of noncarrier cases across all of the models ($P \leq 0.00002$). The familial clustering of breast cancer in noncarriers described by the recessive $\alpha$ model appears distinct from the pattern observed in the WAS families containing *BRCA1/2* founder mutation carriers.

Stratifying noncarrier families into those where probands were pre-, peri-, or postmenopausal at the time of diagnosis provided no evidence ($P = 0.18$) of genetic heterogeneity based on the time of the proband breast cancer diagnosis. We also tested for heterogeneity based on the birth years of all of the women in the family. Assuming that a cohort effect would be most pronounced in women who were at highest risk of breast cancer after 1960, families were separated into those with no women born before 1910 (all reaching highest risk after 1960; $n = 83$), and those containing women born both before and after 1910 ($n = 148$), and the models were reanalyzed. In families containing women born before 1910, the recessive model fit the data best, whereas in families containing only recently born women, the environmental model fit best. However, the differences between the two groups were not statistically significant ($P = 0.84$), so this cannot be considered as evidence of heterogeneity.

### Discussion

Our complex segregation analysis supports the existence of a common, high-risk recessive component to breast cancer susceptibility in addition to *BRCA1/2*. This could be because of a single gene or multiple genes acting in a recessive manner. This finding is consistent with the somewhat higher age-specific probability of developing cancer observed in sisters of the
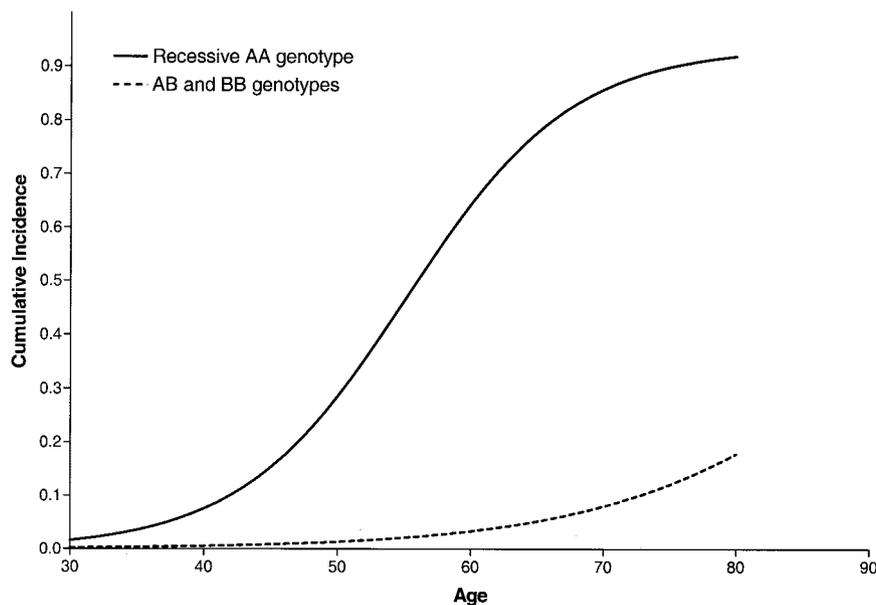
*Fig. 2.* Genotype-specific cumulative incidence of breast cancer predicted under the recessive model. Penetrance for a given age $t$ is calculated by $\gamma [e^{(\beta+t\alpha)}/1 + e^{(\beta+t\alpha)}]$. $\gamma$, $\beta$, and $\alpha$ are parameters from the models: $\gamma$ is lifetime penetrance, $\beta$ defines the mean age of onset, and genotype-specific values of $\alpha$ determine the slope of the cumulative incidence curve.

affected probands compared with their mothers (hazard ratio = 1.5; $P$ = 0.12; 95% CI, 0.9–2.6). The differences in risk between sisters and mothers appear beginning at age 55 and continue through age 80. As a control, sisters of WAS noncarriers without breast cancer were compared with mothers from the same families; the comparable hazards ratio was 1.1 ($P$ = 0.23; 95% CI, 0.9–1.3). Under a recessive model, daughters of the probands should have risks comparable with the mothers of the probands; however, the few cases among daughters in the study population ($n$ = 3) precludes meaningful analysis. However, the apparent increased risk among sisters of probands supports the idea that a recessive gene may be operating in these families.

The cumulative incidence of breast cancer predicted by the recessive and codominant models both matched the observed Kaplan-Meier estimates of risk seen among first-degree relatives. However, when transmission parameters were estimated, only the recessive model closely adhered to Mendelian transmission. Four percent of women would carry the high-risk genotype (95% CI, 0.2–15%), which would result in a 16% population attributable risk, exceeding estimates of the 5–10% of breast cancer generally attributed to hereditary breast cancer. A simulation study of *BRCA1/2* carriers that found that the majority of hereditary cases of breast cancer have no family history of the disease (28) and that only 2% of hereditary cases attributable to a dominant gene have more than two affected relatives suggests that a significant proportion of hereditary breast cancer occurs outside multiplex families. If so, additional susceptibility genes may have greater population attributable risks than previously thought.

One shortcoming of this study is that only carriers of the Jewish *BRCA1/2* founder mutations were excluded, and families segregating undetected *BRCA1/2* mutations could confound this analysis. If the founder mutations represent 80% of mutations in Ashkenazim (17) and we assume a 60% lifetime penetrance of breast cancer among carriers (16), we would expect a maximum of 12 of the 231 affected probands to be carriers. Our exclusion of families with ovarian cancer might reduce this number somewhat, but other probands may represent "sporadic" cases of breast cancer from families segregating a founder

mutation among the other affected members. However, the models based on these families showed significant heterogeneity compared with *BRCA1/2* carrier families, suggesting that a factor distinct from the known genes exists. A segregation analysis of *BRCA1/2* noncarriers by Cui *et al.* (16) provides empirical data supporting this; testing a two-locus model they found evidence of additional undetected *BRCA1/2* mutations against a background of a second, highly penetrant common recessive allele.

To evaluate the findings of this study, additional segregation analyses could be undertaken in other populations showing familial aggregation of breast cancer outside of *BRCA1/2* families, such as the CASH study population (20), and a group of Scandinavian breast and breast-ovarian families studied in Sweden (29). Although self-reports of family history of breast cancer have been deemed reliable (30–32), additional analyses would ideally be based on families where histories of breast cancer could be verified. Recruiting families through a population-based series of incident cases would also be preferable, to reduce volunteer and survival bias. A sufficient number of families should also be studied to permit estimation of a polygenic model.

Finally, collecting data from relatives as well as probands on factors such as age at first birth would allow direct adjustment for risk factors of which the distribution may vary from generation to generation. This could help avoid misinterpreting generational differences in breast cancer risk factors as a genetic effect. Although there was no strong evidence of such a problem in this study, the adjustment and stratification by birth year alone is difficult to interpret, given the collinearity of birth year with age and age-at onset.

Several other segregation analyses (16, 23, 33–37) have provided various degrees of evidence for recessive inheritance of susceptibility to breast cancer. Two of these studies (16, 37) were specifically designed to look for evidence of genes other than BRCA1 and BRCA2. Cui *et al.* (16) compared recessive, dominant, and codominant models, mixed Mendelian-polygenic models, and two-locus models of independent dominant and recessive genes that would model the BRCA genes alongside other loci. Of these, models of a highly penetrant recessive

gene (virtually 100% by age 70) with allele frequency of 7% (95% CI, 0.05–0.10) were the most parsimonious, and remained robust against backgrounds of polygenic and residual dominant inheritance. The study did not publish individual hypothesis tests for these models, nor did it evaluate sporadic or environmental models. A British study of 1484 families ascertained through a proband affected by age 55 tested models including simultaneous effects for *BRCA1*, *BRCA2,* and a third gene (37), and supported recessive, polygenic, and mixed recessive-polygenic models. The recessive models predicted a gene with high allele frequency (24%; 95% CI, 12.4–41.8) and penetrance of 42% by age 70. This method may have precluded detection of any additional dominant genes, because additional dominant effects were likely to be subsumed by the *BRCA1/2* models.

The parameter estimates presented here could be used as the basis for linkage studies to try to localize this putative recessive gene. Data on existing collections of breast cancer families already genotyped for markers could be used to test for linkage under the proposed recessive model. Collections of affected twins and other sib-pairs may be among the most useful to detect a recessive gene (38). Another, nonparametric method, comparative genomic hybridization, may be useful to discover recessive tumor suppressor genes. Comparative genomic hybridization examines tumor tissue in affected relatives for concordant loss of heterozygosity (14, 39). Candidate regions are identified on the genome where early somatic deletions of genetic material have repeatedly occurred (14, 40).

## Acknowledgments

## References

1. American Cancer Society. Cancer facts and figures 2001. Atlanta, GA 2001.

2. Feuer, E. J., and Wen, L. M. DEVCAN: Probability of developing or dying of cancer, Version 4.0. Bethesda, MD: National Cancer Institute, 1999.

3. Collaborative Group on Hormonal Factors in Breast Cancer. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58, 209 women with breast cancer and 101, 986 women without the disease. Lancet, *358:* 1389–1399, 2001.

4. Balmain, A., Gray, J., and Ponder, B. The genetics and genomics of cancer. Nat. Genet., *33*(Suppl.)*:* 238–244, 2003.

5. King, M. C. Breast cancer genes: how many, where and who are they? Nat. Genet., *2:* 89–90, 1992.

6. Claus, E. B., Schildkraut, J. M., Thompson, W. D., and Risch, N. J. The genetic attributable risk of breast and ovarian cancer. Cancer (Phila.), *77:* 2318–2324, 1996.

7. Easton, D. F., Bishop, D. T., Ford, D., and Crockford, G. P. Genetic linkage analysis in familial breast and ovarian cancer: results from 214 families. Am. J. Hum. Genet., *52:* 678–701, 1993.

8. Wooster, R., Neuhausen, S., Mangion, J., Quirk, Y., Ford, D., Collins, N., Nguyen, K., Seal, S., Tran, T., Averill, D., Fields, P., Marshall, G., Narod, S., Lenoir, G. M., Lynch, H., Feunteun, J., Devilee, P., Cornelisse, C. J., Menko, F. H., Daly, P. A., Ormiston, W., McManus, R., Pye, C., Lewis, C. M., Cannon-Albright, L. A., Peto, J., Ponder, A. J., Skolnick, M. H., Easton, D. F., Goldgar, D. E., and Stratton, M. R. Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12–13. Science (Wash. DC), *265:* 2088–2090, 1994.

9. Szabo, C. I., and King, M. C. Population genetics of BRCA1 and BRCA2. Am. J. Hum. Genet., *60:* 1013–1020, 1997.

10. Rebbeck, T. R., Couch, F. J., Kant, J. Calzone, K., DeShano, M., Peng, Y., Chen, K., Garber, J. E., and Weber, B. L. Genetic heterogeneity in hereditary breast cancer: role of BRCA1 and BRCA2. Am. J. Hum. Genet., *59:* 547–553, 1996.

11. Schubert, E. L., Lee, M. K., Mefford, H. C., Argonza, R. H., Morrow, J. E., Hull, J., Dann, J. L., and King, M. C. BRCA2 in American families with four or more cases of breast or ovarian cancer: recurrent and novel mutations, variable expression, penetrance, and the possibility of families whose cancer is not attributable to BRCA1 or BRCA2. Am. J. Hum. Genet., *60:* 1031–1040, 1997.

12. Serova, O. M., Mazoyer, S., Puget, N., Dubois, V., Tonin, P., Shugart, Y. Y., Goldgar, D., Narod, S. A., Lynch, H. T., and Lenoir, G. M. Mutations in BRCA1 and BRCA2 in breast cancer families: are there more breast cancer-susceptibility genes? Am. J. Hum. Genet., *60:* 486–495, 1997.

13. Seitz, S., Rohde, K., Bender, E., Nothnagel, A., Kolble, K., and Schlag, P. M. Strong indication for a breast cancer susceptibility gene on chromosome 8p12–22: linkage analysis in German breast cancer families. Oncogene, *14:* 741–743, 1997.

14. Kainu, T., Juo, S. H., Desper, R., Schaffer, A. A., Gillanders, E., Rozenblum, E., Freas-Lutz, D., Weaver, D., Stephan, D., Bailey-Wilson, J., Kallioniemi, O. P., Tirkkonen, M., Syrjakoski, K., Kuukasjarvi, T., Koivisto, P., Karhu, R., Holli, K., Arason, A., Johannesdottir, G., Bergthorsson, J. T., Johannsdottir, H., Egilsson, V., Barkardottir, R. B., Johannsson, O., Haraldsson, K., Sandberg, T., Holmberg, E., Gronberg, H., Olsson, H., Borg, A., Vehmanen, P., Eerola, H., Heikkila, P., Pyrhonen, S., and Nevanlinna, H. Somatic deletions in hereditary breast cancers implicate 13q21 as a putative novel breast cancer susceptibility locus. Proc. Natl. Acad. Sci. USA, *97:* 9603–9608, 2000.

15. Claus, E. B., Risch, N., and Thompson, W. D. Genetic analysis of breast cancer in the Cancer and Steroid Hormone Study. Am. J. Hum. Genet., *48:* 232–242, 1991.

16. Cui, J., Antoniou, A. C., Dite, G. S., Southey, M. C., Venter, D. J., Easton, D. F., Giles, G. G., McCredie, M. R., and Hopper, J. L. After BRCA1 and BRCA2— what next? Multifactorial segregation analysis of three-generation, population based Australian families affected by female breast cancer. Am. J. Hum. Genet., *68:* 420–431, 2001.

17. Struewing, J. P., Hartge, P. H., Wacholder, S., Baker, S. M., Berlin, M., McAdams, M., Timmerman, M. M., Brody, L. C., and Tucker, M. A. The risk of cancer associated with specific mutations of BRCA1 and BRCA2 among Ashkenazi Jews. N. Engl. J. Med., *336:* 1401–1408, 1997.

18. Frank, T. S., Deffenbaugh, A. M., Reid, J. E., Hulick, M., Ward, B. E., Lingenfelter, B., Gumpper, K. L., Scholl, T., Tavtigian, S. V., Pruss, D. R., and Critchfield, D. C. Clinical characteristics of individuals with germline mutations in BRCA1 and BRCA2: analysis of 10, 000 individuals. J. Clin. Oncol., *20:* 1480–1490, 2002.

19. Shattuck-Eidens, D., Oliphant, A., McClure, M., McBride, G., Gupte, J., Rubano, T., Pruss, D., Tavtigian, S. V., Teng, D. H. F., Adey, N., Staebell, M., Gumpper, K., Lundstron, R., Hulick, M., Kelly, M., Holmen, J., Lingenfelter, B., Manley, S., Fujimura, F., Luce, M., Ward, B., Cannon-Albright, L., Steele, L., Offit, K., Gilewski, T., Norton, L., Brown, K., Schulz, C., Hampel, H., Schulger, A., Giulotto, E., Zoli, W., Ravaioli, A., Nevanlinna, H., Pyrhonen, S., Rowley, P., Loader, S., Osborne, M. P., Daly, M., Tepler, I., Weinstein, P. L. Scalia, J. L., Michaelson, R., Scott, R. J., Radice, P., Pierotti, M. A., Garber, J. E., Issacs, C., Peshkin, B., Lippman, M. E., Dosik, M. H., Caligo, M. A., Greenstein, R. M., Pilarski, R., Weber, B., Burgemeister, R., Frank, T. S., Skolnick, M. H., and Thomas, A. BRCA1 sequence analysis in women at high risk for susceptibility mutations. Risk factor analysis and implications for genetic testing. JAMA, *278:* 1242–1250, 1997.

20. Claus, E. B., Schildkraut, J., Iversen, E. S., Jr., Berry, D., and Parmigiani, G. Effect of BRCA1 and BRCA2 on the association between breast cancer risk and family history. J. Natl. Cancer. Inst., *90:* 1824–1829, 1998.

21. Kaufman, D. J., and Struewing, J. P. Regarding the effect of BRCA1 and BRCA2 on the association between breast cancer risk and family history. J. Natl. Cancer. Inst., *91:* 1250–1251, 1999.

22. Bonney, G. Regressive logistic models for familial disease and other binary traits. Biometrics, *42:* 611–625, 1986.

23. Go, R. C., King, M. C., Bailey-Wilson, J., Elston, R. C., and Lynch, H. T. Genetic epidemiology of breast cancer and associated cancers in high-risk families. I. Segregation analysis. J. Nat. Cancer. Inst., *71:* 455–461, 1983.

24. Department of Epidemiology and Biostatistics, Case Western Reserve University. S. A. G. E. Statistical Analysis for Genetic Epidemiology, release 3.1, Cleveland, OH 1997.

25. Atwood, L. D., Kammerer, C. M., and Mitchell, B. D. Exploring the HDL likelihood surface. Genet. Epidemiol., *10:* 641–645, 1993.

26. Hanson, R. L., Elston, R. C., Pettitt, D. J., Bennett, P. H., and Knowler, W. C. Segregation analysis of non insulin-dependent diabetes mellitus in Pima Indians: evidence for a major gene effect. Am. J. Hum. Genet., *57:* 160–170, 1995.

27. Borecki, I. B., Province, M. A., and Rao, D. C. Inferring a major gene for quantitative traits by using segregation analysis with tests on transmission probabilities: how often do we miss? Am. J. Hum. Genet., *56:* 319–326, 1995.

28. Cui, J., Hopper, J. L. Why are the majority of hereditary cases of early-onset breast cancer sporadic? A simulation study. Cancer Epidemiol. Biomark. Prev., *9:* 805–812, 2000.

29. Hakansson, S., Johannsson, O., Johannsson, U., Sellberg, G., Loman, N., Gerdes, A. M., Holmberg, E., Dahl, N., Pandis, N., Kristoffersson, U., Olsson, H., and Borg, A. Moderate frequency of BRCA1 and BRCA2 germ-line mutations in Scandinavian familial breast cancer. Am. J. Hum. Genet., *60:* 1068–1078, 1997.

30. Love, R., Evans, A., and Josten, D. The accuracy of patient reports of a family history of cancer. J. Chronic Dis., *38:* 289–293, 1985.

31. Bondy, M. L., Strom, S. S., Colopy, M. W., Brown, B. W., and Strong, L. C. Accuracy of family history of cancer obtained through interviews with relatives of patients with childhood sarcoma. J. Clin. Epidemiol., *47:* 89–96, 1994.

32. Novakovi, B., Goldstein, A. M., and Tucker, M. A. Validation of family history of cancer in deceased family members. J. Natl. Cancer Inst., *88:* 1492–1493, 1996.

33. Goldstein, A. M., Haile, R. W. C., Hodge, S. E. Paganini-Hill A., and Spence M. A. Possible heterogeneity in the segregation pattern of breast cancer in families with bilateral breast cancer. Genet. Epidemiol., *5:* 121–133, 1988.

34. Goldstein, A. M., and Amos, C. I. Segregation analysis of breast cancer from the Cancer and Steroid Hormone Study: histologic subtypes. J. Natl. Cancer Inst., *82:* 1911–1917, 1990.

35. Ciske, D. J., Rich, S. S., King, R. A., Elving Anderson, V., Bartow, S., Vachon, C., McGovern, P. G., Kushi, L. H., Zheng, W., and Sellers, T. A.
Segregation analysis of breast cancer: a comparison of type-dependent age-at-onset versus type-dependent susceptibility models. Genet. Epidemiol., *13:* 317–328, 1996.

36. Pankow, J. S., Vachon, C. M., Kuni, C. C., King, R. A., Arnett, D. K., Grabrick, D. M., Rich, S. S., Anderson, V. E., and Sellers, T. A. J. Genetic analysis of mammographic breast density in adult women: evidence of a gene effect. Natl. Cancer Inst., *89:* 531–533, 1997.

37. Antoniou, A. C., Pharoah, P. D. P., McMullan, G., Day, N. E., Ponder, B. A. J., and Easton, D. Evidence for further breast cancer susceptibility genes in addition to BRCA1 and BRCA2 in a population-based study. Genet. Epidemiol., *21:* 1–18, 2001.

38. Khoury, M., Beaty, T. H., and Cohen, B. H. Fundamentals of Genetic Epidemiology, p. 305–306. Oxford: Oxford University Press, 1993.

39. Forsti, A., Soderberg, M., and Hemminki, K. Use of twins in search for tumor suppressor genes. Environ. Mol. Mutagen., *30:* 231–239, 1997.

40. Hemminki, A., Tomlinson, I., Markie, D., Jarvinen, H., Sistonen, P., Bjorkqvist, A. M., Knuutila, S., Salovaara, R., Bodmer, W., Shibata, D., de la Chapelle, A., and Aaltonen, L. A. Localization of a susceptibility locus for Peutz-Jeghers syndrome to 19p using comparative genomic hybridization and targeted linkage analysis. Nat. Genet., *15:* 87–90, 1997.

# Segregation Analysis of 231 Ashkenazi Jewish Families for Evidence of Additional Breast Cancer Susceptibility Genes

David J. Kaufman, Terri H. Beaty and Jeffery P. Struewing

| | |
|---|---|
| **Updated version** | Access the most recent version of this article at:<br>http://cebp.aacrjournals.org/content/12/10/1045 |

| | |
|---|---|
| **Cited articles** | This article cites 34 articles, 4 of which you can access for free at:<br>http://cebp.aacrjournals.org/content/12/10/1045.full#ref-list-1 |
| **Citing articles** | This article has been cited by 2 HighWire-hosted articles. Access the articles at:<br>http://cebp.aacrjournals.org/content/12/10/1045.full#related-urls |

| | |
|---|---|
| **E-mail alerts** | Sign up to receive free email-alerts related to this article or journal. |
| **Reprints and Subscriptions** | To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org. |
| **Permissions** | To request permission to re-use all or part of this article, use this link<br>http://cebp.aacrjournals.org/content/12/10/1045.<br>Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC)<br>Rightslink site. |