ARTICLES

Genetic Analysis of Mammographic Breast Density in Adult Women: Evidence of a Gene Effect

*James S. Pankow, Celine M. Vachon, Christopher C. Kuni, Richard A. King, Donna K. Arnett, Dawn M. Grabrick, Stephen S. Rich, V. Elving Anderson, Thomas A. Sellers**

Background: **The appearance of the female breast viewed by mammography varies considerably from one individual to another because of underlying differences in the relative proportions of fat, connective tissue, and glandular epithelium that combine to create a characteristic pattern of breast density. An association between mammographic patterns and family history of breast cancer has previously been reported. However, this association has not been found in all studies, and few data are available on possible genetic components contributing to mammographic breast density.** ***Purpose:*** **Our purpose was to estimate familial correlations and perform complex genetic segregation analyses to test the hypothesis that the transmission of a major gene influences mammographic breast density.** ***Methods:*** **As part of a cohort study (initiated in 1944) of families with a history of breast cancer, the probands' female relatives who were older than 40 years were asked to obtain a routine mammogram. The mammograms of 1370 women from 258 independent families were analyzed. The fraction of the breast volume occupied by radiographically dense tissue was estimated visually from video displays of left or right mediolateral oblique views by one radiologist experienced in mammography who had no knowledge of individual relationships to the probands. Data on breast cancer risk factors were obtained through telephone interviews and mailed questionnaires. Unadjusted and adjusted familial correlations in breast density were calculated, and complex genetic segregation analyses were performed.** ***Results:*** **Sister–sister correlations in breast density (unadjusted and adjusted for age and either body mass index, menopausal status, hormone replacement therapy, waist-to-hip ratio, number of live births, alcohol consumption, or cigarette smoking status) were all statistically significant (r = .16-.27; all P<.05 [two-sided]). Estimated mother–daughter correlations were smaller in magnitude (r = .01-.17) and not statistically significant. Segregation analyses indicate that a major autosomal gene influences breast density. The mendelian transmission of a dominant gene provided the best fit to the data; however, hypotheses involving the inheritance of either a recessive gene or a codominant gene could not be ruled out. The mendelian dominant hypothesis, accounting for 29% of the variability in breast density, suggests that approximately 12% of the population would be expected to carry at least one variant allele of this putative gene. Women who inherit the variant allele would have a mean breast density about twice that of the rest of the population.** ***Conclusions:*** **Our preliminary findings suggest that, in this cohort of women at risk of breast cancer, mammographic breast density may be genetically influenced. [J Natl Cancer Inst 1997;89:549-556]**

There is considerable interindividual variability in the radiographic appearance of the female breast; this variability results from underlying differences in the relative proportions of fat, connective tissue, and glandular epithelium (i.e., parenchymal tissue) (*1*). The distribution of radiologically lucent fat and radiologically dense connective and epithelial tissues creates a parenchymal pattern.

Wolfe (*2*) described four breast parenchymal patterns, consisting of primarily fatty breasts (N1), breasts with increasing parenchymal prominence (P1 and P2), and breasts in which the parenchyma is characterized by diffuse or nodular densities (DY). A significant limitation of Wolfe's subjective classification system is the potential for misclassification, and more recent studies [reviewed in (*3*)] have used an alternative classification in which the visible mammographic breast density is expressed as a percentage of total breast area or volume. Studies that use both the Wolfe classification and percent density (*4-6*) show that percent density is a better marker of breast cancer risk. The magnitude of the risk associated with breast density is not completely defined; however, the risk of the highest compared with the lowest category of density may be as large as or larger than that for nearly every other established risk factor (*7-9*). Studies

**Affiliations of authors:* J. S. Pankow, C. M. Vachon, D. K. Arnett, D. M. Grabrick, V. E. Anderson (Division of Epidemiology, School of Public Health), C. C. Kuni (Department of Radiology, School of Medicine), R. A. King (Department of Medicine, Division of Genetics and Metabolism, Institute of Human Genetics, and Cancer Center), T. A. Sellers (Division of Epidemiology, School of Public Health, Institute of Human Genetics, and Cancer Center), University of Minnesota, Minneapolis; S. S. Rich, Department of Public Health Sciences, Bowman Gray School of Medicine, Winston-Salem, NC.

Correspondence to: Thomas A. Sellers, Ph.D., M.P.H., Division of Epidemiology, University of Minnesota, 1300 South Second St., Suite 300, Minneapolis, MN 55454-1015.

See ''Notes'' following ''References.''

that have examined the association of other breast cancer risk factors with parenchymal patterns and mammographic density have shown a positive association with younger age, late age at menarche, nulliparity, low body mass index (BMI), small breasts, late age at first live birth, never having breast fed, and history of breast cancer. Conversely, there does not appear to be an association between breast density and type of menopause, age at menopause, or duration of breast feeding (*7*). Although cross-sectional studies have generally not observed an association between replacement hormones and breast density, a recent longitudinal study (*10*) has demonstrated an increase in breast density in women after they begin taking replacement estrogen.

An association between mammographic patterns and a family history of breast cancer has been reported (*11-13*), but this association has not been found in all studies (*4,14,15*). Most of these studies used Wolfe's qualitative classification. Few data are available on the possible genetic components of breast density. A sample of 110 mother–daughter pairs and 122 sister pairs was compared, and a greater agreement in parenchymal patterns than expected by chance was found (*16*). The fact that the subjects in this study were selected by physician referral suggests that the results may not generally apply. More importantly, the data provide no insight into the potential mode of genetic inheritance.

The purpose of our study was to estimate familial correlations and perform a complex segregation analysis to test the hypothesis that transmission of a major gene influences mammographic breast density.

Subjects and Methods

Population

Details of the study design and methods have been published (*17*). The study was approved by the University of Minnesota Institutional Review Board, and written informed consent was obtained from each subject. Briefly, a family study of breast cancer initiated in 1944 ascertained female breast cancer probands at the Tumor Clinic of the University of Minnesota Hospital with microscopic diagnosis or clinical evidence of breast cancer. Information on 544 probands and history of cancer in their relatives was obtained by interviews, medical history questionnaires, and death certificates.

We are now conducting a follow-up study of this cohort of families. Members included in the current investigation were sisters, daughters, nieces, and granddaughters of the breast cancer probands and the spouses of brothers, sons, nephews, and grandsons of the breast cancer probands. The pedigrees of each family were updated to ascertain whether each family member was alive and, if so, his or her current address and telephone number. For deceased individuals, the date of death and address and telephone number of a designated surrogate, preferably a spouse or a first-degree relative, were obtained. Information on any additional family members who were not born at the time of the baseline study or who had subsequently married into the family was procured.

Questionnaire Data

Telephone interviews were completed with all available female relatives older than 18 years. The following specific topics were covered: history of cancer, education, benign or malignant breast disease, mammography, menstrual and pregnancy histories, oral contraceptive use, hormone replacement therapy, physical activity, cigarette smoking, and alcohol consumption. Data on deceased female relatives were obtained through an abbreviated interview completed by the designated surrogate. After the telephone interview, living subjects were mailed a Body Measurements Questionnaire (and tape measure) designed to elicit replicated measures of current height and weight. Circumferences of the waist (2 inches above the umbilicus) and hips (maximal protrusion) were determined according to a validated protocol (*18*).

Estimation of Breast Density

To verify the breast cancer status of relatives and spouses, women older than age 40 were asked to obtain a routine mammogram through their personal physician if one had not been taken in the previous year for women older than age 50 (or the previous 2 years for women under age 50). Most mammograms (94.5%) were taken within 1 year of the telephone survey; only 1.4% were taken in an interval greater than 2 years. Mammograms of poor technical quality (63 on the left breast and 65 on the right breast) were not analyzed. The fraction of the breast volume occupied by fibroglandular stroma was estimated visually by a radiologist experienced in mammography (C. C. Kuni), who had no knowledge of the relationship to the proband. The estimate was made from a video display of the left mediolateral oblique view; the right mediolateral oblique view was used if the left was unavailable. The mammographic films were scanned with a Lumican 100 film digitizer (Lumisys, Inc., Sunnyvale, CA) in $1024 \times 1024 \times 16$ format to generate digital data used to display the images. Software used to analyze the digitized image was created locally. Contrast and brightness of the displayed images were routinely manipulated to facilitate density estimation. A nearly continuous (5% increments) quantitative estimate of breast density was made without consideration of stromal pattern. The intra-reader reliability of estimation, based on re-reading more than 1200 mammograms, was 0.82.

Multiple Regression Analysis

Multiple regression analysis was performed using PROC GLM in SAS (*19*) to identify factors (demographic, anthropometric, lifestyle, and reproductive) associated with mammographic breast density expressed as a continuous trait. Independence of the observations was assumed throughout. Age, BMI (kg/m^2), waist-to-hip ratio (WHR), number of live births, and age at menarche were modeled as continuous variables. Education was classified into three levels (less than a high school graduate, high school graduate, or college graduate). Alcohol consumption was divided into three levels (daily, less than daily, or nondrinker). Cigarette smoking status (ever or never), oral contraceptive use (ever or never), hormone replacement therapy (current user or nonuser), and menopausal status (premenopausal or postmenopausal) were modeled as dichotomous variables. Physical activity was converted into an index (high, moderate, or low) on the basis of reported frequencies of vigorous and moderate activities.

Familial Correlations

Familial correlations of estimated mammographic breast density were estimated for mother–daughter and and sister pairs by use of the FCOR (i.e., familial correlation) program of the Statistical Analysis for Genetic Epidemiology Software Package (*20*). Correlations were computed by three weighting methods: 1) equal weight to pairs, 2) equal weight to pedigrees, and 3) equal weight to nuclear families.

Breast density was inversely correlated with age ($r = -.34$; $P<.001$). Therefore, correlations were computed by each weighting method, with the mean of breast density dependent on 10 age groups (i.e., 40-45 years, 46-50 years, 51-55 years, 56-60 years, 61-65 years, 66-70 years, 71-75 years, 76-80 years, 81-85 years, and $\geq$86 years). Additional analyses were performed in which the means were also dependent on risk factors (i.e., age, BMI, WHR, menopausal status, hormone replacement therapy, number of live births, alcohol consumption, and cigarette smoking status) that have been shown in our multiple regression analyses to have an independent association with breast density. These variables were categorized into groups of five or fewer because of the restrictions of the FCOR program. Fisher's *z* transformation was used to test whether the familial correlations were significantly different from zero.

Segregation Analyses

Segregation analyses were performed by use of the REGC program (Segregation Analysis of a Continuous Trait Under Approximations to Class A, B, C, and D Regressive Models, Version 5.0) in the Statistical Analysis for Genetic Epidemiology Software Package (*20*). This model assumes that mendelian inheritance, if present, is due to a single autosomal locus with two alleles. Class D regressive models, which allow residual correlation between siblings after allowance for common parentage, were employed. Hardy–Weinberg equilibrium was assumed, and no allowance was made for major gene (or type) effect on variance estimates. Moreover, no allowance was made for sex-dependent parameters or residual spouse and father–offspring correlations because breast density measurements were available only for women.

Four mendelian hypotheses, three non-mendelian hypotheses, and one general (unrestricted) hypothesis were fitted to the data. The mendelian hypotheses were as follows: 1) dominant inheritance, 2) recessive inheritance, 3) codominant inheritance, and 4) additive inheritance. The non-mendelian hypotheses were as follows: 1) no major type, with no parent–offspring transmission and a single population mean; 2) random environmental inheritance, with no parent–offspring transmission but possible heterogeneity between founders and nonfounders; and 3) random environmental inheritance, with no parent–offspring transmission and no heterogeneity between founders and nonfounders. The general model imposed no restrictions on the mode of transmission or other parameters and was used to base statistical comparisons.

In the REGC program, parameter estimates were obtained by use of the method of maximum likelihood; likelihood ratio tests were calculated to assess the goodness-of-fit of a particular hypothesis. Twice the difference in $\log_e$ (likelihood) values between the hypothesis of interest and the general model is distributed asymptotically as a chi-squared statistic. The degrees of freedom are given by the difference in the number of parameters estimated in the two hypotheses being compared. Akaike's Information Criterion (AIC) (*21*) was computed to help identify the most parsimonious model. It is equal to minus two times the $\log_e$ likelihood plus two times the number of parameters estimated.

Segregation analyses were initially performed by use of class D models without inclusion of covariates in the model. Covariates were then added by forward selection, beginning with the covariate (BMI) most strongly associated with breast density in multiple regression analysis. Multiple regression coefficients for these covariates were used as initial maximum likelihood parameter estimates. Continuous variables (age, age^2, BMI, WHR, and number of live births) were expressed as deviations from their population means. If the addition of a specific covariate significantly improved the fit of any of the hypotheses ($P<.05$), then the covariate was incorporated into the other hypotheses and retained in all later runs that evaluated additional covariates.

Estimates of any residual polygenic component in regressive models were achieved through familial correlations. Separate parameters for residual mother–daughter and sister–sister correlations were estimated for all of the hypotheses.

Analytic Population

Of the 5915 living female participants, 4943 (83.6%) were older than age 40 years and potentially eligible for mammography. Of these 4943, 2580 (52.2%) had recently had a mammogram and were asked to have the original films submitted for analysis. The remaining 2363 (47.8%) women were invited to have a mammogram as part of the study; a total of 1003 (20.3%) refused. Individuals with incomplete covariate information and pedigrees with only one individual with a breast density measurement were not included in the present analysis. Women with a personal history of breast cancer were also excluded to reduce the potential for misclassification bias. The present analysis was based on mammographic breast densities that had been measured for 1370 women in 258 independent families. A total of 832 women (60.7%) were genetically related to the original breast cancer probands; the remaining 538 women (39.3%) married into the families.

Results

Table 1 presents demographic, anthropometric, lifestyle, and reproductive characteristics of the participants with mammographic breast density measurements. Participants ranged in age from 40 years to 93 years (median, 61 years); approximately 81% were postmenopausal. Ninety-one percent of the women had given birth to at least one child. The mammographic breast density (mean ± standard deviation) was 33% ± 16%. The 25th, 50th, and 75th percentile values were 20%, 30%, and 40%, respectively. The estimated percent density was similar for blood relatives and for persons who had married into the family (33.6% versus 32.1%, respectively; $P = .19$).

Covariate Selection

Thirteen variables were evaluated in multiple regression analysis as possible covariates based on their association with breast density in the present study or other published studies (*6,7,9,15*). These variables were age (and age^2), education level, BMI, WHR, physical activity index, cigarette smoking status, alcohol consumption, menopausal status, oral contraceptive use, hormone replacement therapy, age at menarche, and number of live births. In multiple regression, physical activity index, oral contraceptive use, education level, and age at menarche were removed by use of backward elimination ($P>.10$). All remaining variables were at least weakly associated with breast density ($P\leq.10$) and were retained for evaluation as possible covariates in later segregation analyses. The multiple R^2 for the final model (after backward elimination) was .31.

Table 1. Distribution of demographic, anthropometric, lifestyle, and reproductive variables for women with breast density readings

Age, y (%)	Educational level (%)	Relation to proband (%)
Demographic		
⩽50 (21)	Less than a high school graduate (14)	Blood relative (61)
51-60 (28)	High school graduate (70)	Married into the family (39)
61-70 (30)	College graduate (16)	
71-80 (17)		
⩾81 (4)		

Body mass index, kg/m^2 (%)	Waist-to-hip ratio (%)
Anthropometric	
⩽22.89 (21)	⩽0.76 (17)
22.90-25.04 (18)	0.77-0.80 (17)
25.05-27.43 (19)	0.81-0.85 (25)
27.44-30.69 (19)	0.86-0.90 (17)
⩾30.70 (23)	⩾0.91 (24)

Physical activity index (%)	Cigarette smoking status (%)	Alcohol consumption (%)
Lifestyle		
Low (33)	Never (58)	Never (14)
Moderate (42)	Ever (42)	Less than daily (82)
High (25)		Daily (4)

Menopausal status (%)	Use of oral contraceptives (%)	Hormone replacement therapy (%)
Reproductive		
Premenopausal (19)	Never (51)	Nonuser (71)
Postmenopausal (81)	Ever (49)	Current user (29)

Age at menarche, y (%)	No. of live births (%)
⩽10 (5)	0 (9)
11-13 (66)	1-2 (33)
14-16 (27)	3-4 (41)
17-19 (2)	5-6 (12)
	⩾7 (5)

Familial Correlations

Of the 1370 women with breast density readings, 65 mother–daughter pairs from 41 pedigrees and 275 sister pairs from 112 pedigrees formed the basis for the estimation of familial correlations. Table 2 presents the familial correlations calculated from the equal-weight-to-pairs method only, inasmuch as all three weighting methods produced similar results. Simple and age-

Table 2. Familial correlation of breast density with mean dependent on age, anthropometric, reproductive, and lifestyle factors using the equal-weight-to-pairs weighting method

Adjustment to familial correlation*	Mother–daughter pairs		Sister pairs	
	r	*P*	*r*	*P*
Unadjusted	.11	.39	.27	.0002
Age	.15	.22	.22	.0002
Age, BMI	.02	.99	.16	.014
Age, menopausal status	.14	.28	.21	.0022
Age, hormone replacement therapy	.17	.22	.23	.0011
Age, WHR	.01	.93	.20	.0022
Age, No. of live births	.16	.19	.21	.0010
Age, alcohol consumption	.09	.46	.19	.0026
Age, cigarette smoking status	.16	.22	.22	.0004

*Adjustment to familial correlation consists of mean of breast density dependent on the following groups: age in 10 categories (40-45 years, 46-50 years, 51-55 years, 56-60 years, 61-65 years, 66-70 years, 71-75 years, 76-80 years, 81-85 years, and ⩾86 years); body mass index (BMI; kg/m^2) in five categories (⩽22.89, 22.90-25.04, 25.05-27.43, 27.44-30.69, and ⩾30.70); menopausal status in two categories (premenopausal and postmenopausal); hormone replacement therapy in three categories (current, former, and never users); waist-to-hip ratio (WHR) in five categories (⩽0.76, 0.77-0.80, 0.81-0.85, 0.86-0.90, and ⩾0.91); number of live births in five categories (0, 1-4, 5-8, 9-12, and ⩾13); alcohol consumption in four levels (daily, weekly, monthly, and never); and cigarette smoking in two categories (ever and never).

adjusted correlations were .11 and .15, respectively, for mother–daughter pairs; the corresponding correlations for sister pairs were .27 and .22. Both the unadjusted and age-adjusted sister pair correlations were statistically significant ($P<.05$). Subsequent analyses were performed to adjust singly for several risk factors, in addition to age, that previous analyses suggested were correlated with density. None of the additional analyses resulted in statistically significant mother–daughter correlations, although the mother–daughter and sister–sister correlations were in the same direction and of similar magnitude. The estimated correlation of breast density for the sister pairs, with the mean dependent on age and BMI, was statistically significant ($P<.05$) ($r = .16$), as were the correlations with the mean dependent on age and menopausal status ($r = .21$), age and hormone replacement therapy ($r = .23$), age and WHR ($r = .20$), age and number of live births ($r = .21$), age and alcohol consumption ($r = .19$), and age and cigarette smoking status ($r = .22$).

Segregation Analyses

Initial results of segregation analyses of mammographic breast density, without adjustment for covariates, are shown in Table 3. Parameter estimates for the four mendelian hypotheses, three non-mendelian hypotheses, and the general model are shown, along with a likelihood ratio test to assess the goodness-of-fit test and the AIC for each model. Tests for nonzero residual mother–daughter correlations were statistically significant for all hypotheses. In no hypothesis, however, did the addition of a residual sibling correlation improve the fit. Parameter estimates for residual sibling correlations were therefore fixed at zero in all subsequent runs. All four mendelian hypotheses and all three non-mendelian hypotheses were rejected ($P<.05$). According to AIC, the general model provided the most parsimonious fit to the data.

Risk factors found to be associated with mammographic breast density in multiple regression analysis were then considered for inclusion as covariates in the complex segregation analysis. Covariates were evaluated one at a time, starting with those that exhibited the strongest correlations with breast den-

Table 3. Segregation analysis of mammographic breast density: maximum likelihood estimates and model fit without adjustments for covariates*

Parameter†	Hypothesis							
	Mendelian inheritance				Non-mendelian			
	Dominant	Recessive	Codominant	Additive	No major type	Random environment‡	Random environment§	General model
q_A	0.07	0.36	0.28	0.06		0.27	0.28	0.27
τ_{AA}	1.00‖	1.00‖	1.00‖	1.00‖		0.29	0.28¶	0.00≠
τ_{AB}	0.50‖	0.50‖	0.50‖	0.50‖		0.29¶	0.28¶	0.57
τ_{BB}	0.00‖	0.00‖	0.00‖	0.00‖		0.29¶	0.28¶	0.13
μ_{AA}	63.18	62.86	68.27	97.30	32.91	67.91	67.95	67.94
μ_{AB}	63.18¶	28.47	39.42	62.98‖		39.88	39.90	39.82
μ_{BB}	28.59	28.47¶	21.94	28.66		21.64	21.64	21.63
σ^2	130.75	128.53	69.92	131.41	261.92	59.53	59.51	60.23
ρ parent–offspring	0.24	0.24	0.26	0.24	0.23	0.19	0.20	0.23
χ^2 (*df*)**	40.23 (3-4)	37.28 (3-4)	31.94 (2-3)	47.47 (3-4)	268.85 (5-6)	14.72 (1-2)	16.52 (2-3)	
P††	<.0001	<.0001	<.0001	<.0001	<.0001	.00012	.00026	
Akaike's Information Criterion (*21*)	11 293.23	11 290.28	11 286.94	11 300.47	11 517.85	11 271.72	11 271.52	11 259.00

*Spouse and sibling residual correlations set to zero.

†q_A = estimate of allele frequency; τ_u = probability that an individual of type u transmits the A allele, where $u = AA, AB,$ and BB; μ_u = mean breast density for individuals with type u; σ^2 = variability; ρ = parent–offspring, correlation between parents and offspring.

‡Allowance for heterogeneity between founders and nonfounders ($\tau \neq q_A$).

§No allowance for heterogeneity between founders and nonfounders ($\tau = q_A$).

‖Parameter is fixed at this value and not estimated in this model.

¶Parameter is constrained to equal the preceding parameter and is not estimated.

≠Parameter estimate went to a bound.

**Because one parameter estimate went to a bound, a range of degrees of freedom (*df*) is shown.

††One-sided.

sity. Table 4 summarizes the results of the model-building process for four hypotheses (mendelian dominant, no major type, environmental, and general). The addition of measures of BMI, age, age^2, menopausal status, hormone replacement therapy, WHR, number of live births, alcohol consumption, and cigarette smoking all significantly improved the fit of at least one of the hypotheses tested.

After adjustment for BMI, age, and age^2, estimates of residual mother–daughter correlations were greatly reduced and were no longer statistically significant for any hypothesis (data not shown). For example, the estimate of the residual mother–daughter correlation in the no major type model was reduced from .23 to .08 after adjustment for the BMI and age variables.

Table 5 presents the parameter estimates, goodness-of-fit statistics, and the AIC obtained from segregation analyses of mammographic breast density adjusted for all covariates. Residual mother–daughter correlations were fixed to zero for all hypotheses. The results indicated that breast density was inversely associated with BMI, age, WHR, and number of live births and positively associated with hormone replacement therapy and alcohol consumption. Breast density was lower in postmenopausal women than in premenopausal women and lower in ever smokers than in never smokers.

After adjustment for covariates, the mendelian additive hypothesis and all three non-mendelian hypotheses were rejected ($P<.05$). Conversely, mendelian transmission of either a dominant, recessive, or codominant major gene could not be rejected. The dominant hypothesis, however, had the lowest AIC value and, therefore, provided the best fit under this criterion.

Mean breast density estimates for the final general model were not ordered (μ_{AA}: 56%; μ_{AB}: 21%; μ_{BB}: 33%). Local maxima with ordered means were found after we tried a variety of initial parameter estimates to search the likelihood surface. None of these maxima, however, improved the $-2 \log_e$ likelihood of the general model shown in Table 5.

Parameter estimates from the three mendelian hypotheses that fit the data were consistent. The mendelian dominant hypothesis suggests that approximately 12% of the population would be expected to *carry at least one allele* associated with higher breast density. According to the model, individuals who inherit this putative gene would have a mean breast density about twice as high as that of the rest of the population (54% versus 27%). Parameter estimates from the mendelian recessive hypothesis suggest that 12% of the population are *homozygous* for the allele associated with higher mean density. Similar to the results obtained for the mendelian dominant hypothesis, the mean density in women carrying two recessive alleles was estimated to be 54%, compared with 27% in women presumed to carry at least one wild-type allele. Parameter estimates from the mendelian codominant hypothesis correspond to the estimates for the dominant model, with the mean density of the heterozygotes very similar to that of individuals homozygous for the allele associated with a higher mean density.

The variance attributable to the putative major gene was estimated by use of data from Tables 3 and 5. The variance estimate from the no major gene model with no covariates was 261.9 (Table 3). Addition of covariates to the no major gene model reduced the estimated variance to 179.9 (Table 5). Therefore, the variance in breast density explained by covariates is approximately 31% [i.e., (261.9 − 179.9)/261.9]. The estimated variance in breast density from the mendelian dominant hypothesis in Table 5 (103.3) includes the variance explained by covariates. Therefore, the variance in breast density accounted for by the putative locus is simply the relative proportion of total variance (261.9) comprising by the difference between the no major gene hypothesis (with covariates) and the mendelian dominant hypothesis [i.e., (179.9 − 103.3)/261.9], or 29%. The combination of the putative locus and the covariates accounts for 60% of the total variance, leaving 40% unaccounted for by the model.

Discussion

Although many studies have investigated the possible association of demographic, anthropometric, lifestyle, and reproductive characteristics with breast parenchymal patterns or breast density, virtually none have explored possible genetic and familial inheritance. Several studies (*6,11-13,22-25*) have found mammographic patterns to be associated with family history of breast cancer, but other studies (*4,14,15,26-29*) have found no such relationship. Nearly all of these studies used Wolfe's qualitative classification system rather than a quantitative estimate of mammographic breast density to assess mammographic patterns as used in the current study.

Our study of familial correlations suggests a heritable component of breast density. All correlations for the sister pairs

Table 4. Segregation analysis of mammographic breast density: improvement of fit of selected models after adjustment for covariates*

		Hypothesis							
		Dominant		No major type		Random environment		General	
Model	Covariates in model	IOF χ^2 (*df*)	*P*	IOF χ^2 (*df*)	*P*	IOF χ^2 (*df*)	*P*	IOF χ^2 (*df*)	*P*
1	None								
2	Model 1 + BMI	141.09 (1)	<.0001	204.93 (1)	<.0001	132.34 (1)	<.0001	137.68 (1)	<.0001
3	Model 2 + age, age^2	230.53 (2)	<.0001	254.10 (2)	<.0001	192.62 (2)	<.0001	196.05 (2)	<.0001
4	Model 3 + menopausal status	13.40 (1)	.00025	11.69 (1)	.00063	13.12 (1)	.00029	14.39 (1)	.00015
5	Model 4 + hormone replacement therapy	16.86 (1)	<.0001	15.59 (1)	<.0001	16.38 (1)	<.0001	17.17 (1)	<.0001
6	Model 5 + WHR	9.01 (1)	.0027	9.80 (1)	.0017	8.11 (1)	.0044	8.33 (1)	.0039
7	Model 6 + No. of live births	3.95 (1)	.047	9.42 (1)	.0021	3.32 (1)	.068	2.70 (1)	.10
8	Model 7 + alcohol consumption	8.46 (2)	.015	6.16 (2)	.046	10.62 (2)	.0049	9.18 (2)	.010
9	Model 8 + cigarette smoking	4.06 (1)	.044	3.46 (1)	.063	7.03 (1)	.0080	3.97 (1)	.046

*IOF = improvement of fit; *df* = degrees of freedom; BMI (kg/m^2) = body mass index; WHR = waist-to-hip ratio. IOF compared with previous model (one-sided *P*).

Table 5. Segregation analysis of mammographic breast density: maximum likelihood estimates and model fit with adjustments for covariates*

	Hypothesis							
	Mendelian inheritance				Non-mendelian			
Parameter†	Dominant	Recessive	Codominant	Additive	No major type	Random environment‡	Random environment§	General model
q_A	.06	.35	.06	.06		.27	.29	.31
τ_{AA}	1.00‖	1.00‖	1.00‖	1.00‖		.30	.29¶	1.00≠
τ_{AB}	0.50‖	0.50‖	0.50‖	0.50‖		.30¶	.29¶	.50
τ_{BB}	0.00‖	0.00‖	0.00‖	0.00‖		.30¶	.29¶	.03
μ_{AA}	54.35	54.11	54.64	80.96	29.74	58.52	58.47	56.45
μ_{AB}	54.35¶	26.90	54.32	54.02‖		35.65	35.63	20.72
μ_{BB}	27.01	26.90¶	27.01	27.08		21.04	21.01	33.44
σ^2	103.30	102.09	103.28	104.04	179.86	58.00	57.95	63.46
BMI	−.74	−.75	−.74	−.74	−.90	−.75	−.76	−.74
Age	−.31	−.31	−.31	−.31	−.36	−.30	−.30	−.29
Age^2	+.009	+.010	+.010	+.010	+.011	+.009	+.009	+.010
Menopausal status	−5.24	−5.30	−5.20	−5.24	−5.56	−4.98	−5.00	−5.08
Hormone replacement therapy, current	+2.72	+2.74	+2.70	+2.71	+2.81	+2.53	+2.55	+2.71
WHR	−13.52	−13.48	−13.51	−13.59	−16.21	−13.27	−13.21	−13.01
No. of live births	−.29	−.28	−.29	−.29	−.56	−.24	−.25	−.22
Alcohol consumption, daily	+6.04	+6.05	+6.04	+5.92	+5.55	+6.82	+6.83	+6.32
Alcohol consumption, less than daily	+1.61	+1.66	+1.62	+1.64	+2.26	+2.07	+2.08	+1.67
Cigarette smoking, ever	−1.36	−1.42	−1.36	−1.39	−1.42	−1.79	−1.77	−1.33
χ^2 (*df*)**	3.77 (3-4)	6.34 (3-4)	3.77 (2-3)	9.86 (3-4)	143.13 (5-6)	20.88 (1-2)	22.03 (2-3)	
P††	.29	.096	.15	.020	<.0001	<.0001	<.0001	
Akaike's Information Criterion (*21*)	10 889.88	10 892.45	10 891.88	10 895.97	11 024.24	10 910.99	10 910.14	10 894.11

*q_A = estimate of allele frequency; τ_u = probability that an individual of type *u* transmits the *A* allele, where *u* = *AA*, *AB*, and *BB*; μ_u = mean breast density for individuals with type *u*; σ^2 = variability. All residual family correlations set to zero.

†BMI (kg/m^2) = body mass index; WHR = waist-to-hip ratio.

‡Allowance for heterogeneity between founders and nonfounders ($\tau \neq q_A$).

§No allowance for heterogeneity between founders and nonfounders ($\tau = q_A$).

‖Parameter is fixed at this value and not estimated in this model.

¶Parameter is constrained to equal the preceding parameter and is not estimated.

≠Parameter estimate went to a bound.

**Because one parameter estimate went to a bound, a range of degrees of freedom (*df*) is shown.

††One-sided.

(adjusted and unadjusted) were statistically significant. The estimated mother–daughter correlations in breast density were smaller in magnitude and not statistically significant. Under an hypothesis of genetic transmission, one might have expected the mother–daughter and sister–sister correlations to be of a similar magnitude. Although the exact reasons for this apparent inconsistency are unknown, several possible explanations can be advanced. The first is random variation due to a small sample size; there were 275 sister pairs but only 65 mother–daughter pairs. The second is lack of data on the fathers. If a putative gene for breast density was transmitted in an autosomal dominant manner, roughly half of the time it would occur from father to daughter, a mechanism that would clearly not be reflected in mother–daughter correlations. Also, the software used to estimate familial correlations permitted adjustment for only age and one covariate at a time. Alternatively, it can also be argued that the observed differences in the magnitude of mother–daughter versus sister–sister correlations are consistent with expectations. In particular, if there is no dominance at the major locus with an additive variance of near zero, then there is no dominance contribution to the covariance between mothers and daughters.

Segregation analysis provided additional evidence for a major gene influencing mammographic breast density. After incorporation of covariates on BMI, age, age^2, menopausal status, hormone replacement therapy, WHR, number of live births, alcohol consumption, and cigarette smoking, hypotheses of mendelian transmission of a dominant, recessive, or codominant major gene all fit the data as well as the general model. A possible limitation of these data is the temporal difference between collection of risk factor data and the appointment for mammography. Although for roughly 95% of the study subjects, this interval was less than 1 year, it is conceivable that some exposures, in particular oral contraceptive use and hormone replacement therapy, have been misclassified. It should also be pointed out, however, that all of the other covariates incorporated into the segregation analyses would not be affected in this manner.

Two findings from the segregation analysis are particularly noteworthy. First, residual mother–daughter correlations (the equivalent of the polygenic component of the trait) were virtually eliminated after demographic, anthropometric, lifestyle, and reproductive variables were added to the models. This result suggests that, together with possible genetic influences, other

characteristics shared by family members may account for some of the familial aggregation of breast density. Second, statistical evidence for a major gene emerged only *after* breast density was adjusted for covariates, particularly BMI and age. This finding underscores the value of including relevant covariates in genetic analyses, inasmuch as they may represent confounders of the familial pattern of a trait; i.e., to the extent that these covariates are associated with breast density but differ between family members, the degree of familial correlation will be reduced. Furthermore, these environmental covariates can influence the variability of the trait. By accounting for these factors, the ''noise'' is reduced, and the ability to detect the familial pattern is enhanced.

The present analyses were conducted without a correction for ascertainment. Although nonrandom sampling can lead to biased genetic parameter estimates and incorrect inferences (*30*), the magnitude of such bias in this case may be small. No probands with measured breast density were included in the present analyses, and relatives were measured, on average, two generations after proband ascertainment. Also, the variable (breast cancer) used for ascertainment of probands in this study was different from the phenotype (breast density). When the phenotype is not identical to the variable used in selecting probands, the potential for bias in analyses based exclusively on non-probands is minimal, unless there is a strong correlation between the phenotype and selection variable (*31*).

In our study, the limited ability to distinguish between competing mendelian hypotheses may be due primarily to the inability to obtain information on breast density for all female relatives and the complete absence of breast density data for male family members. Of all the hypotheses, mendelian dominant inheritance appears to provide the best fit to the data for two reasons. First, the AIC for the dominant model was lowest in adjusted analyses. Second, the maximum likelihood estimates for the codominant model, which allows for an arbitrary mean density in heterozygotes, were remarkably similar to the corresponding estimates for the dominant model. Under the mendelian dominant hypothesis, women carrying one or two dominant alleles would have a mean breast density approximately 27% higher than the rest of the female population.

Mammographic breast density primarily reflects proliferation of breast stroma through collagen formation and fibrosis (*9*). The formation and maintenance of breast densities likely depend on the complex interplay of hormones, such as estrogen, and growth factors, such as epidermal growth factor, transforming growth factor, and insulin-like growth factors I and II (*9*). Genes involved in the regulation of hormone metabolism and the activity of growth factors, therefore, may be important in determining mammographic breast density.

The current finding of a major gene influencing mammographic breast density should be viewed as preliminary. The fact that the best fitting hypothesis explained only 60% of the variability in breast density (29% was genetic only; when covariates were included, the total is 60%) implies that the models are missing a number of key variables, such as other genes, environment, or gene × environment interactions. These findings should be replicated. If the tendency toward dense breasts is related to an inherited tendency to develop breast cancer, then a random sample of families may provide a better estimate of the genetic component. Ultimate verification of the major gene hypothesis, however, requires confirmation by identification of a gene by linkage analysis.

References

(*1*) Feig SA. Breast masses. Mammographic and sonographic evaluation. Radiol Clin North Am 1992;30:67-92.

(*2*) Wolfe JN. Breast patterns as an index of risk for developing breast cancer. AJR Am J Roentgenol 1976;126:1130-7.

(*3*) Byng JW, Boyd NF, Fishell E, Jong RA, Yaffe MJ. The quantitative analysis of mammographic densities. Phys Med Biol 1994;39:1629-38.

(*4*) Saftlas AF, Hoover RN, Brinton LA, Szklo M, Olson DR, Salane M, et al. Mammographic densities and risk of breast cancer. Cancer 1991;67: 2833-8.

(*5*) Boyd NF, Byng JW, Jong RA, Fishell EK, Little LE, Miller AB, et al. Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening Study. J Natl Cancer Inst 1995;87:670-5.

(*6*) Byrne C, Schairer C, Wolfe J, Parekh N, Salane M, Brinton LA, et al. Mammographic features and breast cancer risk: effects with time, age, and menopause status. J Natl Cancer Inst 1995;87:1622-9.

(*7*) Saftlas AF, Szklo M. Mammographic parenchymal patterns and breast cancer risk. Epidemiol Rev 1987;9:146-74.

(*8*) Warner E, Lockwood G, Tritchler D, Boyd NF. The risk of breast cancer associated with mammographic parenchymal patterns: a meta-analysis of the published literature to examine the effect of method of classification. Cancer Detect Prev 1992;16:67-72.

(*9*) Oza AM, Boyd NF. Mammographic parenchymal patterns: a marker of breast cancer risk. Epidemiol Rev 1993;15:196-208.

(*10*) Laya MB, Larson EB, Taplin SH, White E. Effect of estrogen replacement therapy on the specificity and sensitivity of screening mammography. J Natl Cancer Inst 1996;88:643-9.

(*11*) Wilkinson E, Clopton C, Gordonson J, Green R, Hill A, Pike MC. Mammographic parenchymal pattern and the risk of breast cancer. J Natl Cancer Inst 1977;59:1397-400.

(*12*) Hainline S, Myers L, McLelland R, Newell J, Grufferman S, Shingleton W. Mammographic patterns and risk of breast cancer. AJR Am J Roentgenol 1978;130:1157-8.

(*13*) Saftlas AF, Wolfe JN, Hoover RN, Brinton LA, Schairer C, Salane M, et al. Mammographic parenchymal patterns as indicators of breast cancer risk. Am J Epidemiol 1989;129:518-26.

(*14*) Kaufman Z, Garstin WI, Hayes R, Michell MJ, Baum M. The mammographic parenchymal patterns of nulliparous women and women with a family history of breast cancer. Clin Radiol 1991;43:385-8.

(*15*) Brisson J, Sadowsky NL, Twaddle JA, Morrison AS, Cole P, Merletti F. The relation of mammographic features of the breast to breast cancer risk factors. Am J Epidemiol 1982;115:438-43.

(*16*) Wolfe JN, Albert S, Belle S, Salane M. Familial influences on breast parenchymal patterns. Cancer 1980;46:2433-7.

(*17*) Sellers TA, Anderson VE, Potter JD, Bartow SA, Chen PL, Everson L, et al. Epidemiologic and genetic follow-up study of 544 Minnesota breast cancer families: design and methods. Genet Epidemiol 1995;12:417-29.

(*18*) Weaver TW, Kushi LH, McGovern PG, Potter JD, Rich SS, King RA, et al. Validation study of self-measures of fat distribution. Int J Obes Metab Disord 1996;20:644-50.

(*19*) SAS Institute Inc. SAS/STAT user's guide, version 6, 4th edition, vol 2. Cary (NC): SAS Institute Inc., 1989.

(*20*) S.A.G.E. Statistical analysis for genetic epidemiology, release 2.2. Cleveland (OH): Department of Epidemiology and Biostatistics, Case Western Reserve University, 1994.

(*21*) Akaike H. A new look at the statistical model identification. IEEE Trans Automatic Control 1974;19:716-26.

(*22*) Boyd NF, O'Sullivan B, Fishell E, Simor I, Cooke G. Mammographic patterns and breast cancer risk: methodologic standards and contradictory results. J Natl Cancer Inst 1984;72:1253-9.

(*23*) Buchanan JB, Weisberg BF, Sandoz JP, Gray LA Sr, Bland KI. Selected prognostic variables for mammographic parenchymal patterns. Cancer 1981;47:2135-7.

(*24*) Krook PM. Mammographic parenchymal patterns as risk indicators for

incident cancer in a screening program: an extended analysis. AJR Am J Roentgenol 1978;131:1031-5.
(*25*) Ernster VL, Sacks ST, Peterson CA, Schweitzer RJ. Mammographic parenchymal patterns and risk factors for breast cancer. Radiology 1980;134: 617-20.
(*26*) Gravelle IH, Bulstrode JC, Wang DY, Bulbrook RD, Hayward JL. The relation between radiographic features and determinants of risk of breast cancer. Br J Radiol 1980;53:107-13.
(*27*) de Waard F, Rombach JJ, Collette HJ, Slotboom B. Breast cancer risk associated with reproductive factors and breast parenchymal patterns. J Natl Cancer Inst 1984;72:1277-82.
(*28*) Brisson J. Family history of breast cancer, mammographic features of breast tissue, and breast cancer risk. Epidemiology 1991;2:440-4.
(*29*) Breuer B, Miller DG, Salane M, Wolfe JN. Mammographic parenchymal patterns and family history of breast cancer [letter]. Cancer 1992;69:602-3.
(*30*) Chakraborty R, Hanis CL. Non-random sampling in human genetics: estimation of familial correlations, model testing, and interpretation. Stat Med 1987;6:629-46.
(*31*) Bucher KD, Schrott HG. The effect of nonrandom sampling on familial correlations. Biometrics 1982;38:249-53.

Notes

Supported by Public Health Service grants R01CA55747 (National Cancer Institute) and M01RR00400 and S07RR055448 (National Center for Research Resources), National Institutes of Health, Department of Health and Human Services; and by grant IN-13-30-14 from the American Cancer Society. C. M. Vachon is supported in part by Public Health Service training grant T32CA09607 from the National Cancer Institute. The results of this study were obtained by use of the Statistical Analysis for Genetic Epidemiology Program Package, which is supported by Public Health Service grant 1P41RR03655 from the National Center for Research Resources.

Manuscript received August 6, 1996; revised January 16, 1997; accepted February 13, 1997.