CYP17 Promoter Polymorphism and Breast Cancer in Australian Women Under Age Forty Years

Amanda B. Spurdle, John L. Hopper, Gillian S. Dite, Xiaoqing Chen, Jisheng Cui, Margaret R. E. McCredie, Graham G. Giles, Melissa C. Southey, Deon J. Venter, Douglas F. Easton, Georgia Chenevix-Trench

Background: **The cytochrome P450c17α enzyme functions in the steroid biosynthesis pathway, and altered endogenous steroid hormone levels have been reported to be associated with a T to C polymorphism in the 5′ promoter region of the CYP17 gene. Because steroid hormone exposure is known to influence breast cancer risk, we conducted a population-based, case–control-family study to assess the relationship between the CYP17 promoter polymorphism and early-onset breast cancer.** ***Methods:*** **Case subjects under 40 years of age at diagnosis of a first primary breast cancer, population-sampled control subjects, and the relatives of both case and control subjects were interviewed to record family history of breast cancer and other risk factors. CYP17 genotype was determined in 369 case subjects, 284 control subjects, and 91 relatives of case subjects. Genotype distributions were compared by logistic regression, and cumulative risk was estimated by a modified segregation analysis. All statistical tests were two-tailed.** ***Results:*** **Compared with the TT genotype (i.e., individuals homozygous for the T allele), the TC genotype was not associated with increased breast cancer risk ($P = .7$). Compared with the TT and TC genotypes combined, the CC genotype was associated with a relative risk of 1.81 (95% confidence interval [CI] = 1.15–2.86; $P = .01$) before adjustment for measured risk factors and 1.63 (95% CI = 1.00–2.64; $P = .05$) after adjustment. There was an excess of CC genotypes in case subjects who had at least one affected first- or second-degree relative, compared with control subjects unstratified by family history of breast cancer (23% versus 11%; $P = .006$), and these case subjects had a threefold to fourfold higher risk than women of other groups defined by genotype and family history of breast cancer. Analysis of breast cancer in first- and second-degree relatives of case subjects with the CC genotype, excluding two known carriers of a deleterious mutation in BRCA1 or BRCA2, gave a relative hazard in women with the CC genotype of 3.48 (95% CI = 1.13–10.74; $P = .04$), which is equivalent to a cumulative risk of 16% to age 70 years.** ***Conclusions:*** **The CC genotype may modify the effect of other familial risk factors for early-onset breast cancer. [J Natl Cancer Inst 2000;92:1674–81]**

On a population basis, female breast cancer is a familial disease, in that having a first-degree relative with breast cancer is associated with an increased risk of about 1.5-fold to 2.0-fold, on average *(1)*. The increased risk is greater the younger the age at diagnosis of the affected relative. This degree of familial aggregation for disease could exist only if there are substantial underlying familial risk factors *(2–4)*. It could be, in part, a consequence of nongenetic factors shared by relatives, since established lifestyle risk factors identified to date by epidemiologic studies also display a modest degree of familial aggregation. However, mathematical models suggest that such lifestyle factors might explain less than 15% of the familial aggregation of breast cancer, although this may be greater once measurement error and misclassification of these questionnaire-derived surrogates for an underlying hormonal etiology are taken into account *(2)*.

Genetic risk factors for breast cancer include deleterious mutations in the genes BRCA1 and BRCA2 that are associated with a dominantly inherited increased risk of disease of at least 10- to 20-fold in female carriers *(5–7)*. However, the rarity of such carriers (somewhere in the order of one in 100 to one in 1000) means that these genes explain less than 20% of familial aggregation of breast cancer *(8)*. This leaves the possibility that "low-risk" genetic factors, an order of magnitude more common than the "high-risk" mutations in BRCA1 or BRCA2, explain most of the familial aggregation of breast cancer.

Candidates for such low-risk genes include those involved in cancer predisposition pathways, for which there are common but subtle functional variants resulting from genetic polymorphisms. Variants of particular interest would include polymorphisms that affect gene expression or function through modified transcription of DNA, altered stability, processing or translation of messenger RNA, or by amino acid substitution in the expressed protein. Because exposures to endogenous and exogenous steroid hormones are known to influence breast cancer risk, genes in the hormone biosynthesis pathway are currently being considered as candidates for low-risk breast cancer genes.

One such candidate is the CYP17 gene, which encodes the enzyme cytochrome P450c17α that functions at two different points in the steroid biosynthesis pathway. A 5′-promoter-region single-nucleotide T to C substitution polymorphism occurs in women from the United States and Europe with an allele frequency of around 0.3–0.4 and creates an additional Sp1-type promoter site

Affiliations of authors: A. B. Spurdle, X. Chen, G. Chenevix-Trench, Cancer Unit, Joint Experimental Oncology Programme, The Queensland Institute of Medical Research, and The University of Queensland, Brisbane, Australia; J. L. Hopper, G. S. Dite, J. Cui, Centre for Genetic Epidemiology, The University of Melbourne, Carlton, Australia; M. R. E. McCredie, The New South Wales Cancer Council, Kings Cross, Australia, and Department of Preventive and Social Medicine, University of Otago, Dunedin, New Zealand; G. G. Giles, The Anti-Cancer Council of Victoria, Carlton, Australia; M. C. Southey, D. J. Venter, Peter MacCallum Cancer Institute, Melbourne, Australia, and Department of Pathology, The University of Melbourne, Parkville, Australia; D. F. Easton, Cancer Research Campaign Genetic Epidemiology Unit, Strangeways Research Laboratory, Cambridge, U.K.

Correspondence to: Amanda B. Spurdle, Ph.D., Cancer Unit, The Queensland Institute of Medical Research, P.O. Royal Brisbane Hospital, Queensland, 4029, Australia (e-mail: mandyS@qimr.edu.au).

See "Notes" following "References."

(CCACT → CCACC) 34 base pairs (bp) upstream of the initiation of translation but downstream from the transcription start site *(9).* The T allele (CCACT sequence) and the C allele (CCACC sequence) are also reported in the literature as A_1 and A_2 alleles, respectively. Recent *in vitro* data suggest that the 5′ Sp1-type site resulting from the T to C substitution does not actually bind transcription factor Sp1 *(10),* but there is still some evidence to indicate that this polymorphism may influence endogenous steroid hormone levels *(11,12).* Premenopausal nulliparous women with the CC (or A_2A_2) genotype have been shown to have higher mean levels of serum estradiol than those with a TT (or A_1A_1) genotype *(11),* whereas postmenopausal women have been shown to have higher mean levels of serum estrone *(12).* In addition, postmenopausal women with the CC genotype are less likely to be current users of hormone replacement therapy *(13).* There is also a suggestion that the C allele may modify the risk of familial polycystic ovary disease *(9),* and examination of published polycystic ovary syndrome case–control data suggests that, although a dominant effect of the C allele is not apparent *(14,15),* the homozygous CC genotype specifically may be associated with the expression of this disease.

The CYP17 polymorphism has been investigated as a risk factor for female breast cancer, but conflicting results have been obtained. The initial case–control study *(16)* involved African-American, Latino, and Japanese women aged 45–75 years and living in Los Angeles, CA, or in Hawaii. It suggested that inheritance of at least one C allele was associated with an increased (dominantly inherited) risk of advanced breast cancer, based on a subset of 40 women with regional or metastatic breast cancer. However, regardless of whether all cases or advanced cases were considered, the C allele was not associated with risk of breast cancer in numerous subsequent case–control and cohort studies of predominantly postmenopausal women of similar or considerably larger sample size *(10,12,17–20).* An exception is the recent Swedish case–control study *(21)* of 109 women with breast cancer diagnosed before the age of 37 years and 117 control subjects; this study found that the increased risk of such early-onset breast cancer was 2.0-fold for having at least one C allele ($P = .03$) and 2.8-fold for having two C alleles ($P = .06$).

We have undertaken a large population-based, case–control-family study to assess the CYP17 5′ T to C transition polymorphism as a risk factor for breast cancer in Australian women diagnosed before the age of 40 years. The Australian Breast Cancer Family Study has focused on this age group principally because the contribution of familial factors to breast cancer risk is greatest in cases of earlier onset *(22–24).* We have used population registry-sampled case subjects and have compared them with population-sampled control subjects from the electoral rolls. Analyses were also stratified by family history of breast cancer, since, if inheriting one or both copies of a particular variant is truly associated with an increased risk of the disease, then the frequency of this variant should be greater in women with a family history of the disease. To assess if any increased risk associated with the CC genotype was also evident in the female relatives of case subjects, we exploited two important aspects of our design *(22,23)*: 1) Information on cancers in family members was asked of probands *and* of their living relatives, and verification of all reported cancers was sought; and 2) blood samples were collected from participating relatives. We estimated the increased risk and age-specific penetrance of genotypes defined by this polymorphism in the CYP17 gene by a modified segregation analysis used previously for estimating the average penetrance of protein-truncating mutations in BRCA1 and BRCA2 *(5).*

Subjects and Methods

Subjects

A population-based, case–control-family study of early-onset breast cancer was carried out in Melbourne and Sydney, Australia, during the period 1992 through 1995 *(22–24).* Case subjects were women under the age of 40 years at diagnosis of their first primary breast cancer identified though the Victorian and New South Wales cancer registries. Control subjects were women without breast cancer who were selected from the electoral roll (adult registration for voting is compulsory in Australia) with the use of stratified random sampling, frequency matched for age. Subjects were excluded if they could not speak English or if they had been diagnosed previously with breast cancer.

With the permission of case subjects and control subjects, all living parents, aunts, grandparents, and adult siblings were also asked to participate, and these relatives were administered the same risk factor questionnaire as that administered to the case subjects and control subjects *(24).* In particular, ancestry was asked by an open-ended question, and country of birth was asked for the subjects, their parents, and their grandparents. For the great majority of subjects, their parents and grandparents were born in Australia, the British Isles, or Western Europe. For the purpose of the subanalyses restricted to Caucasian subjects only, subjects were excluded if any of the ancestry or country of birth fields mentioned Australian aboriginal, Torres Strait Island or Maori heritage, Asia, South Pacific, Indian Ocean, or Caribbean islands.

For each case subject and control subject, a detailed family history was systematically recorded for all first- and second-degree relatives and was subsequently checked with their living relatives at the time of their interview. Unless otherwise stated, women who reported having at least one first- or second-degree relative with breast cancer were considered to have a family history of breast cancer. Verification of all family cancers reported by subjects and their relatives was sought through personal interview, cancer registries, pathology reports, hospital records, the clinicians who treated subjects, and death certificates.

Of 643 eligible case subjects, 466 (72.5%) were interviewed. Attrition was due to death (n = 11; 1.7%), refusal by the surgeon (n = 54; 8.4%), refusal by the patient (n = 76; 11.8%), nonresponse by the surgeon (n = 4; 0.6%), nonresponse by the patient (n = 9; 1.4%), or having changed residence (n = 23; 3.6%). Of the 633 eligible control subjects, refusals (n = 163; 25.8%) and nonresponse (n = 62; 9.8%) resulted in 408 (64.5%) control subjects being interviewed. For the 466 case subjects and 408 control subjects, details of the measured characteristics including family history, as well as extensive case–control analyses, have been published *(23,24).*

Blood samples were available from 393 case subjects (84.3% of participating and 61.1% of eligible case subjects) and 295 control subjects (72.3% of participating and 46.6% of eligible control subjects). Blood samples were collected from case subjects and control subjects at the time of interview and, depending on the family history of breast cancer, resources, and availability, from participating relatives.

CYP17 genotype was determined on the basis of DNA availability for 369 case subjects (79.2% of participating and 57.4% of eligible case subjects) and 284 control subjects (69.6% of participating and 44.9% of eligible control subjects). Comparing genotyped with nongenotyped subjects showed that selection was independent of factors previously shown to be associated with breast cancer in this study *(24),* except for height (those measured were, on average, 2.3 cm shorter; $P = .02$). Case–control analyses suggested a recessive effect (*see below*). An independent test of that putative effect was conducted by measurement of the CYP17 genotype of all participating relatives of the case subjects found to be of the CC genotype. After exclusion of two families due to non-mendelian inconsistencies in alleged relatives, CYP17 genotype information was available from 91 of these relatives (18 males and 73 females) and could be inferred probabilistically for the other relatives under the assumptions of different genetic models.

To date, a deleterious mutation in either BRCA1 or BRCA2 has been detected in 21 case subjects by protein-truncation testing of all case subjects in specific exons covering about 70% of the coding regions and by manual sequencing of BRCA1 in a

subset [*(5,25);* our unpublished data]. Of these 21 case subjects, eight (38%) had a family history of breast cancer.

Approval of the study protocol was obtained from the ethics committees of The University of Melbourne, the New South Wales Cancer Council, The Anti-Cancer Council of Victoria, and The Queensland Institute of Medical Research. Written informed consent was obtained from each participating subject.

Molecular Analysis

Collection of peripheral blood and DNA extraction were done as described previously *(26)*. The CYP17 5′ C-T polymorphism *(9)* was detected with the use of the ABI Prism 7700 Sequence Detection System. A 102-bp polymerase chain reaction (PCR) product was amplified with the use of the primers GCCTCCTTGTGCCCTAGAGTT and AGCAAGAGAGCCACGAGCTC. *Msp*A1 restriction enzyme digestion and high-resolution agarose gel electrophoresis were used to identify TT and CC homozygote DNA controls required for the Sequence Detection System allelic discrimination assay. With the use of the standard protocol for the Sequence Detection System allelic discrimination assay, fluorescently labeled probes 5′-6-carboxyfluorescein (FAM)-TCTACTCCACCGCTGTCTATCTTGCC-6-carboxy-tetramethyl-rhodamine (TAMRA)-3′ and 5′-tetrachloro-6 carboxyfluorescein (TET)-TTCTACTCCACTGCTGTCTATCTTGCCTG-TAMRA-3′ were utilized to detect the C and T alleles, respectively. Reaction volumes varied between experiments from 25 μL to 15 μL, and the final concentration of reagents in the PCR mix was 1× TaqMan Universal PCR Master Mix (catalog No. 4304437; The Perkin-Elmer Corp., Foster City, CA), 900 n*M* each primer, 200 n*M* FAM-C probe, and 100 n*M* TET-T probe. The reaction mix was added to 30 ng of genomic sample DNA that had been predried in 96-well plates. Reactions were amplified in the ABI Prism 7700 Sequence Detection System PCR machine for 2 minutes at 50 °C, 10 minutes at 95 °C, followed by 40–50 cycles of 15 seconds at 95 °C and 1 minute at 66 °C. Genotype analysis was performed on amplified samples with the use of the ABI PRISM 7700 software by the standard procedures for automated allelic discrimination. In brief, by comparison to the fluorescence signals in known controls (eight each of homozygote allele 1, homozygote allele 2, and no template), the software will call each "unknown" sample as homozygote allele 1, heterozygote, homozygote allele 2, undetermined, or no amplification.

Statistical Methods

Under Hardy–Weinberg equilibrium, the maximum likelihood estimator of the frequency of a particular allele is $f = (2n_{11} + n_{01})/2n$, where $n = n_{11} + n_{01} + n_{00}$ and n_{ij} is the observed number of subjects with the "ij" genotype ($i = 0; j = 1$), where 1 represents the presence of the allele or a group of alleles and 0 represents the absence and has asymptotic standard error (SE) $= [(f[1 - f])/2n]^{1/2}$. The 95% confidence interval (CI) was calculated by $f \pm 1.96$ SE.

The Hardy–Weinberg equilibrium assumption was assessed by the standard method of matching the observed numbers of individuals in the different genotype categories with those expected under Hardy–Weinberg equilibrium for the estimated allele frequency and comparing the Pearson goodness-of-fit statistic with the χ^2 distribution with 1 *df*. Genotype distributions were compared with the use of contingency table analysis. The influence of CYP17 genotype on the risk of breast cancer was assessed, as in standard case–control analyses, by use of unconditional multiple linear logistic regression, with and without adjustment for measured risk factors. The effect of genotype was modeled four ways: 1) by actual number of C alleles (codominant inheritance; a different risk for all three genotypes; two parameters); 2) by a linear effect on the log odds scale for each C allele (multiplicative risk; increase in risk multiplies with each additional C allele; one parameter); 3) by an effect of having any C allele (dominant inheritance; individuals with at least one C allele at increased risk; one parameter); and 4) by an effect of having two C alleles (recessive inheritance; individuals with two C alleles at increased risk; one parameter).

By use of the method described by Hopper et al. *(5),* the cumulative risk of breast cancer in women according to the CYP17 genotype was estimated from a modified segregation analysis of the families of case subjects found to have the CC genotype with the use of the statistical package MENDEL *(27)*. The joint likelihood of each family was expressed as a function of the observed disease status, age at interview, death, or diagnosis, and genotype of family members, conditional on the population-sampled case subject being a known CC homozygote with diagnosis before age 40 years. The relative hazard rate(s)—the risk of breast cancer in women with a defined number of C alleles compared with the risk for women with a baseline genotype appropriate to the inheritance model—was estimated on the logarithmic scale, separately for each of five decades of age, as in Ford et al. *(7)*. The C-allele frequency was assumed to be that found in the control sample, and the relative hazard(s) and baseline rates were constrained so that the average risk, weighted by genotype, was equal to that for the Australian population *(28)*. We fitted this model assuming *(a)* dominant inheritance (i.e., women with one or two C alleles had the same relative hazard), *(b)* codominant inheritance (i.e., the relative hazard differed between women with one C allele compared with those with two C alleles), or *(c)* recessive inheritance (i.e., only women with two C alleles were at increased risk). We also fitted models in which the relative hazard was assumed to be a constant over all ages. The cumulative risk for each genotype was calculated by first deriving the genotype-specific incidence by multiplying the incidence for that baseline genotype for each age by the estimated relative hazard for that age and genotype. The baseline incidence was derived from the population incidence so that the sum of these genotype-specific incidences each multiplied by the population frequency of that genotype is equal to the population incidence. The age-specific cumulative risk for a given genotype was then calculated as 1 − exp(cumulative incidence to that age). The likelihood ratio test was used to compare nested models, and Akaike's information criterion (AIC) *(29)* was used to compare the fits of non-nested models of inheritance. The AIC measure assesses the relative fits of different models by adding a penalty to the maximized log likelihood to reflect the number of parameters tested. It is defined by AIC = −2(maximum log likelihood) + 2(number of parameters estimated), and smaller values reflect a better fitting model. CIs for relative hazard estimates were calculated with the use of the likelihood profile method *(30)*.

All other analyses were performed by use of Stata statistical software *(31)*. All statistical tests and *P* values were two-tailed; following convention, statistical significance was taken as a nominal *P* value of less than .05.

For the codominant inheritance model (*see* Table 2 and Fig. 1), floating SEs and CIs were calculated to help evaluate linear trends in log relative risk with number of C alleles *(17,32,33)*. This method of analysis allows for the construction of SEs and CIs for the parameters of all groups without the need to select a baseline group, and unlike standard methods that compare risk in each group to a baseline group, the SEs and CIs are not dependent on precision within the baseline group *(32)*. We used a simple method for calculating floating SEs and CIs that utilizes standard logistic regression software *(34)*. For the multiplicative risk, dominant inheritance and recessive inheritance models shown in Table 2, CIs were derived for estimates of genotypic risk relative to the appropriate baseline genotype.

Results

Table 1 shows the CYP17 genotypes for case subjects and control subjects, broken down by whether or not they had a family history of breast cancer, defined by any reported breast cancer in a first- or second-degree female relative. There was evidence of deviation from Hardy–Weinberg equilibrium in case subjects with a family history of breast cancer ($\chi^2 = 5.0$; $P = .03$), for whom there was an excess of CC genotypes, namely, 23% compared with 8% in control subjects with a family history of breast cancer ($P = .02$) and 11% in the larger sample of control subjects unstratified by family history of breast cancer ($P = .006$). There was no deviation from Hardy–Weinberg equilibrium in case subjects without a family history ($\chi^2 = 1.1$; $P = .3$) or in control subjects ($\chi^2 = 0.7$; $P = .4$).

The C-allele frequency was 0.347 (95% CI = 0.308–0.386) in control subjects and was independent of family history of breast cancer ($P = .5$). In comparison, it was marginally higher at 0.427 (95% CI = 0.364–0.490) in case subjects with a family history ($P = .03$), but it was no different at 0.377 (95% CI = 0.334–0.420) in case subjects without a family history ($P = .3$). There was no difference in genotype distribution between case subjects with or without a family history of breast cancer ($\chi^2 = 3.0$; $P = .2$) or between control subjects with or without a family history ($\chi^2 = 0.8$; $P = .7$). In a

Table 1. CYP17 5′ promoter polymorphism genotype distribution and allele frequency in case subjects and control subjects, stratified by family history of breast cancer

Genotype	Family history*—yes			Family history*—no			Total		
	Case subjects, No. (%)	Control subjects, No. (%)	Total No. (%)	Case subjects, No. (%)	Control subjects, No. (%)	Total No. (%)	Case subjects, No. (%)	Control subjects, No. (%)	Total No. (%)
TT	44 (38)	32 (43)	76 (40)	102 (40)	86 (41)	188 (41)	146 (40)	118 (42)	264 (40)
TC	45 (39)	36 (49)	81 (43)	111 (44)	99 (47)	210 (45)	156 (42)	135 (48)	291 (45)
CC	27 (23)	6 (8)	33 (17)	40 (16)	25 (12)	65 (14)	67 (18)	31 (11)	98 (15)
Total	116	74	190	253	210	463	369	284	653
C-allele frequency	0.427	0.324	0.387	0.377	0.355	0.367	0.393	0.347	0.373
95% confidence interval	0.364–0.490	0.250–0.398	0.338–0.436	0.334–0.420	0.310–0.378	0.337–0.399	0.358–0.411	0.308–0.386	0.348–0.398

*Family history = reported first- or second-degree female relative with breast cancer. In a comparison of the genotype distribution among subjects stratified by family history of breast cancer, there was no significant difference within case subjects ($P = .2$), within control subjects ($P = .7$), or within all subjects ($P = .5$). When case subjects were compared with control subjects, there was a significant difference between all case subjects and all control subjects ($P = .03$) and between family history-positive case subjects and control subjects ($P = .02$), but there was not a significant difference between family history-negative case subjects and control subjects ($P = .5$). Results were similar when case subjects stratified by family history of breast cancer were compared with the larger sample of all control subjects, with a significant difference for case subjects with a family history ($P = .006$) but not for case subjects without a family history ($P = .2$).

comparison of the genotype distribution in case subjects with that in all control subjects, there was a difference for all case subjects ($\chi^2 = 6.8$; $P = .03$) and for case subjects with a family history of breast cancer ($\chi^2 = 10.3$; $P = .006$) but not for case subjects without a family history ($\chi^2 = 2.9$; $P = .2$).

The above findings were robust to definition of family history of breast cancer (any first- or second-degree relative verified, any first-degree relative reported, or any first-degree verified) and *(a)* when analyses were restricted to the 327 case subjects and 261 control subjects of Caucasian ancestry (as defined in the "Subjects and Methods" section) or *(b)* when the 21 case subjects known to carry a deleterious mutation in either BRCA1 or BRCA2 were excluded. For example, there was still a deviation from Hardy–Weinberg equilibrium in case subjects with a family history of breast cancer [$P = .01$ and .04 for *(a)* and *(b)*, respectively], and the C-allele frequency remained at 0.347 in Caucasian control subjects. Of the 21 mutation-carrying case subjects, 11 (52%) had the TT genotype, eight (38%) had the TC genotype, and two (10%) had the CC genotype. This genotype frequency was no different from case subjects not known to carry a mutation ($P = .4$) or from control subjects ($P = .6$).

Table 2 shows the results of case–control analyses of CYP17 genotype as a risk factor for breast cancer, with adjustment for measured covariates and with stratification by family history of breast cancer. Odds ratios (ORs) adjusted for age and for other covariates did not differ greatly from crude risk estimates. For the total sample, the codominant inheritance model showed that, compared with the baseline TT genotype, the TC genotype was not associated with an increased risk ($P = .7$). The CC genotype, however, was associated with an increased risk ($P = .03$), although this was not strictly significant after adjustment for other risk factors ($P = .09$). The recessive inheritance model (now comparing CC with TC and TT combined) gave a crude OR of 1.81 ($P = .01$) and an adjusted OR of

Table 2. Risk of breast cancer associated with CYP17 genotype according to different models of inheritance*

	Total				Family history of breast cancer—yes		Family history of breast cancer—no	
	Crude OR (95% CI)	*P*	Adjusted OR (95% CI)	*P*	Adjusted OR (95% CI)	*P*	Adjusted OR (95% CI)	*P*
Codominant inheritance								
TT	1.00 (0.78–1.28)		1.00 (0.77–1.30)		1.00 (0.60–1.66)		1.00 (0.73–1.36)	
TC	0.93 (0.74–1.18)	.7	0.93 (0.73–1.19)	.7	0.90 (0.54–1.50)	.8	0.95 (0.78–1.41)	.8
CC	1.75 (1.14–2.76)	.03	1.57 (1.00–2.47)	.09	3.05 (1.17–7.95)	.04	1.24 (0.73–2.12)	.5
Multiplicative risk per C allele	1.21 (0.97–1.51)	.1	1.16 (0.91–1.48)	.2	1.47 (0.92–2.35)	.1	1.07 (0.80–1.43)	.6
Dominant inheritance								
TT	1.00 (referent)		1.00 (referent)		1.00 (referent)		1.00 (referent)	
TC or CC	1.09 (0.79–1.49)	.6	1.05 (0.75–1.48)	.8	1.23 (0.63–2.41)	.5	1.02 (0.69–1.52)	.9
Recessive inheritance								
TT or TC	1.00 (referent)		1.00 (referent)		1.00 (referent)		1.00 (referent)	
CC	1.81 (1.15–2.86)	.01	1.63 (1.00–2.64)	.05	3.27 (1.18–9.08)	.02	1.26 (0.71–2.23)	.8

**P* values are two-tailed. Floating confidence intervals (CIs) were calculated for codominant inheritance only, according to the method of Easton et al. *(32)* and as elaborated under the "Subjects and Methods" section. Odds ratios (ORs) were adjusted for age, country of birth, state, educational level, marital status, number of live births, height, weight, age at menarche, and oral contraceptive use and, for the combined analysis, for family history (reported first- or second-degree relative) of breast cancer. For the total sample, the codominant inheritance model showed that, compared with the baseline TT genotype, the TC genotype was not associated with an increased risk ($P = .7$), whereas the CC genotype was associated with an increased risk ($P = .03$). After stratification by family history, this increased risk associated with the CYP17 CC genotype was evident only for the comparison of case subjects and control subjects who had a family history—the CC genotype was associated with an increased risk for the codominant inheritance model ($P = .04$ adjusted) and the recessive inheritance model ($P = .02$ adjusted).

1.63 ($P = .05$). Neither the multiplicative risk nor the dominant inheritance models gave evidence of a CYP17 effect, and the SEs were about 0.12 on the log OR scale, so that effects equivalent to an OR of $\exp(0.12 \times 2.5) = 1.35$ or more would have been detectable at the .05 level of significance (one-tailed) with more than 80% power. Analysis of subjects with stratification by family history of breast cancer revealed that the increased risk associated with the CYP17 CC genotype was not apparent for subjects without a family history ($P>.5$ for all models) but was evident for the comparison of case subjects and control subjects who had a family history. Although the point estimates were relatively imprecise because of the small sample numbers of case and control subjects with a family history, the CC genotype was associated with an increased risk for the codominant inheritance model ($P = .04$, adjusted) and the recessive inheritance model ($P = .02$, adjusted).

Stratification by genotype revealed no significant effect on breast cancer risk of age at menarche (≥13 years versus <13 years), with an OR (95% CI) of 1.10 (0.66–1.86) ($P = .7$) and 0.66 (0.42–1.03) ($P = .06$) within the TT and pooled TC/CC genotype groups, respectively, and age at menarche effects were independent of genotype (test of interaction, $P = .13$). There was also no difference in the average age at menarche between case and control subjects for the different genotype groups (data not shown).

Given that the CC genotype appears to be associated with risk of breast cancer (Table 2) and the proportion of women with the CC genotype is greater in case subjects with a family history of breast cancer (*see* Table 1), we examined risk by family history. Fig. 1 shows the adjusted ORs, with women with the TT genotype and no family history arbitrarily assigned as the referent group. It can be seen that there is no gradient in risk with number of C alleles in women without a family history of breast cancer (left-hand side of Fig. 1), and the risk associated with the CC genotype appears to be most evident in women with a family history of breast cancer (right-hand side of Fig. 1), for whom it is 3.6 times that of the referent group (floating 95% CI = 1.42–9.11). There was little change in risk estimates for women with the CC genotype who had a family history of breast cancer if *(a)* the analysis was limited to families of Caucasian descent (OR = 4.10; floating 95% CI = 1.42–11.81) or *(b)* the analysis excluded the two case subjects in whom a deleterious BRCA1 or BRCA2 mutation had been detected (OR = 3.69; floating 95% CI = 1.45–9.41). That is, case subjects with the CC genotype *and* a family history of breast cancer were at a threefold to fourfold increased risk compared with other groups defined by genotype and family history.

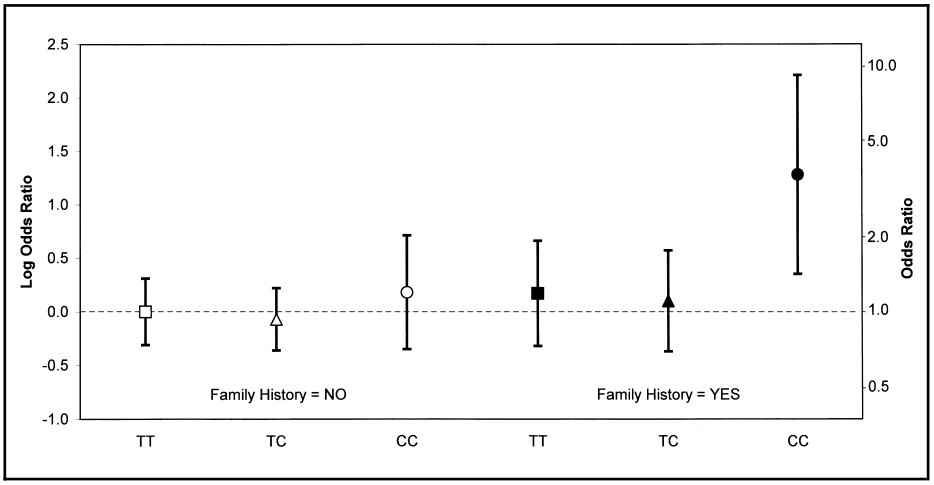

Fig. 1. Association of the CYP17 C allele and family history of breast cancer. The influence of CYP17 genotype on risk of breast cancer was assessed, as in standard case–control analyses, by use of unconditional multiple linear logistic regression, with adjustment for age, country of birth, state, educational level, marital status, number of live births, height, weight, age at menarche, and oral contraceptive use. Women with the TT genotype and no family history of breast cancer were arbitrarily assigned as the referent group. Floating confidence intervals were calculated according to the method of Easton et al. *(32)* and as elaborated on in the "Subjects and Methods" section. **Open and closed symbols** represent individuals without and with, respectively, a family history of breast cancer, where a positive family history is defined as having a first- or second-degree relative with the disease. Values for risk are presented as both log odds ratios (**scale on left-hand side**) and odd ratios (**scale on right-hand side**).

We then made an independent assessment of the above putative risk associated with the CC genotype by studying the occurrence of breast cancer in the female relatives (i.e., sisters, mothers, aunts, and grandmothers) of 65 case subjects with the CC genotype. CYP17 genotype information was available from 91 of these relatives (18 males and 73 females). The proportion of female relatives who were affected was 12% (four of 33) for those with a CC genotype, 12% (three of 26) for those with a TC genotype, 0% (0 of 14) for those with a TT genotype, and 6% (24 of 427) for those with an unknown genotype.

Table 3 shows the results of modified segregation analyses. These analyses use the known genotypes of the case subjects and relatives and the underlying disease model to infer probabilistically the genotypes of the relatives for whom a blood sample was not available. The best fitting model, as judged by the likelihood ratio test and AIC *(29),* was for a recessive risk associated with the CC genotype, irrespective of whether the analysis included the two families in which a deleterious BRCA1 or BRCA2 had been detected in the case subject. The dominant inheritance model gave a smaller log likelihood (−304.69 minus −305.35 = 0.66 for all families, and −283.59 minus −284.67 = 1.08 excluding BRCA1 or BRCA2 families), yet it used the same number of parameters as the recessive inheritance model. The codominant inheritance model used one extra parameter, but the increase in log likelihood compared with the recessive inheritance model was compatible with chance, being 0.00 (−304.69 minus −304.69) and 0.15 (−283.44 minus −283.59), respectively ($P = 1.0$ and .7). For the recessive inheritance model, the hazard for women with the CC genotype compared with women with the TT or TC genotype was 3.4 times higher ($P = .04$), equivalent to a cumulative risk to age 70 years of 15% (compared with 7.5% in the Australian population). These results were little different when the analysis excluded the two families carrying a BRCA1 or BRCA2 mutation; the hazard for the CC genotype was 3.5 times higher ($P = .04$), equivalent to a cumulative risk to age 70 years of 16%. That is, the female relatives of a woman with a CC genotype who had been diagnosed with breast cancer before the age of 40 years

Table 3. Segregation analysis of case subjects with the CC genotype and their families

Model	Relative hazard (95% CI)*	Cumulative risk to age 70 y, %	Log likelihood	P†,‡	AIC‡,§
All families (n = 65)					
Codominant inheritance					
TC versus TT	1.97 (0.34–9.95)	6	−304.69	.11	613.38
CC versus TT	5.05 (1.28–19.91)	19			
Dominant inheritance:					
TC/CC versus TT	4.36 (0.89–17.10)	9	−305.35	.08	612.71
Recessive inheritance:					
CC versus TC/TT	3.42 (1.14–10.28)	15	−304.69	.04	611.39
Excluding families carrying a BRCA1 or BRCA2 mutation (n = 63)					
Codominant inheritance					
TC versus TT	1.08 (0.13–8.73)	5	−283.44	.10	570.88
CC versus TT	4.66 (1.15–18.85)	18			
Dominant inheritance:					
TC/CC versus TT	3.21 (0.77–13.45)	9	−284.67	.14	571.34
Recessive inheritance:					
CC versus TC/TT	3.48 (1.13–10.74)	16	−283.59	.04	569.18

*CI = confidence interval. Hazard for at risk genotype relative to baseline genotype. Actual genotype depends on inheritance model. For codominant inheritance, the relative hazard differs between women with one C allele compared with those with two C alleles, with TT genotype as baseline; for dominant inheritance, women with one or two C alleles have the same relative hazard, with TT baseline; and for recessive inheritance, only women with two C alleles are considered to be at increased risk, with combined TC/TT genotypes as baseline. Cumulative risk to age 70 years was calculated as elaborated in the "Subjects and Methods" section.

†*P* value was derived by the likelihood ratio test compared with the null model of no effect of genotype.

‡The best fitting model was for a recessive risk associated with the CC genotype, with this model yielding the smallest *P* value for the log likelihood ratio test and the smallest value derived by AIC.

§Akaike's information criterion (AIC) compares fits of different models of inheritance by adding a penalty to the maximized log likelihood to reflect the number of parameters tested.

were at increased risk of breast cancer themselves if they also had the CC genotype.

Discussion

Case–control analysis of our data suggested that women with the CYP17 CC genotype are at an increased risk of early-onset breast cancer (Table 2). This recessively inherited risk appears to be most evident in women who have a family history of the disease, in whom the excess of homozygotes for the C allele is more than expected under Hardy–Weinberg equilibrium (Fig. 1, Table 1). An independent confirmation of this genetic risk came from a modified segregation analysis of the families of the case subjects with the CC genotype, which suggested that female relatives with the CC genotype were at an increased risk compared with those with the TC or the TT genotype (Table 3). These findings imply that the CC genotype may modify the effect of other familial risk factors for breast cancer. Although we have not genotyped all case subjects and control subjects, there was no evidence of selection differences between genotyped and nongenotyped subjects on any of the important measured risk factors, and we adjusted for risk factors in the analyses.

Our findings are somewhat in contrast to most previous published analyses of CYP17 and risk of breast cancer, which have failed to find convincing evidence for an association. A meta-analysis *(35)* of data published before July 1998, limited by publication of some incomplete data through considering dominant inheritance only, found a pooled risk estimate of 1.10 (95% CI = 0.93–1.30). Three issues, however, need to be considered: 1) the age at onset, 2) the involvement of family history of breast cancer, and 3) the mode of inheritance.

1) Age at onset. Previously published studies have been of predominantly postmenopausal women. An exception was a Swedish study *(21)* of 109 case subjects and 117 control subjects that, as in this report, investigated women under the age of 40 years. That study found marginal evidence of increased risks of about twofold in both women with the TC genotype and women with the CC genotype compared with women with the TT genotype. We have identified published complete genotype data on a further 151 affected women who were either premenopausal or under the age of 45 years at diagnosis, as well as 75 unaffected women of the same age or menopausal status *(10,18,21).*Combining those data with ours, compared with the TT genotype, the crude OR associated with the TC genotype was 1.09 (95% CI = 0.84–1.43) ($P = .5$), whereas it was 1.59 (95% CI = 1.08–2.35) ($P = .02$) for the CC genotype. That is, the pooled data are not inconsistent with a recessively inherited risk for early-onset disease, although the majority of the pooled data came from the current study.

2) Involvement of family history of breast cancer. Apart from the current report, to our knowledge, there appears to be only one published study stratifying risk according to family history of breast cancer *(18),* and three quarters of the subjects in that study were postmenopausal (average age, 60 years). Helzlsouer et al. *(18)* defined family history of breast cancer as having an affected mother, sister, daughter, or grandmother and found no difference by family history in the distribution of CYP17 genotype in just 31 case subjects with and 78 without a family history.

The strength of our study is that the family data, based on reports by both case subjects and their relatives, have been used in two essentially independent ways. First, comparison of case subjects with control subjects showed that the frequency of the CC genotype was higher in case subjects with a family history of breast cancer (χ^2 test; $P = .006$). If that genotype is truly associated with an increased risk of breast cancer, then their relatives will be more likely to have the disease in part *because* the relatives are more likely to share that genetic risk. Second, studying these relatives as individuals provided statistical confirmation that those having, or likely to have, the CC genotype were at an increased risk of breast cancer. The CYP17 genotype of relatives was determined either directly or inferred probabilistically from analysis of the DNA of their relatives, so that all of the family data could be used in the maximum likelihood analysis. Another strength of having used relatives is that it may have reduced problems of stratification bias that can be encountered in association studies in nonhomogeneous populations.

The higher point estimate of relative risk for CC homozygotes seen in the segregation analysis than in the case–control analysis could be due to chance or to the

fact that close relatives of index cases are at increased risk of breast cancer by virtue of other genetic or familial risk factors *(23,24).* The latter may have led to a larger estimated relative risk if CC homozygotes were overrepresented in close relatives. A more complex analysis to adjust for this effect, incorporating the effects of other familial factors in addition to CYP17, BRCA1, and BRCA2, is now under way.

3) Mode of inheritance. There is some support for a recessive effect from measures of biologic variables. The mean levels of serum estrogen metabolites tend to increase with the number of C alleles; however, in some measures, the effect is nonlinear because of a considerably higher value in women with the CC genotype [e.g., *see* Table 2 of Feigelson et al. *(11);* Table 6 of Haiman et al. *(12)*]. In terms of breast cancer risk, some previous studies have pooled the TC and CC genotypes in their data presentation, making it difficult to consider evidence for a recessively inherited risk. The meta-analysis by Dunning et al. *(35)* did not consider recessive inheritance. In the three largest studies *(10,12,17),* containing mostly postmenopausal women, the point estimate of the relative risk for the CC genotype was always numerically greater than that for the TC genotype, but none of these homozygote versus heterozygote differences were statistically significant, and even the point estimates themselves were not statistically different from unity. That is, although our data and the pooled data on premenopausal women mentioned above appear to be more compatible with recessive inheritance than with dominant inheritance, pooling of large studies is required to resolve this issue with any certainty.

In conclusion, our data suggest that the CYP17 CC genotype may be implicated in early-onset familial breast cancer. We will attempt to replicate this finding by conducting similar genetic analyses in an independent case–control-family study of breast cancer in women under 40 years of age *(22).*

REFERENCES

(1) Pharoah PD, Day NE, Duffy S, Easton DF, Ponder BA. Family history and the risk of breast cancer: a systematic review and meta-analysis. Int J Cancer 1997;71:800–9.

(2) Hopper JL, Carlin JB. Familial aggregation of a disease consequent upon correlation between relatives in a risk factor measured on a continuous scale. Am J Epidemiol 1992;136: 1138–47.

(3) Easton D, Peto J. The contribution of inherited predisposition to cancer incidence. Cancer Surv 1990;9:395–416.

(4) Aalen OO. Modelling the influence of risk factors on familial aggregation of disease. Biometrics 1991;47:933–46.

(5) Hopper JL, Southey MC, Dite GS, Jolley DJ, Giles GG, McCredie MR, et al. Population-based estimate of the average age-specific cumulative risk of breast cancer for a defined set of protein-truncating mutations in BRCA1 and BRCA2. Australian Breast Cancer Family Study. Cancer Epidemiol Biomarkers Prev 1999;8:741–7.

(6) Narod SA, Ford D, Devilee P, Barkardottir RB, Lynch HT, Smith SA, et al. An evaluation of genetic heterogeneity in 145 breast–ovarian cancer families. Breast Cancer Linkage Consortium. Am J Hum Genet 1995;56:254–64.

(7) Ford D, Easton DF, Stratton M, Narod S, Golgar D, Devilee P, et al. Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium. Am J Hum Genet 1998;62:676–89.

(8) Peto J, Collins N, Barfoot R, Seal S, Warren W, Rahman N, et al. Prevalence of BRCA1 and BRCA2 gene mutations in patients with early-onset breast cancer. J Natl Cancer Inst 1999; 91:943–9.

(9) Carey AH, Waterworth D, Patel K, White D, Little J, Novelli P, et al. Polycystic ovaries and premature male pattern baldness are associated with one allele of the steroid metabolism gene CYP17. Hum Mol Genet 1994;3:1873–6

(10) Nedelcheva Kristensen V, Haraldsen EK, Anderson KB, Lonning PE, Erikstein B, Karesen R, et al. CYP17 and breast cancer risk: the polymorphism in the 5′ flanking area of the gene does not influence binding to Sp-1. Cancer Res 1999;59:2825–8.

(11) Feigelson HS, Shames LS, Pike MC, Coetzee GA, Stanczyk FZ, Henderson BE. Cytochrome P450c17α gene (CYP17) polymorphism is associated with serum estrogen and progesterone concentrations. Cancer Res 1998;58: 585–7.

(12) Haiman CA, Hankinson SE, Spiegelman D, Colditz GA, Willett WC, Speizer FE, et al. The relationship between a polymorphism in CYP17 with plasma hormone levels and breast cancer. Cancer Res 1999;59:1015–20.

(13) Feigelson HS, McKean-Cowdin R, Pike MC, Coetzee GA, Kolonel LN, Nomura AM, et al. Cytochrome p450c17alpha Gene (CYP17) polymorphism predicts use of hormone replacement therapy. Cancer Res 1999;59: 3908–10.

(14) Gharani N, Waterworth DM, Williamson R, Franks S. 5′ polymorphism of the CYP17 gene is not associated with serum testosterone levels in women with polycystic ovaries [letter]. J Clin Endocrinol Metab 1996;81:4174.

(15) Diamanti-Kandarakis E, Bartzis MI, Zapanti ED, Spina GG, Filandra FA, Tsianateli TC, et al. Polymorphism T→C (−34 bp) of gene promoter in Greek patients with polycystic ovary syndrome. Fertil Steril 1999;71:431–5.

(16) Feigelson HS, Coetzee GA, Kolonel LN, Ross RK, Henderson BE. A polymorphism in the CYP17 gene increases the risk of breast cancer. Cancer Res 1997;57:1063–5.

(17) Dunning AM, Healey CS, Pharoah PD, Foster NA, Lipscombe JM, Redman KL, et al. No association between a polymorphism in the steroid metabolism gene CYP17 and risk of breast cancer. Br J Cancer 1998;77:2045–7.

(18) Helzlsouer KJ, Huang HY, Strickland PT, Hoffman S, Alberg AJ, Comstock GW, et al. Association between CYP17 polymorphisms and the development of breast cancer. Cancer Epidemiol Biomarkers Prev 1998;7:945–9.

(19) Weston A, Pan CF, Bleiweiss IJ, Ksieski HB, Roy N, Maloney N, et al. CYP17 genotype and breast cancer risk. Cancer Epidemiol Biomarkers Prev 1998;7:941–4.

(20) Huang CS, Chern HD, Chang KJ, Cheng CW, Hsu SM, Shen CY. Breast cancer risk associated with genotype polymorphism of the estrogen-metabolizing genes CYP17, CYP1A1, and COMT: a multigenic study on cancer susceptibility. Cancer Res 1999;59:4870–5.

(21) Bergman-Jungestrom M, Gentile M, Lundin AC, Wingren S. Association between CYP17 gene polymorphism and risk of breast cancer in young women. Int J Cancer 1999;84: 350–3.

(22) Hopper JL, Giles GG, McCredie MR, Boyle P. Background, rationale and protocol for a case–control-family study of breast cancer. Breast 1994;3:79–86.

(23) Hopper JL, Chenevix-Trench G, Jolley DJ, Dite GS, Jenkins MA, Venter DJ, et al. Design and analysis issues in a population-based, case–control-family study of the genetic epidemiology of breast cancer and the Co-operative Family Registry for Breast Cancer Studies (CFRBCS). J Natl Cancer Inst Monogr 1999; 26:95–100.

(24) McCredie MR, Dite G, Giles GG, Hopper JL. Breast cancer in Australian women under the age of 40. Cancer Causes Control 1998;9: 189–98.

(25) Southey MC, Tesoriero AA, Andersen CR, Jennings KM, Brown SM, Dite GS, et al. BRCA1 mutations and other sequence variants in a population-based sample of Australian women with breast cancer. Br J Cancer 1999; 79:34–9.

(26) Southey MC, Batten LE, McCredie MR, Giles GG, Dite G, Hopper JL, et al. Estrogen receptor polymorphism at codon 325 and risk of breast cancer in women before age forty. J Natl Cancer Inst 1998;90:532–6.

(27) Lange K, Boehnke M, Weeks D. Programs for pedigree analysis. Los Angeles (CA): Department of Biomathematics, University of California at Los Angeles; 1987.

(28) Jelfs P, Coates M, Giles G, Shugg D, Threlfall T, Roder D, et al. Cancer in Australia 1989–1990, Australian Institute of Health and Welfare, Cancer Series No. 5. Canberra (Australia): Australian Government Publishing Service; 1996. p. 149.

(29) Akaike H. Information theory and an extension of the maximum likelihood principle. In: Petrov EB, Csaki F, editors. 2nd International Symposium of Information Theory and Con-

trol. Budapest (Hungary): Akademia Kiado; 1973. p. 276–81.
(30) Clayton D, Hills M. Statistical methods in epidemiology. Oxford (U.K.): Oxford University Press; 1993. Chapter 3.
(31) StataCorp. Stata statistical software: release 5.0. College Station (TX): Stata Corp.; 1997.
(32) Easton DF, Peto J, Babiker AG. Floating absolute risk: an alternative to relative risk in survival and case–control analysis avoiding an arbitrary reference group. Stat Med 1991;10: 1025–35.
(33) Greenland S, Michels KB, Robins JM, Poole C, Willett WC. Presenting statistical uncertainty in trends and dose–response relations. Am J Epidemiol 1999;149:1077–86.
(34) Hopper JL, Dite GS. Re: Presenting statistical uncertainty in trends and dose–response relationships. Am J Epidemiol. In press 2000.
(35) Dunning AM, Healey CS, Pharoah PD, Teare MD, Ponder BA, Easton DF. A systematic review of genetic polymorphisms and breast cancer risk. Cancer Epidemiol Biomarkers Prev 1999;8:843–54.

NOTES

Supported by the National Health and Medical Research Council of Australia, the Victorian Health Promotion Foundation, the New South Wales Cancer Council, and the Peter MacCallum Cancer Institute. M. R. E. McCredie is funded in New Zealand by the Inkster-Ross Memorial Fund.

We thank Antonis Antoniou for help with computing analysis. We are grateful to physicians, surgeons, and oncologists in Victoria and New South Wales who endorsed this project; to the interviewing staff; and to the many women and their relatives who participated in this research.

Manuscript received February 29, 2000; revised August 11, 2000; accepted August 15, 2000.