

Genetic Linkage Analysis in Familial Breast and Ovarian Cancer: Results from 214 Families

D. F. Easton,* D. T. Bishop,† D. Ford,* G. P. Crockford,† and the Breast Cancer Linkage Consortium¹

*Cancer Research Campaign and Institute of Cancer Research, Section of Epidemiology, Belmont, Surrey, England; and †Imperial Cancer Research Fund, Genetic Epidemiology Laboratory, St. James's Hospital, Leeds

Summary

Breast cancer is known to have an inherited component, consistent in some families with autosomal dominant inheritance; in such families the disease often occurs in association with ovarian cancer. Previous genetic linkage studies have established that in some such families disease occurrence is linked to markers on chromosome 17q. This paper reports the results of a collaborative linkage study involving 214 breast cancer families, including 57 breast-ovarian cancer families; this represents almost all the known families with 17q linkage data. Six markers on 17q, spanning approximately 30 cM, were typed in the families. The aims of the study were to define more precisely the localization of the disease gene, the extent of genetic heterogeneity and the characteristics of linked families and to estimate the penetrance of the 17q gene. Under the assumption of no genetic heterogeneity, the strongest linkage evidence was obtained with D17S588 (maximum LOD score [Z_{\max}] = 21.68 at female recombination fraction [θ_f] = .13) and D17S579 (Z_{\max} = 13.02 at θ_f = .16). Multipoint linkage analysis allowing for genetic heterogeneity provided evidence that the predisposing gene lies between the markers D17S588 and D17S250, an interval whose genetic length is estimated to be 8.3 cM in males and 18.0 cM in females. This position was supported over other intervals by odds of 66:1. The location of the gene with respect to D17S579 could not be determined unequivocally. Under the genetic model used in the analysis, the best estimate of the proportion of linked breast-ovarian cancer families was 1.0 (lower LOD - 1 limit 0.79). In contrast, there was significant evidence of genetic heterogeneity among the families without ovarian cancer, with an estimated 45% being linked. These results suggest that a gene(s) on chromosome 17q accounts for the majority of families in which both early-onset breast cancer and ovarian cancer occur but that other genes predisposing to breast cancer exist. By examining the fit of the linkage data to different penetrance functions, the cumulative risk associated with the 17q gene was estimated to be 59% by age 50 years and 82% by age 70 years. The corresponding estimates for the breast-ovary families were 67% and 76%, and those for the families without ovarian cancer were 49% and 90%; these penetrance functions did not differ significantly from one another.

Received June 15, 1992; revision received October 30, 1992.

Address for correspondence and reprints: D. Timothy Bishop, ICRF Genetic Epidemiology Laboratory, Ashley Wing, St. James's Hospital, Beckett Street, Leeds LS9 7TF, England.

1. Collaborating groups' abbreviations, used throughout this report, and members of the Breast Cancer Linkage Consortium are as follows (for a full list of affiliations, see accompanying papers): Aberdeen = University of Aberdeen, Aberdeen—N. Haites, B. Milner, and L. Allan; Berkeley = University of California, Berkeley—M.-C. King, A. Bowcock, and L. Anderson; CRC = Cancer Research Campaign and Institute of Cancer Research, London and Cambridge—D. F. Easton, B. A. J. Ponder, J. Peto, S. Smith, K. Anderson, D. Ford, and M. Stratton; FCN = French Cooperative Network, Lyon, Paris, and Clermont-Ferrand—H. Sobol, S. Mazoyer, D. Lyonnet, Y. J. Bignon, and P. Lalle; IARC = International Agency for Research on Cancer, Lyon—S. A. Narod, G. M. Lenoir, J. Feunteun, and H.

Lynch; Iceland = University of Iceland, Reykjavik—A. Arason, R. Barkardottir, and V. Egilsson; ICRF = Imperial Cancer Research Fund, Leeds and London—D. T. Bishop, D. M. Black, D. Kelsell, V. A. Murday, E. Solomon, N. Spurr, and G. Turner; Leiden = Leiden University, Leiden—P. Devilee; Manchester = Christie Hospital and Holt Radium Institute, Manchester—J. M. Birch, M. S. Santibáñez-Koref, and M. D. Teare; MDC = Max-Delbrück-Centrum für Molekulare Medizin, Berlin—S. Scherneck and W. Zimmerman; MRC = Medical Research Council Human Genetics Unit, Edinburgh—M. Steel, D. Porter, B. B. Cohen, and A. Carothers; Stockholm = Karolinska Hospital, Stockholm—A. Lindblöm and C. Larsson; and Utah = University of Utah, Salt Lake City—L. A. Cannon-Albright, D. Goldgar, and M. Skolnick.

© 1993 by The American Society of Human Genetics. All rights reserved.
0002-9297/93/5204-0005\$02.00

Introduction

A family history of breast cancer is a now well-established risk factor for the disease (Kelsey 1979). Many previous epidemiological studies have shown a significantly increased risk of breast cancer in the relatives of breast cancer cases as compared with the general population (e.g., see Adami et al. 1980; Claus et al. 1990). With some exceptions (e.g., Mettlin et al. 1990), these studies generally show that the risk of breast cancer in relatives increases with decreasing age of the index case and is greater for women with several affected relatives than for women with one affected relative (Claus et al. 1990, 1991). In their analysis of the CASH study, based on the families of over 4,500 breast cancer cases and matched controls, Claus et al. (1991) found that familial clustering of breast cancer could be best explained by a model in which there are one or more autosomal dominant predisposing genes conferring a high risk of the disease; germ-line mutations in these genes must, however, only account for a minority of cases of the disease, the remaining cases occurring sporadically; but the proportion of cases occurring in predisposed individuals increases with decreasing age at onset. Similar models have been obtained by several other authors (Williams and Anderson 1984; Bishop et al. 1988; Newman et al. 1988; Iselius et al. 1991).

Ovarian cancer is also known to have a familial component, and there is evidence that this is also at least partly the result of an autosomal dominant gene(s) with high penetrance (Schildkraut and Thompson 1988; D. F. Easton, unpublished data). Moreover, the risk of breast cancer is increased in ovarian cancer relatives, and vice versa (Schildkraut et al. 1989; D. F. Easton, F. Matthews, A. J. Swerdlow, and J. Peto, unpublished data). This suggests the existence of a gene(s) predisposing to both breast cancer and ovarian cancer. In addition, many impressive pedigrees supporting the existence of such dominant highly penetrant breast and/or ovarian cancer genes have been published (e.g., see Gardner and Stephens 1950; Fraumeni et al. 1975; Lynch et al. 1978).

The first convincing localization of a breast cancer gene by genetic linkage was achieved by Hall et al. (1990), who obtained a LOD score of 2.35 at $\theta = .20$ with D17S74 located in chromosomal region 17q21. There was significant evidence of heterogeneity, with an estimated 40% of families being linked. Analysis by age showed that linkage evidence was clearest in those families with earliest age at onset (Mérette et al. 1992). Among families with an age at onset of less than 46

years, the maximum LOD score was 5.98 at $\theta = .001$ (Hall et al. 1990). Families with later onset showed evidence against linkage. However, subsequent reanalysis (using analytical methods similar to those in the present paper) has shown that the evidence against linkage in later-onset families disappears if sporadic cases are allowed for appropriately, since these families contain a high proportion of sporadic cases (Margaritte et al. 1992).

Confirmation of this linkage result was provided by Narod et al. (1991), who obtained a LOD score of 2.20 at $\theta = .20$ with D17S74 in five breast-ovarian cancer families. That study also found significant evidence of genetic heterogeneity, linkage being apparently restricted to three of the five families tested. There was no apparent difference in average age at onset between linked and unlinked families in that study. This breast-ovarian cancer locus has been formally labeled "BRCA1" (Solomon and Ledbetter 1991).

The aims of this collaborative study were fourfold: first, to attempt to confirm the previously published linkage results on a much larger set of breast cancer families; second, to localize the gene to a particular chromosomal interval, by using several markers spanning the region of interest; third, to define the extent of genetic heterogeneity and to attempt to characterize linked and unlinked families; and, fourth, to use the marker information to provide an estimate of the penetrance of the 17q gene.

In order to provide as unbiased an assessment as possible of the extent of heterogeneity, the aim was to encourage participation by as many groups as possible, worldwide, who are conducting breast cancer linkage studies. Groups were asked to submit linkage information on all families studied, regardless of the strength of their evidence for or against linkage. The majority of the groups have been collaborating over the past 3 years by participating in biennial meetings at which linkage results have been shared. When the linkage to D17S74 was reported by Hall et al. (1990) and replicated by Narod et al. (1991), the groups all typed D17S74. Heterogeneity analysis for the combined set of families was significant, but there was no resolution of the location of the breast cancer locus (data not shown). In brief, the data were consistent with, at one extreme, the disease being tightly linked to D17S74 in 20% of the families and, at the other extreme, 90% percent being linked but with the disease gene 20 cM from D17S74.

To resolve this and to facilitate a joint multipoint linkage analysis, a common set of highly polymorphic markers was recommended for typing across the family

set. Initially, these markers were chosen to be up to 20 cM either side of D17S74, but early results suggested that only markers proximal to D17S74 were relevant. A modified set of six markers focusing on proximal locations was drawn up, and we report here on the results of these analyses.

Families, Material, and Methods

Families

Thirteen groups contributed a total of 214 families to this study. Accompanying papers by the collaborating groups describe in detail the methods used for selecting families, data collection, and confirmation of diagnoses. Criteria for family selection differed between groups. Two of the groups (Utah and Iceland) ascertained their families by using record linkage of population genealogies to cancer registries (Arason et al. 1993; Goldgar et al. 1993). Most of the remaining families were ascertained through collaborating physicians. The majority of families were ascertained predominantly through a family history of breast cancer, though one group (CRC) selected some families on the basis of ovarian cancers in the family (Smith et al. 1993)

while another group (IARC) selected most of their families for both breast and ovarian cancer (Feunteun et al. 1993). One family was ascertained and typed independently by two groups (reference numbers IARC 2775 and Utah 1901). All the analyses for this family are based on the IARC data.

In 10 of the families one breast cancer occurred in an individual on the opposite side of the family from the remaining cancers; these breast cancer cases are ignored in the summaries of numbers and ages of cases in the families (but not in the linkage analysis). In one family (see Arason et al. 1993) there were three breast cancer cases on the opposite side of the family. In the remaining 203 families all cases of breast and ovarian cancer occurred in a pattern consistent with segregation of a single rare autosomal dominant gene.

Table 1 summarizes basic details about the families studied. All the families except three contained at least two breast cancer cases (one family contained four ovarian cancers and no breast cancers, one contained three ovarian cancers and one breast cancer, and one contained seven ovarian cancers and one breast cancer). One hundred twenty-four (58%) contained at least four cases, and 77 (36%) contained at least three cases under

Table 1

Summary of Types of Family in the Study

GROUP ^a	NO. OF FAMILIES									NO. OF PEDIGREE MEMBERS						
	Total	Average Age at Diagnosis of Breast Cancer (years)			No. of Ovarian Cancers				Male Breast Cancers	Total	Breast Cancer			Ovarian Cancer	Typed	
		<45	45-60	≥60	0	1	2	>2			Total	Bilateral	Diagnosis at <45 Years		Total	Affected
Aberdeen	1	1						1		50	4		3	8	28	4
Berkeley	23	5	17	1	20	2		1	2	356	141	7	56	7	256	117
CRC	43	23	19	1	31	2	2	8	1	599	192	28	99	49	245	153
IARC	19	13	6		4	7	2	6		570	119	19	76	42	356	85
Iceland	7	2	5		3	3		1	1	261	48	7	20	7	156	37
ICRF	16	9	7		12	2	2			330	64	12	34	6	157	45
Leiden	13	5	6	2	11		2			353	61	2	23	4	246	39
FCN	19	9	10		14	3	1	1		303	76	8	35	8	192	69
MDC	9	6	3		6	3				92	30		16	3	62	25
MRC	15	5	9	1	10	2	1	2	1	426	72	9	34	10	239	47
Manchester	7	2	3	2	6	1				59	26	3	8	1	25	16
Stockholm	29	5	13	11	29					205	82	6	15		89	68
Utah	13	8	4	1	11	1		1		351	72		39	8	209	44
Total	214	93	102	19	157	26	10	21	5	3,955	987	101	458	153	2,260	749

^a One family was independently identified by both the IARC group and the Utah group. In this table, the family is only listed under the IARC group, as the latter had sampled more extensive branches of the pedigree.

the age of 45 years. Consequently, in light of the known epidemiological data, few of the families could be chance aggregations. Fifty-seven (27%) of the families contained at least one case of ovarian cancer (These families are referred to hereafter as “breast-ovary families.”), and 31 (14%) contained two cases. Of the remaining families, four contained one or more cases of male breast cancer and are referred to as “male breast cancer families.” (One of the breast-ovary families also contained one case of male breast cancer.) The remaining 153 families which contained only female breast cancer are referred to as “breast-only families.” Seven hundred forty-nine (66%) of the 1,140 breast and ovarian cancer cases in the families were typed for at least one marker.

The numbers in table 1 summarize the status of the families as recorded at the time of this analysis (in March 1992). Some collaborating groups have accumulated further information since this date, and the totals in the accompanying papers therefore differ slightly from those given in table 1.

There are clear differences between the ages at onset of the breast and ovarian cancers in the families ascertained. Among the diagnoses of breast and ovarian cancers, 4% of those made before the age of 30 years are ovarian cancer, compared with 11% of diagnoses made between 30 and 59 years of age and 15% of those made at age 60 years or older. For the CRC and IARC families which focused attention for ascertainment on both ovarian and early-onset breast cancer, the corresponding figures were 7%, 18%, and 25%.

Markers

Six polymorphic markers (GH, D17S74, NME1, D17S588, D17S579, and D17S250) were typed in the families, although not all of the families were typed for all of the markers. All of these markers were CA-repeat polymorphisms, with the exception of the VNTR cMM86 (D17S74) used in the original reports of linkage. These markers were chosen because of both their known high levels of polymorphism and their locations. Details on the markers, as well as on others subsequently used by consortium members, are summarized in the Appendix, which is supplied by Dr. Nigel Spurr. The laboratory methods used by individual groups to type these markers are described in the accompanying papers. In total, 194 families (91%) were typed for D17S588, 189 (88%) for D17S74, 189 (88%) for D17S579, 172 (80%) for D17S250, 163 (76%) for NME1, and 141 (66%) for GH.

Marker Allele Frequencies

One disadvantage of using microsatellite and VNTR markers was that consistent scoring of marker alleles, between groups and even between families being studied by the same group, could not be guaranteed. This is an important issue, since misspecifying allele frequencies can lead to exaggerated LOD scores and, in exceptional cases, to a biased recombination fraction estimate (Ott 1992). Some of the groups did attempt to score D17S588 and NME1 consistently, but we found discrepancies in the sizes of the most common alleles reported. For these reasons, our analyses assume equal allele frequencies for each marker. The markers GH and D17S74 are known to be highly polymorphic (some groups scored 12 or more distinct alleles within some families), so we assumed that all alleles had frequency .05. The remaining markers were less polymorphic, and we chose allele frequencies to reflect their observed polymorphism. For D17S579 we assumed 10 equally frequent alleles, for D17S250 and D17S588 6 alleles, and for NME1 5 alleles. In some families, more alleles were recognized, and recoding of the marker alleles within such families was necessary (see Statistical Methods section, below). This solution is clearly not ideal, but we made these assumptions to avoid the greatly exaggerated LOD scores which would result from underestimating the frequency of an allele observed in two distantly related affected individuals. As a check on the robustness of the results to misspecification of the allele frequencies, we also repeated the two-point analyses for D17S588 and D17S579, assuming different allele frequencies. Reducing the allele frequencies increased the maximum LOD scores, but the recombination estimates remained essentially unchanged.

Statistical Methods

The two-point and multipoint linkage analyses were carried out using the LINKAGE program (Lathrop and Lalouel 1984; Lathrop et al. 1984). As a standard genetic model for breast cancer, for use in the linkage analysis, we used the model derived by Claus et al. (1991) from the Cancer and Steroid Hormone Study. In this model, breast cancer susceptibility is conferred by an autosomal dominant allele, with population frequency .0033, such that the breast cancer risk is 67% by age 70 years in carriers and is 5% in noncarriers. A minor modification to this model was made, in order to remove the large increase in risk to gene carriers over the age of 70 years as compared with the risk at age 60–69 years. In our model, the age-specific risk of

breast cancer after age 70 years in gene carriers was taken to be the same as the risk at age 60–69 years. This model is referred to as the “CASH” model throughout. To implement this model, individuals were assigned to one of seven age groups (less than 30 years, 30–39 years, 40–49 years, 50–59 years, 60–69 years, 70–79 years, and 80 or more years), on the basis of either the age at diagnosis of breast cancer, if affected, or the age at death or last observation, if unaffected. Because the disease status is censored at the age of last observation for unaffected individuals, the likelihood must be calculated according to survival analysis methods (Elston 1973). This was achieved by assigning affected and unaffected individuals to different liability classes, so that 14 liability classes (7 age groups \times 2 disease classifications) were used in total. The “penetrance” probabilities assigned to these classes were the genotype-specific cumulative risk, for unaffecteds, and the genotype-specific density, for affecteds, appropriate to each age group (with the values at ages 25, 35, 45, 55, 65, 75 and 85 years being used, assuming a constant incidence rate over the decade). These values are shown in table 2.

Since male breast cancer is so rare in the general population, male cases in these families are likely to be the result of genetic predisposition. We therefore assigned

affected males to the “female affected before age 30” liability class, to maximize their probability of being carriers. Unaffected males were assigned to the “female unaffected at age 30” liability class, approximately equivalent to being of unknown disease status.

Ovarian cancers presented a problem, since no genetic model specifically applying to breast-ovary families has been derived. We considered that ovarian cancers should be treated differently than breast cancers, first, because the disease is much rarer in the general population (and therefore cases in the families are more likely to be genetic) and, second, because examination of these families and of those previously published suggests that (in contrast to breast cancer) ovarian cancers occurring at older ages are also part of the syndrome. We therefore assigned ovarian cancer cases to the “female affected before age 30” liability class. However, we also carried out analyses in which women with ovarian cancer were assigned risks based on the age at diagnosis, in the same manner as for breast cancer. If both cancers occurred in the same woman, the age of the first cancer was taken. The differences in results obtained by using these two approaches were very minor (see Results). All other cancers were ignored in the analyses. There is no clear evidence that any other cancers should be regarded as part of the syndrome, and, consequently, information on other cancers was not consistently reported. Excesses of other cancers could not therefore be reliably examined in this data set.

Linkage analyses in these families and in the CEPH families (Fain et al. 1991) shows that the female genetic map is considerably longer than the male map, over this region. Over the whole region, the distance ratio is approximately twofold, although there is variation across the region. For this reason, all two-point LOD scores have been calculated on the assumption that the female distance is twice the male distance. Results are summarized in the tables, according to the female recombination fraction (θ_f).

Given both the large number of alleles present for each marker and the complexity of many of the families, a full multipoint analysis including all markers and BRCA1 was not feasible. Instead, we performed a series of multipoint analyses. The first involved D17S250, D17S588, and BRCA1—except for the IARC families, for which no D17S250 data were available. In the IARC families, the analysis involved D17S74, D17S588, and BRCA1, with the LOD scores being evaluated at the equivalent locations on the genetic map, to allow a combined analysis of all families. In subsequent analy-

Table 2

Penetrance Values Used in the LINKAGE Program

AGE GROUP (years)	PENETRANCE OF GENOTYPE ^a		
	dd	Dd	DD
Unaffected females:			
<3000009	.008	.008
30–3900146	.083	.083
40–490083	.269	.269
50–59021	.469	.469
60–69039	.616	.616
70–79061	.724	.724
≥80082	.801	.801
Affected females:			
<3000002	.00167	.00167
30–3900026	.01276	.01276
40–4900112	.02305	.02305
50–5900137	.01711	.01711
60–6900226	.01260	.01260
70–7900218	.00908	.00908
≥8000213	.00654	.00654

NOTE.—Male breast cancers and ovarian cancers were assigned to the youngest age group.

^a D refers to the disease allele conferring susceptibility to breast and ovarian cancers, and d is the “normal” allele.

ses to further resolve the location, we analyzed D17S588, D17S579, and BRCA1. The position of the disease locus was also assessed by a more informal examination of recombinant events.

Multipoint analyses were performed using LINK-MAP (Lathrop et al. 1984), again by assuming a two-fold-increased map distance for females versus males. The computational efficiency of multipoint linkage analysis depends on the number of alleles at each of the loci. Analyses take progressively longer, the larger the number of alleles—while ignoring the phenotypic differences between alleles loses linkage information. The compromise which we adopted involved recoding the markers to six alleles for D17S588 and D17S250, to seven alleles for D17S579 (six “observed” alleles each with frequency .1 and a seventh “unobserved” allele with frequency .4), and to seven alleles for D17S74 (six “observed” alleles each with frequency .05 and a seventh “unobserved” allele with frequency .70). In total, 11 families show more than six alleles for D17S588, as do 19 for D17S250 and 16 for D17S579. For all except two of these families, recoding could be performed without change of linkage information, by considering those nuclear families in which the allele transmitted from each parent to child could be unequivocally determined. Alleles carried by individuals marrying into such nuclear families were then renumbered while maintaining evidence of transmission. As in the paper by Ott (1978), such recoding changes the absolute value of the likelihood but does not change the LOD score. In the two exceptions, some linkage information was lost, but this was shown to be trivial by comparing pairwise LOD scores before and after recoding. For D17S74, the large number of alleles recognized made recoding necessary in 35 families. In this case, recoding was conducted in such a way as to preserve as much linkage information as possible; consequently, some of the marker phase information from spouses of presumed carriers was lost. Specifically, in some families a noncarrier spouse heterozygous for D17S74 was recoded as being homozygous. Even with this recoding, each multipoint analysis took approximately 3 wk on a dedicated SUN 4 workstation.

Heterogeneity of linkage for various subgroups of families was examined by comparing log likelihoods calculated separately under the assumptions of heterogeneity and homogeneity. The subgroups considered were based on considerations of the genetic epidemiology of breast and ovarian cancer and so included an investigation both of age-at-onset effects for breast cancer and of the relevance of ovarian cancer cases

within a family. Following preliminary evidence that different proportions of breast-ovary families and breast-only families were linked to 17q, we estimated the proportion of each set of families linked, assuming that a single gene was responsible for both sets of families. This analysis was performed by computing the log likelihood over all families, postulating separate linked proportions for the two classifications of families but the same 17q susceptibility locus.

The linkage data were used to provide an estimate of the age-specific penetrance of the 17q gene, by maximizing the LOD score over different values of the penetrance function, which is equivalent to maximizing the likelihood conditional on all the disease phenotypes (Risch 1984). This method allows the penetrance to be estimated free of bias due to ascertainment of families on the basis of multiple affected individuals. This analysis was carried out using a modified version of the ILINK program (Lathrop et al. 1984). The analyses were based on the linkage results with D17S588, the marker showing the strongest linkage evidence—except for the Berkeley families, for which D17S588 was not completely typed; D17S579 was used for these families (For other groups, D17S588 was more completely typed than D17S579). For this analysis the recombination fraction between BRCA1 and D17S588 was fixed at .04 in males and .08 in females, as predicted by the multipoint analysis. For the breast-ovary families, penetrance was modeled assuming a separate parameter for the ratio of the incidence rates in gene carriers versus that in noncarriers, for each of the seven 10-year age groups defined previously. The rates in noncarriers were fixed at the rates in noncarriers in the CASH model. Since the aim of this analysis is to estimate the overall penetrance, age at onset was defined by the age at diagnosis of the first cancer (either breast or ovary). This assumption is in contrast to the previous linkage analyses, where ovarian cancers were distinguished from breast cancers (see above).

For the breast cancer families without ovarian cancer there was evidence of genetic heterogeneity. In order to obtain an estimate of the penetrance, we assumed, for simplicity, that the families resulted from the segregation of either BRCA1 or a second, unlinked autosomal dominant gene; these two genes were allowed to confer different penetrances. The penetrance of the linked gene was modeled in terms of seven age-specific risks, as for the breast-ovary families. The penetrance conferred by the unlinked gene was initially assumed to be equal to the penetrances given by Claus et al. (1991) and used in the previous analyses, but different penetrances

were also considered. It was also necessary to estimate a parameter $\gamma = p_1/(p_1 + p_2)$, where p_1 and p_2 are the gene frequencies of the linked and unlinked genes, respectively; γ is not, of course, the same as the proportion of linked families, since the two genes can confer different penetrances, and the probability of a family being linked will therefore depend on the number and ages at onset in affected individuals.

Results

Linkage Map

As a basis for the multipoint linkage analysis, we first constructed a multipoint linkage map of the six markers, based on the marker data in the breast cancer families and using the program CRIMAP (Barker et al. 1987; Lander and Green 1987). The preferred order for the markers, together with the corresponding sex-specific genetic distances between markers, is shown in figure 1. This order is preferred to all other orders with odds of more than $10^{14}:1$. It also corresponds to the order generated by analysis of recombinants in CEPH families (Fain et al. 1991), and is consistent with the order of loci generated by somatic cell and radiation hybrids (Black et al. 1993). The corresponding two-point marker-marker LOD scores and recombination fraction estimates are shown in table 3.

Two-Point Linkage Results under Homogeneity

Table 4 summarizes the overall two-point LOD scores for linkage between BRCA1 and the six chromosome 17q markers, by the female recombination fraction. Significant evidence of linkage is seen with all markers except growth hormone (GH). The strongest evidence of linkage is with D17S588 ($Z_{\max} = 21.68$ at $\theta_f = .13$) and D17S579 ($Z_{\max} = 13.02$ at $\theta_f = .16$). In terms of the maximum LOD score, this represents more than

a fourfold increase in the linkage evidence over that previously reported (Hall et al. 1990; Narod et al. 1991).

Table 5 presents a more detailed breakdown of the results for D17S588 and D17S579, according to the types of cancer occurring in the families. The majority of the linkage evidence comes from the breast-ovary families ($Z_{\max} = 18.44$ at $\theta_f = .08$, for D17S588; and $Z_{\max} = 11.24$ at $\theta_f = .08$, for D17S579). There is also significant linkage evidence within the families containing only female breast cancer cases, but the LOD scores are maximized at significantly greater distances (3.60 at $\theta_f = .22$, for D17S588; and 3.74 at $\theta_f = .22$, for D17S579). This is suggestive of genetic heterogeneity, which is examined in more detail below.

Of the four families containing male breast cancer (but not ovarian cancer), there is some evidence of linkage with D17S588, though not with D17S579. One family shows evidence of linkage (see Cohen et al. 1993), and the male breast cancer case shares the disease-bearing chromosome, while one particularly impressive family (see Hall et al. 1990) shows strong evidence against linkage. This latter family was not typed with D17S588, which explains the discrepancy in results between the two markers. These male breast cancer families have been excluded from all subsequent analyses.

Among the breast-ovary families, no difference is apparent between families with just one ovarian cancer case and those with multiple ovarian cancer cases. We repeated the analysis for the breast-ovary families, altering the liability class for the ovarian cancer cases to that appropriate to the age at diagnosis, as for the breast cancer cases; but this made little difference to the results ($Z_{\max} = 19.18$ at $\theta_f = .06$, for D17S588; and $Z_{\max} = 11.88$ at $\theta_f = .06$, for D17S579). In the breast cancer-only families, evidence of linkage to D17S588 is confined to families with at least four cases of breast cancer

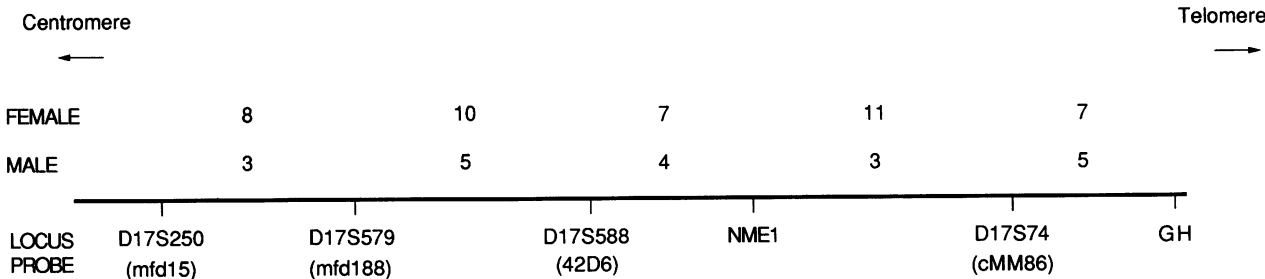


Figure 1 Map of 17q, showing both the order obtained from CRIMAP and the sex-specific genetic distances (in cM) between the markers.

Table 3

Estimated Sex-specific Recombination Fractions between Markers and Corresponding LOD Scores, Based on the Data Set of Breast Cancer Families

	D17S250	D17S579	D17S588	NME1	D17S74	GH
D17S250		77.79	58.80	34.30	41.13	22.35
D17S57908, .03		66.75	34.85	47.93	15.09
D17S58816, .03	.08, .06		48.43	77.16	27.53
NME121, .04	.15, .05	.08, .05		51.23	20.41
D17S7429, .07	.20, .08	.13, .06	.13, .03		71.64
GH26, .10	.28, .11	.19, .09	.16, .08	.06, .06	

NOTE.—Data below the diagonal are estimated sex-specific recombination fractions (in each pair, the first number is the estimate for females, and the second number is the estimate for males); and data above the diagonal are the corresponding LOD scores.

under the age of 60 years. There is also a significant effect of age at onset. For D17S579, linkage evidence was confined to families whose average age at onset was under 45 years. For D17S588, there was some evidence of linkage for families with older average ages, but the maximum LOD score occurred at a recombination fraction greater than that for the earlier-onset families.

Multipoint Analysis and Heterogeneity

The first multipoint analysis involved D17S588 and D17S250 for all families except the IARC families, for which an analysis of D17S74 and D17S588 was performed. Analysis of all 214 families jointly suggests the location of the gene to be between D17S588 and D17S250. Under the maximum-likelihood solution, the proportion of linked families (α) is estimated to be 62%, and the most likely location for BRCA1 is 9 cM proximal to D17S588. The odds in favor of this order are

250:1 over the location being distal to D17S588 and are 475,000:1 over the location being proximal to D17S250.

Figure 2 shows the LOD scores, by location, for the breast-only families, for the breast-ovary families, and for both categories combined. When only the 57 breast-ovary families are considered, the most plausible location is again 9 cM centrometric to D17S588, but in this group the best estimate of α is 1.0. The odds in favor of BRCA1 being located between D17S588 and D17S250 are reduced from those in the analysis of the whole set of families (86:1 over a location distal to D17S588 and 8,050:1 over a location proximal to D17S250). Figure 3 shows a contour plot of the LOD score over possible locations and α , for the breast-ovary families. The lower limit of the LOD – 1 support region, for α , is .71.

Analysis of the 153 breast-only families suggests a

Table 4

LOD Scores for Linkage between Breast and Ovarian Cancers and Markers on Chromosome 17q

Locus	LOD SCORE AT FEMALE RECOMBINATION FRACTION OF						Z_{\max}	θ_{\max}
	.001	.01	.05	.1	.2	.3		
D17S250	–11.98	–8.96	–1.20	3.81	7.30	6.65	7.42	.23
D17S579	–1.43	1.62	8.55	12.08	12.55	9.17	13.02	.16
D17S588	8.23	11.39	18.35	21.33	20.15	14.79	21.68	.13
NME1	–1.41	.75	6.01	8.70	9.13	6.76	9.45	.16
D17S74	–39.15	–31.73	–13.34	–2.73	6.32	7.50	7.67	.27
GH	–34.19	–30.24	–19.49	–11.80	–4.11	–1.03	.00	.50

^a The female map distance is assumed to be equal to twice the male map distance.

Table 5

LOD Scores between D17S588 and D17S579 and the Disease Locus, Subdivided by Family Type

LOD SCORES BETWEEN DISEASE LOCUS AND														
</														

^a With nonzero LOD scores.^b The female genetic distance is assumed to be twice the male distance.

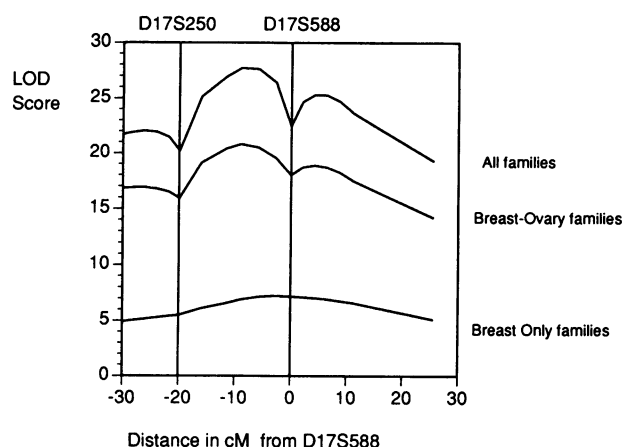


Figure 2 LOD score between BRCA1 and the 17q markers, obtained from an analysis of D17S588 and D17S250. The LOD score is defined as the maximum over α , the proportion of linked families. LOD scores for the total set of families, for the breast-ovary families, and for the breast-only families are shown.

gene location which is 5 cM proximal to D17S588, but there is little evidence over a more distal location (odds of 1.37:1). The location proximal to D17S250 is again the least plausible (odds of 55:1 for proximal versus distal to D17S250). In this case, 45% of breast-only families are estimated to be linked (LOD = 1 support limits 27%–65%), a proportion lower than that for breast-ovary families (LOD difference 1.21). Figure 2 also shows the location scores for these families, while figure 4 shows a contour plot of the LOD score over possible locations and α .

Finally, in figure 2, we also show the location score for the analysis of all 214 families, assuming that the proportion of breast-only families due to BRCA1 is different from the proportion of breast-ovary families due to this gene. In this analysis, the most plausible location for BRCA1 is again between D17S588 and D17S250, at a distance 9 cM from D17S588, with odds of 66:1 over a location distal to D17S588 and 490,000:1 over a location proximal to D17S250. At this location, the estimate of α for the breast-only families is .45.

Table 6 gives details of the multipoint heterogeneity analysis. Analysis by average age at onset showed some evidence that the proportion of linked breast-only families is age dependent; among families with an average age at onset before age 45 years, an estimated 67% were linked, compared with 19% of the families with mean onset age between 45 and 55 years and 38% for families with average onset age after 55 years ($\chi^2 = 4.65$, $P <$

.05, when families with average onset age under 45 years are compared with those with average onset age of 45 years or older). There is some suggestion that the proportion of linked families is higher among families with five or more cases with onset age under 45 years (the estimate of α is .72) than it is among families with fewer early-onset cases. There is little evidence of linkage among breast-only families with three or fewer cases in total ($Z = 0.36$ with α being estimated as .26).

We also conducted a multipoint analysis based on D17S588 and D17S579. This gave slight evidence in favor of a location for BRCA1 between D17S588 and D17S579, with odds of 30:1 over a location proximal to D17S579. The most likely location was 2 cM distal to D17S579.

Critical Recombinants

The multipoint analysis presented above required various statistical assumptions about the mode of inheritance of BRCA1 and only took into account two of the markers in each analysis. In an attempt to include information from more of the markers, we identified those between-marker recombination events which suggested a location for BRCA1 with respect to the markers. The search was restricted to apparently linked families, defined to be breast-only families with a maximum LOD score of more than 1.0 for D17S588, D17S579, or D17S250, and all breast-ovary families. Furthermore, since penetrance is incomplete, only recombinations which occurred either in affected individuals or in ancestors of affected individuals were considered. Even so, these observations need to be interpreted cautiously, for two reasons. First, there is no guarantee that a case occurring within a linked family is due to BRCA1, although recombinants in young affected cases provide more convincing evidence since they are more likely to be genetic. Second, some of the apparently linked families may not be due to BRCA1. No one observation can therefore provide conclusive information about gene location.

Table 7 contains lists of key recombination events occurring in families likely to be linked, which support the conclusion, from the multipoint analysis, that BRCA1 lies between D17S588 and D17S250. This table reports the families concerned, the type of family involved (breast only or breast-ovary), and a brief description of the informative individual. In one family (IARC 1816), two separate branches of the family share a haplotype proximal to D17S588 but are discordant for D17S588, suggesting that there has been a recombina-

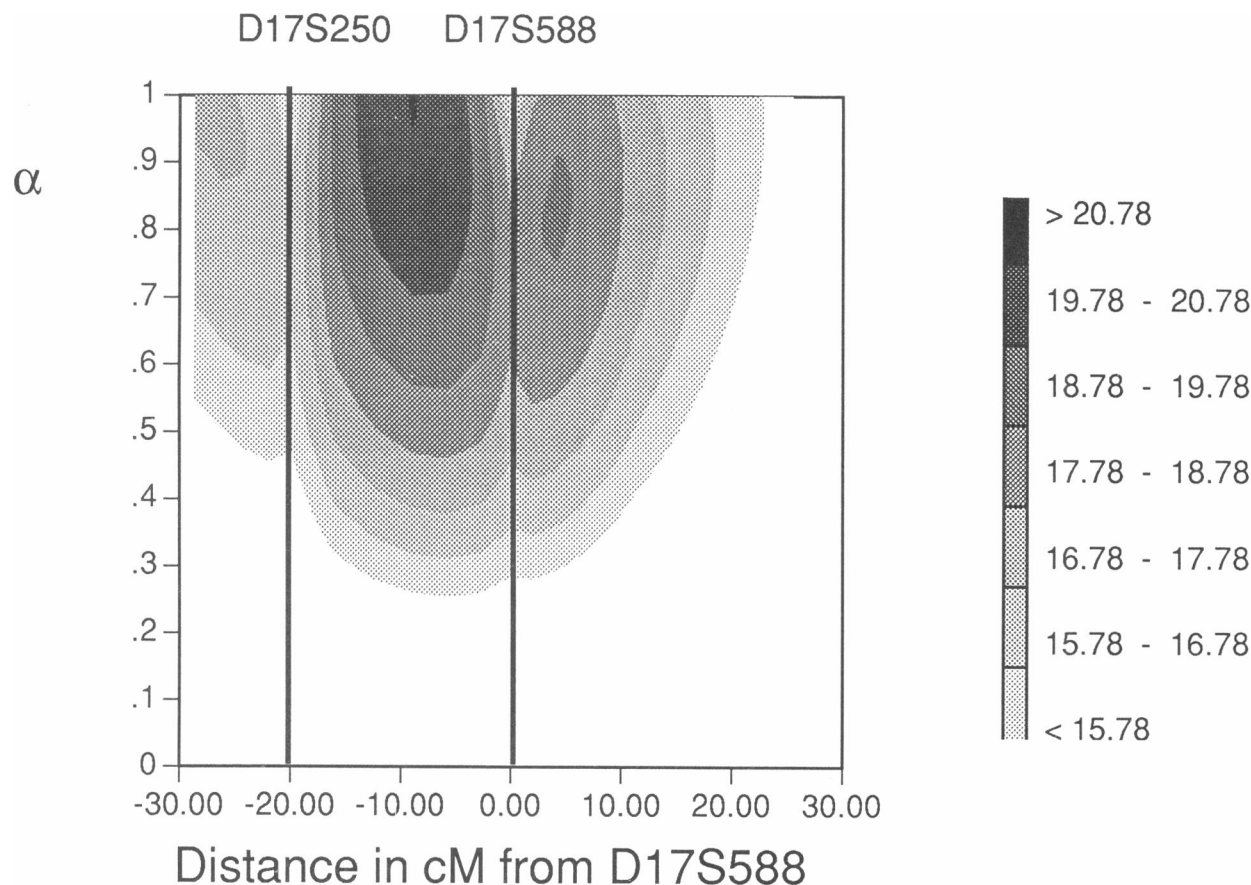


Figure 3 Contour plot showing LOD scores, by location and proportion of linked families, α , for the 57 breast-ovary families. The maximum LOD score is 20.79, which is obtained when all of the families are linked to BRCA1. The blackened region of this figure has LOD scores of at least 20.78, while all unshaded regions have LOD scores less than 15.78; and shaded regions have intermediate LOD scores, as indicated in the key to the right. The regions with LOD scores of at least 19.78 form the LOD - 1 support region.

nation event involving D17S588 in the common line of ancestry.

In table 7 we also report the recombinants which suggest a location with respect to D17S579, assuming that the gene is between D17S588 and D17S250. We include alternative explanations of the recombination events, since several occur either in families which are not clearly linked or in individuals who could be sporadic cases. The recombinants included in table 7 are described in more detail in the accompanying papers.

Penetrance

Inspection of the apparently linked families (as described above) suggests that the penetrance conferred by BRCA1 is close to 100%, since there are very few unaffected female obligate gene carriers. This can be confirmed by examining the carrier status of first-de-

gree relatives of breast or ovarian cancer cases. Relatives of breast cancer cases under age 60 years, or of ovarian cancer cases at any age, were included. In these families 201 relatives could be identified whose carrier status could be identified unambiguously (see table 8). Of these, 77 (39%) were determined as being gene carriers. However, of the unaffected women over age 60 years, only $2/32$ (6%) were carriers, indicating a high penetrance. Penetrance estimates were also obtained by maximizing the LOD score over different genetic models, in order to utilize all the available information. For the breast-ovary families, we first fitted a model in which the age-specific rates in carriers were those given by the CASH model used in the previous analysis, but multiplied by a constant factor r . The maximum conditional likelihood estimate obtained for r was 1.54, which yields a marginally significant improvement in

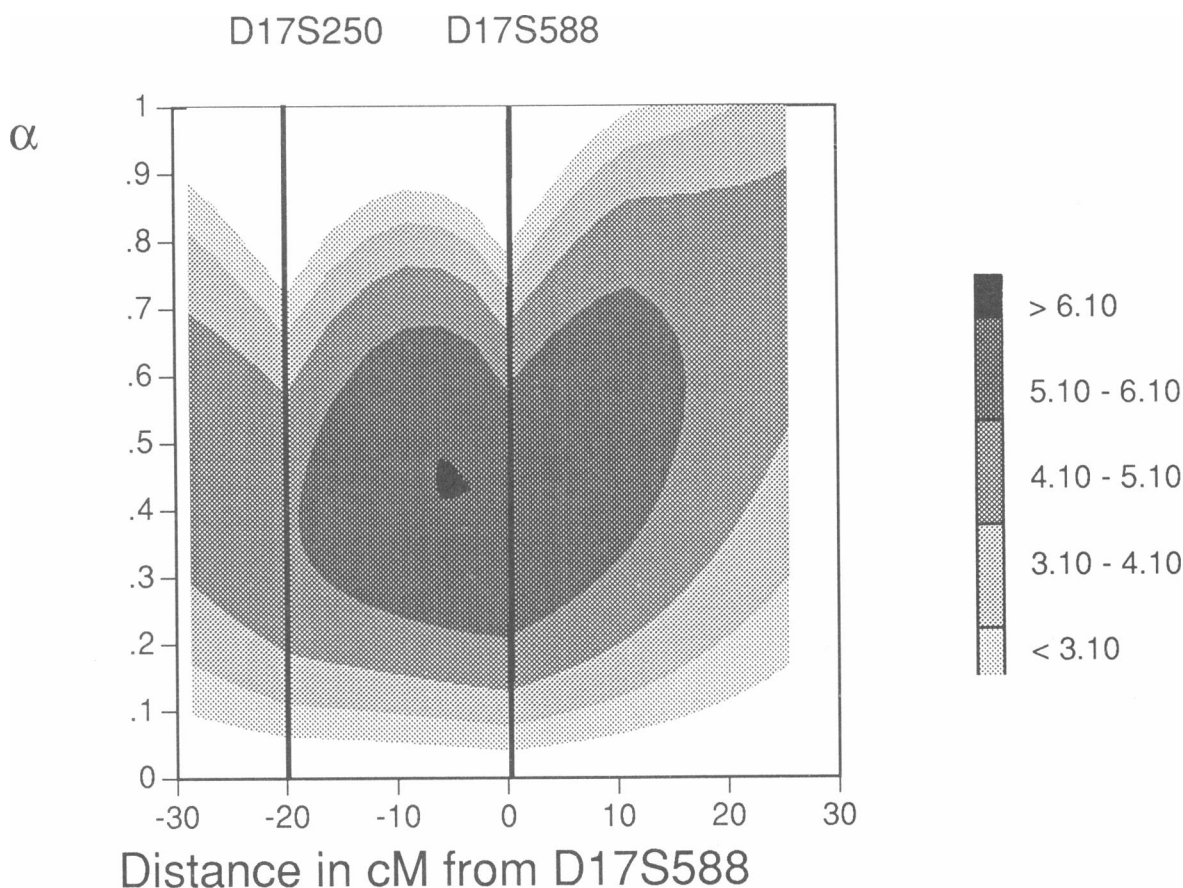


Figure 4 Contour plot showing the LOD scores, by location and proportion of linked families, α , for the 153 breast-only families. The maximum LOD score is 6.11, which is obtained when 45% of the families are linked to BRCA1. The blackened region of this figure has LOD scores of at least 6.10, while all unshaded regions have LOD scores less than 3.10; and shaded regions have intermediate LOD scores, as indicated in the key to the right. The regions with LOD scores of at least 5.10 form the LOD - 1 support region.

the LOD score (22.14, compared with 21.32; $\chi^2 = 3.78$, $P = .05$). This model would correspond to penetrances of 52% by age 50 years and 82% by age 70 years, compared with 38% and 67%, respectively, under the CASH model. If the age-specific rates in all age groups are allowed to vary, we obtain the penetrance function illustrated by the dotted line in figure 5. The estimated penetrance is 67% by age 50 years (95% confidence limits [CL] 44%–80%) and 76% by age 70 years (95% CL 52%–88%).

For the breast-only families we first performed analyses assuming that the risks caused by the unlinked gene(s) followed the CASH model as previously, varying only the penetrance of the linked gene. The result of fitting age-specific penetrances is shown by the dashed line in figure 5; the resulting penetrances are 49% by age

50 years and 90% by age 70 years. Under this model, γ , the gene frequency of BRCA1 as a proportion of the total frequency all breast cancer genes (constrained to be .0033 as in the CASH model), is estimated as .39. The shape of the penetrance function appears somewhat different from that of the CASH model, with high age-specific risks up to age 60 years but with little risk thereafter. If the penetrance of BRCA1 is assumed to be the same in both breast-only and in breast-ovary families, the best estimates are 59% by age 50 years (95% CL 39%–72%) and 82% by age 70 years (95% CL 64%–91%), with γ for the breast cancer estimated to be .35. This penetrance function is illustrated by the continuous line in figure 5.

We also performed analyses in which the risks conferred by the non-17q genes were reduced to $\frac{1}{2}$ of the

Table 6**Multipoint Heterogeneity Analysis**

Group (no. of families)	LOD Score	Proportion Linked ^a
Breast-ovary (57)	20.79	1.00
Breast-only (153): ^b	6.01	.45
Average age (years) at diagnosis:		
<45 (54)	6.56	.67
45–54 (63)40	.19
≥55 (36)08	.38
No. of cases diagnosed:		
Before age 45 years:		
Two or fewer (110)	1.46	.44
Three or four (36)	1.89	.39
Five or more (7)	2.88	.72
Before age 60 years:		
Three or fewer (98)36	.26
Four or five (38)	3.29	.60
Six or more (17)	2.72	.45

^a All estimates are at the maximum-likelihood position 9 cM proximal to D17S588 on the female map.

^b Four families with male breast cancer cases were excluded.

risks implied by the CASH model. This reduced slightly the estimated penetrance of BRCA1. γ was, however, reduced to .05.

Haplotype Sharing among Affected Relatives

The haplotype sharing at the D17S74 locus, reported by Hall et al. (1990), suggested that young affected sisters shared the allele from the nontransmitting parent more often than would be expected by chance, perhaps indicating a recessive component to early-onset breast cancer, conferred by the second 17q allele (if BRCA1 were acting in a dominant fashion, the allele from the nontransmitting parent should be shared by 50% of affected sister pairs). To test this hypothesis by using the closer markers, we identified 30 sister-sister pairs in “linked families” (as defined previously) where both sisters were typed and affected with breast cancer; of these, 16 (53%) shared the haplotype based on D17S588, D17S579, and D17S250 from the nontransmitting parent. The proportions for sister pairs in which both members were affected with ovarian cancer and for sister pairs in which one sister had breast cancer and one had ovarian cancer were $3/7$ (43%) and $6/16$ (38%), respectively. Thus there is no evidence of a recessive effect being conferred by the homologous BRCA1 allele.

Affected-Sib-Pair Analysis

The linkage analysis gives little support for linkage among the families containing, at most, three cases of breast cancer diagnosed under age 60 years. One concern is that this might be an artifact due to an inappropriate genetic model (e.g., if such families were the result of much lower penetrance mutations). As an alternative approach, we examined sharing of the 17q haplotypes between sister pairs affected with breast cancer, subdivided by the number of breast cancer cases diagnosed under age 60 years in the family, to determine whether there was any evidence of excess sharing. The identity-by-descent results are shown in table 9. For this analysis, haplotypes based on D17S588, D17S579, and D17S250 were examined assuming no recombination within the sibship. The results of this analysis are in agreement with the likelihood analysis, in showing no evidence of linkage for families with two or three breast cancer cases under age 60 years at diagnosis, while families with four cases did show some evidence of linkage.

Sporadic Cases of Breast and Ovarian Cancer in the Families

Within the families, some cases of either breast or ovarian cancer which did not appear to carry the high-risk haplotype could be picked out. Table 10 lists apparently sporadic cases which were, at most, third-degree relatives of other affected individuals in linked families (as defined previously). Of the six cases identified, only two were first-degree relatives of other affected individuals.

Discussion

In this study we have attempted to obtain 17q linkage data on breast cancer and breast-ovary families from as many groups as possible, worldwide. This represents a very considerable increase in the total available linkage data (the maximum LOD score is increased more than fourfold over that from previous studies) and provides conclusive proof of the existence, on chromosome 17q, of a gene predisposing to breast and ovarian cancers.

All the collaborating groups contributed data on all their breast cancer families for which 17q marker typing was available, regardless of whether there was any evidence for linkage. In particular, typings for either D17S588 or D17S250 (or both), the markers used in our main multipoint analysis, were available for almost

Table 7**Recombinants with Relevance to the Fine-Scale Mapping of BRCA1**

Group	Family	Family Type	Maximum LOD Score	Critical Observation	
Recombinants putting BRCA1 proximal to D17S588: ^a					
Berkeley	7	Breast-only	1.75 ^b	Breast cancer at age 51 years	
	5	Breast-ovary	2.01 ^c	Breast cancer at age 40 years	
	3	Breast-ovary	.64 ^b	Breast cancer at age 31 years	
CRC	925	Breast-ovary	.21 ^d	Breast cancer at age 33 years	
	871	Breast-ovary	.57 ^b	Breast cancer at age 55 years	
IARC	1816	Breast-ovary	3.51 ^c	Two sides recombinant ^e	
MRC	1	Breast-ovary	.31 ^f	Breast cancer at age 35 years	
Utah	1927	Breast-only	1.16 ^c	Breast cancer at age 38 years	
Recombinants putting BRCA1 distal to D17S250: ^{a,g}					
Berkeley	1	Breast-only	1.68 ^b	Breast cancer at age 37 years	
CRC	801	Breast-ovary	.62 ^b	Ovarian cancer at age 50 years	
FCN	4	Breast-ovary	.04 ^f	Ovarian cancer at age 32 years ^h	
	17	Breast-ovary	.61 ^f	Breast cancer at age 38 years	
ICRF	BOV2	Breast-ovary	.19 ^b	Breast cancer at age 44 years, or breast cancer at age 46 years, and ovarian cancer at age 52 years	
Utah	1901	Breast-only	1.52 ^f	Breast cancer at age 45 years	
	2082	Breast-ovary	3.84 ^f	Ovarian cancer at age 45 years	
					Possible Explanation
Recombinant putting BRCA1 proximal to D17S579: ⁱ					
Leiden	19	Breast-ovary	.82 ^d	Ovarian cancer at age 52 years	Sporadic case?
Recombinants putting BRCA1 distal to D17S579:					
IARC	1816	Breast-ovary	3.51 ^c	Breast cancer at age 57 years	Sporadic case?
MRC	11	Male breast	1.81 ^c	Two sides recombinant ^j	Unlinked family

^a Recombinants in breast-ovary families and breast-only families with a maximum LOD score greater than 1.0 for one of the markers in the region.

^b With D17S579.

^c Multipoint.

^d With D17S250.

^e The two sides of this family have an identical haplotype centromeric to D17S588 but are discordant for D17S588.

^f With D17S588.

^g An additional recombinant, not included in this data set, has been noted by Bowcock et al. (1993).

^h The most likely recombinant individual, although the recombination could have occurred in a woman with bilateral breast cancer at ages 29 and 33 years.

ⁱ An additional recombinant has been identified in the breast-only family Utah 1901 (Goldgar et al. 1993), which suggests that BRCA1 is proximal to D17S579.

^j The recombinant individual cannot be determined.

every family in the study. We have no reason to believe, therefore, that our estimates of the proportions of families (of different types) which are linked are materially biased.

In particular, this study provides much clearer evidence than was previously available that BRCA1 predisposes to ovarian cancer as well as to early-onset breast cancer. This was an implicit assumption in the

Table 8**Carrier Status of First-Degree Relatives of Breast and Ovarian Cancer Cases in Apparently Linked Families**

Age Group (years)	No. of Carriers/ Total No. of First-Degree Relatives (%)
<40	50/99 (51%)
40-49	14/41 (34%)
50-59	11/29 (38%)
60-69	1/14 (7%)
≥70	1/18 (5%)

above-cited linkage analyses, but this assumption can be verified by examining haplotype sharing between the breast and ovarian cancers. Of the 24 ovarian cancer cases who were first-degree relatives of breast cancer cases in families where the disease-bearing chromosome could be identified from the breast cancer cases

alone and who were not obligatory carriers of the disease gene, only 2 cases did not appear to carry the disease chromosome, although one could also be an informative recombinant (table 7, IARC family 1234). The other, in ICRF family BOV3, is discussed below.

Although this study did not specifically aim to investigate ovarian cancer susceptibility, it is worth noting that some of the families in the data set were specifically ascertained through multiple cases of ovarian cancer. Only one family in the study did not contain any cases of breast cancer, and this family was consistent with linkage. These results raise the possibility, therefore, that BRCA1 explains most dominant susceptibility to ovarian cancer. Further families selected solely for the presence of multiple cases of ovarian cancer need to be studied to test this hypothesis.

When we consider first the question of gene localization, the results of this study provide evidence that the cancer-susceptibility gene lies between D17S250 and

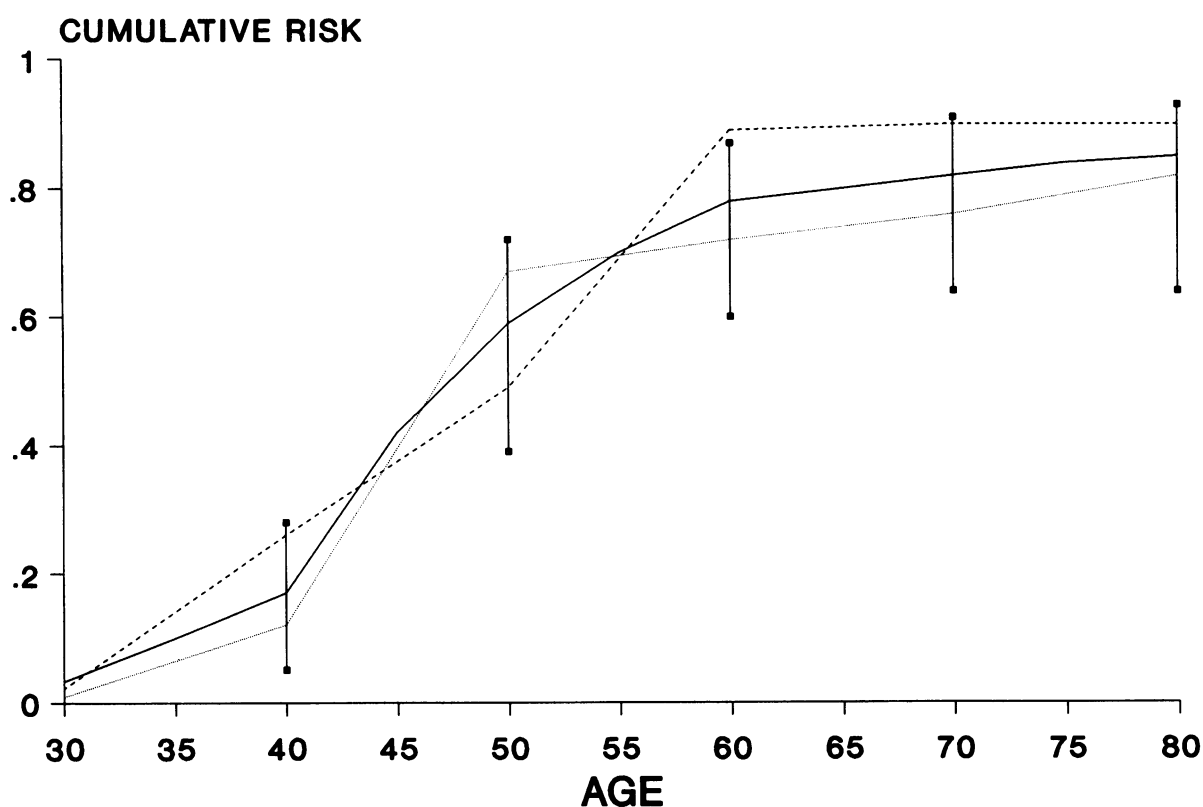


Figure 5 Estimated cumulative risk for either breast or ovarian cancer in female carriers of the BRCA1 mutation. The dotted line indicates the estimated cumulative risk for breast-ovary families; the dashed line indicates the cumulative risk for breast-only families; and the continuous line (with CL) indicates the cumulative risk for all families combined.

Table 9

Observed and Expected Haplotype Sharing among Relative Pairs in Which Both Members Are Affected with Breast Cancer at <60 Years of Age, in Families Containing up to Four Breast Cancers

No. OF CASES	NO. OF HAPLOTYPES SHARED BY TWO AFFECTED RELATIVES													
	Both Relatives <45 Years of Age at Diagnosis							Both Relatives <60 Years of Age, at Least One ≥45 Years of Age						
	Haplotype Shared by First- Degree Relatives			Haplotype Shared by Second- Degree Relatives		Haplotype Shared by Third- Degree Relatives		Haplotype Shared by First- Degree Relatives			Haplotype Shared by Second- Degree Relatives		Haplotype Shared by Third- Degree Relatives	
	2	1	0	1	0	1	0	2	1	0	1	0	1	0
Two	2	2	0	0	0	1	0	1	4	3	2	0	1	2
Three	1	5	3	3	5	0	0	7	17	5	6	9	0	4
Four	4	3	0	0	3	2	1	2	1	2	7	1	2	1
Total no. observed	7	10	3	3	8	3	1	10	22	10	15	10	3	7
No. expected ^a	5	10	5	5.5	5.5	1	3	10.5	21	10.5	12.5	12.5	2.5	7.5

^a Computed under the hypothesis of no linkage to the 17q interval from D17S250 to D17S588.

D17S588, thus confirming the results of Hall et al. (1992). This interval is approximately 8.3 cM in males and 18.0 cM in females. Evidence for this localization is provided both by the summary of key recombinant events and by a formal multipoint analysis which favors this interval with odds of more than 66:1 over either adjacent interval. Nine recombinant events in appar-

ently linked families indicate a location proximal to D17S588, and seven indicate a location distal to D17S250. The apparent sporadic cases listed in table 10 might be considered as evidence against this interval, since they could be explained as gene carriers if the gene were proximal to D17S250. Note, however, that only two of the six individuals listed are first-degree relatives

Table 10

Apparent Sporadic Cases of Ovarian Cancer or Breast Cancer Diagnosed at <60 Years of Age, in Breast-Ovary Families and Breast Cancer Families with LOD Scores Greater than 1.0

Group	Family	Sporadic Case	Closest Affected Carrier	Comment
Aberdeen	1	Breast cancer at age 35 years	Second-degree relative	Does not share D17S250-GH region
Berkeley	5	Breast cancer at age 59 years	Second-degree relative	Intervening unaffected 86-year-old female
Berkeley	5	Breast cancer at age 36 years	Third-degree relative	Intervening unaffected 92-year-old female
IARC	2651	Breast cancer at age 48 years	Second-degree relative	Linkage implies that BRCA1 enters into family from spouse of this case's offspring
ICEL	6	Breast cancer at age 43 years	First-degree relative	Two branches separated by unaffected females, ages 75 and 81 years
ICRF	6	Breast cancer at age 49 years	Third-degree relative	Intervening unaffected 81-year-old female
	BOV3	Ovarian cancer at age 59 years	First-degree relative	Shares NME1 and telomeric region but not THRA1-D17S588 region

of affected gene carriers. The first (in Iceland family 6) was diagnosed with breast cancer at age 43 years and is the daughter of an untyped breast cancer case, and so we cannot be sure that either carries the disease-bearing chromosome. The second (in ICRF family BOV3) was diagnosed with ovarian cancer at age 59 years (Spurr et al. 1993). This case does not share with her affected sister or mother the region from THRA1 to D17S588.

We attempted to define further the location of BRCA1 within the interval D17S250-D17S588 but found that the orientation with respect to D17S579 could not be determined. Multipoint analysis gave some evidence for the interval D17S588-D17S579, but with odds of only 30:1 over the interval D17S579-D17S250. A subsequent examination of the key recombinant events with D17S579 showed no consistent pattern, with two recombinants supporting a location proximal to D17S579 (one of the two occurred in Utah 1901 and is based on data unavailable to us at the time of the analysis) and two suggesting a distal location. None of these observations is entirely convincing, as the families may be unlinked or the important cases may be sporadic. However, a further D17S579 recombinant in an apparently linked family, not included in this data set, has been reported by Chamberlain et al. (1993). This recombinant would place BRCA1 proximal to D17S579. The recombinant occurred in an individual with both breast and ovarian cancer by age 35 years and therefore provides more convincing evidence of location than the other recombinants. Further evidence which includes D17S579 comes from Bowcock et al. (1993), who report a likely recombinant which indicates a location for BRCA1 distal to THRA1, which is known to lie between D17S250 and D17S579.

Turning to the issue of heterogeneity, a clear distinction is apparent in this study, between breast-ovary families and breast-only families. All or almost all families in the former group appear to be compatible with linkage. In this regard, it is interesting to note that the two families, IARC 2651 and IARC 2770, which gave evidence against linkage with D17S74 in the previous analysis by Narod et al. (1991) both gave positive multipoint LOD scores in our analysis. However, it is important to bear in mind that the confidence interval for the proportion of linked families is wide (lower LOD - 1 support limit 0.71) and that a number of these families may ultimately turn out not to be the result of BRCA1 mutations. Indeed, we do not expect all such families to be linked, since ovarian cancer cases

must sometimes occur by chance in breast cancer families. Although none of the breast-ovary families gave clear evidence against linkage in our analysis, some evidence against linkage was found in three breast-ovary families from IARC, in the analysis by Feunteun et al. (1993); their analysis included data on cLB17.1 (close to D17S579), as well as some further typings not included in our data set. In addition, ovarian cancer cases were treated slightly differently by Feunteun et al. (1993) than in our analysis. All three of these families contained just one ovarian cancer. One of these (in family IARC 1252) was diagnosed at age 84 years; the second (in family IARC 2850) occurred in an individual with a previous breast cancer, in a family with 14 other breast cancer cases; the third (in family IARC 2651) was diagnosed at age 26 years, but the individual was a fourth-degree relative of the nearest early-onset breast cancer case, with two intervening unaffected women. In this family in particular, the evidence for linkage depends critically on what assumption is made about the probability that such a young ovarian cancer case is a gene carrier. Smith et al. (1993) also report one breast-ovary family with some evidence against linkage (family CRC128). The evidence against linkage in this family derives from a breast cancer case diagnosed at age 35 years who does not share the putative disease chromosome with her affected sisters. This result also depends on marker typings not available at the time of the present analysis. Finally, some suggestive evidence against linkage was also found in three of the Utah breast-ovary families studied by Goldgar et al. (1993).

In contrast to the breast-ovary families, a substantial proportion of families segregating early-onset breast cancer but not ovarian cancer do not appear to be linked to 17q. It might be argued that this heterogeneity is an artifact of the assumed genetic model and that apparently unlinked families are merely chance aggregations of cases. This seems unlikely, however, both because there are a few impressive families which do not appear compatible with linkage and because the haplotype sharing in the smaller families shows no evidence of linkage, even though epidemiological studies would indicate that such families are not chance aggregations. This indicates that some—possibly the majority—of the genetic variation in breast cancer is not the result of 17q loci. The precise genetic model underlying these unlinked families remains unclear, however, and need not necessarily involve penetrant dominant genes. Among the families without ovarian cancer, there is

partial confirmation of the “age at onset” effect suggested by Hall et al. (1990). Only a proportion of the families with average age at onset under 45 years appear to be linked (67% by the multipoint analysis), but the proportion is lower among families with a higher average age at onset.

The probability of linkage also appears to be related to the number of breast cancer cases. While there are many small families in which affected individuals do not share a consistent 17q haplotype, there are few large families which are clearly incompatible with linkage (the most negative multipoint LOD score at the putative location for BRCA1 is -2.03 in family FCN1, and only four families have LOD scores less than -1.0). The obvious explanation of this phenomenon is that BRCA1 confers a higher age-specific breast cancer risk than do other breast cancer genes. We have addressed this question formally by attempting to estimate the penetrance in the linked families, using the marker data. This provides evidence that the penetrance of BRCA1 is high; of 32 unaffected individuals age 60 years and older who were first-degree relatives of affected individuals in linked families, only 2 are gene carriers. We estimated a penetrance of 59% by age 50 years and 82% by age 70 years in BRCA1 carriers. This compares to 38% and 67%, respectively, predicted by the CASH model (it should, however, be noted that these latter figures relate solely to breast cancer penetrance, while our estimates are for either breast or ovarian cancer). There was no obvious difference in penetrance between the linked breast cancer families and the breast-ovary families. Clearly, better estimates of penetrance will ultimately be obtained from population-based studies of mutations, once the gene has been cloned; but this study provides useful estimates in the interim. In particular, the confirmation that the gene confers a high lifetime risk of cancer has immediate implications for genetic counseling in linked families. The degree of heterogeneity indicates that counseling on the basis of 17q markers will only be appropriate in families with a high posterior probability of linkage (in particular, breast-ovary families) and will therefore not be appropriate for the much larger group of women with only two or three affected relatives.

In addition to the questions alluded to above, a number of other issues remain to be addressed. One is whether BRCA1 confers a risk of any cancer other than breast and ovarian cancer. A further study based on this set of families is attempting to answer this question.

Another question is whether there is any allelic heterogeneity among linked families. For example, some of the linked families contain no ovarian cancers, whereas in others the ovarian cancers outnumber the breast cancers. It is impossible to determine unambiguously from this study to what extent these differences are a result of differences in ascertainment. There is also the possibility that there are mutations in the BRCA1 gene which confer much lower cancer risks. Such mutations would not, in general, give rise to the heavily loaded families which provide most of the linkage evidence in this data set. The lack of evidence of linkage in the smaller families, however, suggests that such mutations do not explain a significant proportion of familial breast cancer. If different mutations prove to confer different risks, this would provide further complications for genetic counseling. Finally, this analysis suggests that only a fraction of the familial aggregation of breast cancer is due to BRCA1 and that other breast cancer-susceptibility genes may be mappable; population-based analyses of familial risks, allowing for the effect of BRCA1, will be necessary to elucidate the contribution of other genes.

Acknowledgments

This study was originated as a collaboration by the UK Cancer Family Study Group. The analysis of this study was supported by the Cancer Research Campaign and the Imperial Cancer Research Fund. The financial support for two meetings of the consortium group was provided by the International Agency for Research on Cancer. During this analysis, extensive use of the Pedigree/Draw program, the LINKAGE program, and CRIMAP was made. We would like to thank Dr. Mark Lathrop and colleagues for providing us with the LINKAGE package used, and we would like to thank Dr. Phil Green for CRIMAP. We would like to give special thanks to Paul Mamelka, Dr. Jean MacCluer, and Dr. Bennett Dyke for their willingness to modify the Pedigree/Draw software to satisfy our requirements. Drs. David Barker, Donny Black, Jeff Hall, Ken Kidd, Michael Polymeropoulos, Mark Skolnick, Jim Weber, and Ray White provided markers that were used by the consortium members; we would like to thank each of them.

Appendix

Data in these two tables were supplied by Dr. Nigel Spurr of the Imperial Cancer Research Fund, London.

Table A1**DNA Probes Screened during the Consortium Linkage Study**

Probe	Human Gene Mapping Name of Locus	Enzyme	Alleles (frequency) (kb)	Reference ^a
pBS3	HOX2B	<i>SacI</i>	{ A1, 2.0 (.85) A2, .9, 1.1 (.15) }	Human Gene Mapping 11 1991
pH2-1-3Bg2	HOX2F	<i>MspI</i>	{ A1, 2.0 (.59) A2, 1.8 (.41) }	Human Gene Mapping 11 1991
pH2-1-9	HOX2G	<i>TaqI</i>	{ A1, 3.8 (.15) A2, 3.2 (.85) }	Human Gene Mapping 11 1991
NM23	NME1	<i>BglII</i>	{ A1, 4.8 (.8) A2, 2.1 (.2) }	Varesco et al. 1992
pTHH59	D17S4	<i>PvuII</i>	VNTR with six alleles of .8–1.8 kb	Human Gene Mapping 11 1991
pRMU3	D17S24	<i>PvuII</i>	VNTR with >10 alleles	Human Gene Mapping 11 1991
LEW101	D17S40	<i>MspI</i>	{ A1, 15.0 (.7) A2, 7.0 (.3) }	Human Gene Mapping 11 1991
LEW102	D17S41	{ <i>PstI</i> <i>TaqI</i> }	{ A1, 12.0 (.25) A2, 8.0 (.75) B1, 8.0 (.25) B2, 5.5 (.75) }	Human Gene Mapping 11 1991
cMM86	D17S74	{ <i>HinfI</i> <i>TaqI</i> }	VNTR with >20 alleles of 1.0–3.5 kb	Human Gene Mapping 11 1991
LCW5A2	D17S308	{ <i>PvuII</i> <i>TaqI</i> <i>BamHI</i> }	VNTR with alleles of 2–5 kb	Human Gene Mapping 11 1991

^a Human Gene Mapping 11 (1991) is the publication of the proceedings of Human Gene Mapping Workshop 11, held in London in August 1991.

Table A2

PCR-formatted Markers Screened during the Consortium Linkage Study

Primer	Human Gene Mapping Name for Locus	Primer Sequences	Expected Product Size	No. of Alleles	Reference
COL1A1	COL1A1	{1. CCGAAGGTTCCCTGGACGA 2. CGCCTGTTCGCCTGTCTCA}	A1 = 356 bp A2 = 96 bp, 61 bp plus 10 constant bands after digestion of 850- bp fragment by <i>HaeIII</i>	A1 = .91 A2 = .09	MacKay et al. 1990
HER2	ERBB2	{1. CTGGAATGGGAAGCA 2. GCCAGCAAGAAATCTTAGACGT}	1.1 kb	{A1 = .7} ^a A2 = .3	Hall et al. 1992
GH	GH	{1. (AAG) TCCAGCCTCGGAGACAGAAT 2. (CTT) AGTCCTTTCTCCAGAGCAGGT}	~243 bp	~18	Polymeropoulos et al. 1991
GIP	GIP	{1. CACAATGGGCTCGACTTAGCATAA 2. CTTGCTGGATCAGACAAACCTCTG}	{A1 = 1.95 kb A2 = 1.8 kb A3 = 1.65 kb}	A1 = .44 A2 = .5 A3 = .06	Human Gene Mapping 11 1991
MPO	MPO	{1. (CA) ACATCTGGTGCTTCAATGCC 2. (GT) AAAACTACCTTAGCACCGA}	~100 bp	~5	M. H. Polymeropoulos, personal communication
NM23	NME1	{1. (CA) TCTCAGTACTTCCGTGACC 2. (GT) TTGACGGGGTAGAAGCTC}	~106 bp	5-6	Hall et al. 1992
THRA1	THRA1	{1. GGGCAAAAATGTCTTAAGC 2. CAGCATAGCATTGCCTTC}	~120 bp	5-7	Bowcock et al. 1993
C27.2	D17S170	{1. (CA) GTGTGAAACTCCATCTCTGC 2. (GT) TTACAGGCACACGGCACCAT}	80 bp	~3	Black et al. 1993
E39	D17S178	{1. (CA) CCTGGCCTGGACTCTCTTT 2. (GT) CTCATGTACCGCCCTCA}	~100 bp	~5	Black et al. 1993
E55	D17S180	{1. (CA) CTCACCTGGTTCAGCACTC 2. (GT) TTATTGAATAGCTGATAGGGAC}	~100 bp	3	Black et al. 1993

(continued)

Table A2 (continued)

Primer	Human Gene Mapping Name for Locus	Primer Sequences	Expected Product Size	No. of Alleles	Reference
SCG40	D17S181	{1. (CA) CCTACAAGCTTTACTGCACTA 2. (GT) AAGTAGTGAAGGATCATATGTC}	~114-150 bp	<div> 1. = .25 2. = .05 3. = .05 4. = .1 5. = .18 6. = .2 7. = .1 8. = .25 9. = .075 10. = ND 11. = .025 12. = .05 13. = .025 14. = ND 15. = .075 16. = ND 17. = .25 Heterozygosity = >.90 </div>	Black et al. 1993
B43	D17S183	{1. (CA) ACAAACTGATGTGGGCTCTAG 2. (GT) GTACATAGCATGGGTGCAGCT}	~100 bp	<div> 1. = .1 2. = .05 3. = .15 4. = .7 Heterozygosity = .4 </div>	Black et al. 1993
C54	D17S190	{1. (CA) ACCTGACCCATCTCTGTC 2. (GT) GTGACCTGTCATATGCACG}	~100 bp	<div> 1. = .1 2. = .1 3. = .2 4. = .6 Heterozygosity = .5 </div>	Black et al. 1993

D58	D17S194	$\left\{ \begin{array}{l} 1. (CA) GAACTCACAGCCTTCAGCCT \\ 2. (GT) GCAATGAATCGTGTGAATGGA \end{array} \right\}$	~100 bp	~3	Black et al. 1993
Mfd15	D17S250	$\left\{ \begin{array}{l} 1. (CA) GGAAGAATCAAATAGACAAT \\ 2. (GT) GCTGGCCATATATATTTAAACC \end{array} \right\}$	162 bp	10	Human Gene Mapping 11 1991
G1	D17S287	$\left\{ \begin{array}{l} 1. (CA) GGAGAAACTATGGCTGGAG \\ 2. (GT) GATATGATTCAGAAAACAGCC \end{array} \right\}$	~75 bp	~6	Black et al. 1993
6C1	D17S293	$\left\{ \begin{array}{l} 1. (CA) ACAGTGCCAGAGATATACCG \\ 2. (GT) GCTATGAGCCTGGCAGACC \end{array} \right\}$	~120 bp	$\left\{ \begin{array}{l} 1. = .09 \\ 2. = .3 \\ 3. = .11 \\ 4. = .11 \\ 5. = .3 \\ 6. = .04 \\ 7. = .05 \end{array} \right\}$ Heterozygosity = .75	Hall et al. 1992
Mfd188	D17S579	$\left\{ \begin{array}{l} 1. (CA) AGTCCTGTAGACAAAACCTG \\ 2. (GT) CAGTTTCATACCAAGTTCT \end{array} \right\}$	~120 bp	12	Hall et al. 1992
46E6	D17S587	$\left\{ \begin{array}{l} 1. (CA) TTCATGGGGCTTACTGTGTTC \\ 2. (GT) TAGCACTCTGCCCTTCCAACATAC \end{array} \right\}$	~100 bp	>6	M. H. Skolnick, personal communication
42D6	D17S588	$\left\{ \begin{array}{l} 1. (CA) CCTGGTCTAGGAAGAGTGCTCA \\ 2. (GT) GTGTAAAGCATCTGTGTATACTAC \end{array} \right\}$	~100 bp	~7	Goldgar et al. (1993)
S52		$\left\{ \begin{array}{l} 1. (CA) TGGAGGATGCAGTGAGCTGAAATG \\ 2. (GT) GACCTCCCCAAAGTACTGAGTTAC \end{array} \right\}$	~150 bp	8-10	R. White, personal communication
MR322		$\left\{ \begin{array}{l} 1. CCCCCCCCCCATACACAGGCT \\ 2. CCTGGCCATGTTTTTTTAAATGTC \end{array} \right\}$	~766 bp	4-5	R. White, personal communication

* From *Mbol* digestion.

References

- Adami HO, Hansen H, Jung B, Rimsten A (1980) Familiality in breast cancer: a case-control study in a Swedish population. *Br J Cancer* 42:71-77
- Arason A, Barkadóttir RB, Egilsson V (1993) Linkage analysis of chromosome 17q markers and breast-ovarian cancer in Icelandic families, and possible relationship to prostatic cancer. *Am J Hum Genet* 52:711-717
- Barker D, Green P, Knowlton R, Schumm J, Landers E, Oliphant A, Willard H, et al (1987) Genetic linkage map of human chromosome 7 with 63 DNA markers. *Proc Natl Acad Sci USA* 22:8006-8010
- Bishop DT, Cannon-Albright L, McLellan T, Gardner EJ, Skolnick MH (1988) Segregation and linkage analysis of nine Utah breast cancer pedigrees. *Genet Epidemiol* 5:151-169
- Black DM, Nicolai H, Borrow J, Solomon E (1993) A somatic cell hybrid map of the long arm of human chromosome 17, containing the familial breast cancer locus (BRCA1). *Am J Hum Genet* 52:702-710
- Bowcock AM, Anderson LA, Friedman LS, Black DM, Osborne-Lawrence S, Rowell SE, Hall JM, et al (1993) THRA1 and D17S183 flank an interval of <4 cM for the breast-ovarian cancer gene (BRCA1) on chromosome 17q21. *Am J Hum Genet* 52:718-722
- Chamberlain JS, Boehnke M, Frank TS, Kiouisis S, Xu J, Guo S-W, Hauser ER, et al (1993) BRCA1 maps proximal to D17S579 on chromosome 17q21 by genetic linkage analysis. *Am J Hum Genet* 52:792-798
- Claus EB, Risch NJ, Thompson WD (1990) Age at onset as an indicator of familial risk of breast cancer. *Am J Epidemiol* 131:961-972
- (1991) Genetic analysis of breast cancer in the Cancer and Steroid Hormone Study. *Am J Hum Genet* 48:232-242
- Cohen BB, Porter DE, Wallace MR, Carothers A, Steel CM (1993) Linkage of a major breast cancer gene to chromosome 17q12-21: results from 15 Edinburgh families. *Am J Hum Genet* 52:723-729
- Elston RC (1973) Ascertainment and age of onset in pedigree analysis. *Hum Hered* 23:105-112
- Fain PR, Solomon E, Ledbetter DH (1991) Second International Workshop on Human Chromosome 17. *Cytogenet Cell Genet* 57:66-77
- Feunteun J, Narod SA, Lynch HT, Watson P, Conway T, Lynch J, Parboosingh J, et al (1993) A breast-ovarian cancer susceptibility gene maps to chromosome 17q21. *Am J Hum Genet* 52:736-742
- Fraumeni JF, Grundy GW, Creagan ET, Everson RB (1975) Six families prone to ovarian cancer. *Cancer* 36:364-369
- Gardner EJ, Stephens FE (1950) Breast cancer in one family group. *Am J Hum Genet* 2:30-40
- Goldgar DE, Cannon-Albright LA, Oliphant A, Ward JH, Linker G, Swenson J, Tran TD, et al (1993) Chromosome 17q linkage studies of 18 Utah breast cancer kindreds. *Am J Hum Genet* 52:743-748
- Hall JM, Friedman L, Guenther C, Lee MK, Weber JL, Black DM, King M-C (1992) Closing in on a breast cancer gene on chromosome 17q. *Am J Hum Genet* 50:1235-1242
- Hall JM, Lee MK, Morrow J, Newman B, Anderson L, Huey B, King MC (1990) Linkage analysis of early onset familial breast cancer to chromosome 17q21. *Science* 250:1684-1689
- Human Gene Mapping 11 (1991) *Cytogenet Cell Genet* 58:1-2200
- Iselius L, Slack J, Littler M, Morton NE (1991) Genetic epidemiology of breast cancer in Britain. *Ann Hum Genet* 55:151-159
- Kelsey JL (1979) A review of the epidemiology of human breast cancer. *Epidemiol Rev* 1:74-109
- Lander ES, Green P (1987) Construction of multilocus linkage maps in humans. *Proc Natl Acad Sci USA* 84:2363-2327
- Lathrop GM, Lalouel JM (1984) Easy calculations of lod scores and genetic risks on small computers. *Am J Hum Genet* 36:460-465
- Lathrop GM, Lalouel JM, Julier C, Ott J (1984) Strategies for multilocus linkage analysis in humans. *Proc Natl Acad Sci USA* 81:3443-3446
- Lynch HT, Harris RE, Guirgis HA, Maloney KM, Carmody LL, Lynch JF (1978) Familial association of breast/ovarian cancer. *Cancer* 41:1543-1549
- MacKay K, Hawkins JR, Superti-Furga A, Steinmann B, Dalgleish R (1990) A Hae III RFLP in COL1A1. *Nucleic Acids Res* 18:5926
- Margaritte P, Bonaiti-Pellie C, King M-C, Clerget-Darpoux F (1992) Linkage of familial breast cancer to chromosome 17q21 may not be restricted to early-onset disease. *Am J Hum Genet* 50:1231-1234
- Mérette C, King M-C, Ott J (1992) Heterogeneity analysis of breast cancer families by using age at onset as a covariate. *Am J Hum Genet* 50:515-519
- Mettlin C, Croghan I, Natarajan N, Lane W (1990) The association of age and familial risk in a case-control study of breast cancer. *Am J Epidemiol* 131:973-983
- Narod SA, Feunteun J, Lynch HT, Watson P, Conway T, Lynch J, Lenoir GM (1991) Familial breast-ovarian cancer locus on chromosome 17q12-23. *Lancet* 338:82-83
- Newman B, Austin MA, Lee M, King M-C (1988) Inheritance of human breast cancer: evidence for autosomal dominant transmission in high risk families. *Proc Natl Acad Sci USA* 85:3044-3048
- Ott J (1978) A simple scheme for the analysis of HLA linkages in pedigrees. *Ann Hum Genet* 42:255-257
- (1992) Strategies for characterizing highly polymorphic markers in human gene mapping. *Am J Hum Genet* 51:283-290

- Polymeropoulos MH, Rath DS, Xiao H, Merrill CR (1991) A simple sequence repeat at the human growth hormone locus. *Nucleic Acids Res* 19:689
- Risch N (1984) Segregation analysis incorporating linkage markers. I. Single-locus models with an application to type I diabetes. *Am J Hum Genet* 36:363–386
- Schildkraut JM, Thompson WD (1988) Familial ovarian cancer: a population based series. *Am J Epidemiol* 128:456–466
- Schildkraut JM, Risch N, Thompson WD (1989) Evaluating genetic association among ovarian, breast and endometrial cancer: evidence for a breast-ovarian cancer relationship. *Am J Hum Genet* 45:521–529
- Smith SA, Easton DF, Ford D, Peto J, Anderson K, Averill D, Stratton M, et al (1993) Genetic heterogeneity and localization of a familial breast-ovarian cancer gene on chromosome 17q12-q21. *Am J Hum Genet* 52:767–776
- Solomon E, Ledbetter DH (1991) Report of the Committee on the Genetic Constitution of Chromosome 17. *Cytogenet Cell Genet* 58:686–738
- Spurr NK, Kelsell DP, Black DM, Murday VA, Turner G, Crockford GP, Solomon E, et al (1993) Linkage analysis of early-onset breast and ovarian cancer families, with markers on the long arm of chromosome 17. *Am J Hum Genet* 52:777–785
- Varesco L, Caligo MA, Simi P, Black DM, Nardini V, Casarino L, Rocchi M, et al (1992) The NM23 gene maps to human chromosome band 17q22 and shows a restriction fragment length polymorphism with Bgl II. *Genes Chromosom Cancer* 4:84–88
- Williams WR, Anderson DE (1984) Genetic epidemiology of breast cancer: segregation analysis of 200 Danish pedigrees. *Genet Epidemiol* 1:7–20