Am. J. Hum. Genet. 52:743–748, 1993

Chromosome 17q Linkage Studies of 18 Utah Breast Cancer Kindreds

D. E. Goldgar, L. A. Cannon-Albright, A. Oliphant, J. H. Ward, G. Linker, J. Swensen, T. D. Tran, P. Fields, P. Uharriet, and M. H. Skolnick

Genetic Epidemiology Group, University of Utah, Salt Lake City

Summary

In this paper we present linkage results from the analysis of 18 Utah breast cancer kindreds, for three 17q markers. Four kindreds had LOD scores greater than 1.0 for at least one of the marker loci. One of these kindreds has a LOD score of 6.07 with D17S579, and we believe it to be the most informative 17q family reported to date. Among the kindreds which appear unlinked to 17q were an early-onset breast cancer family, a large breast-ovarian family, and a kindred with mixed age at onset. Analysis of individual recombinants in the linked families localizes the BRCA1 gene between THRA1 and D17S579 (Mfd188). A comparison of the Cancer and Steroid Hormone Study (CASH) model and a model which assumes a rare dominant susceptibility locus with low penetrance and no phenocopies stresses the difficulties in assessing linkage if the assumptions of the CASH model in terms of age at onset of breast cancer are not appropriate for the BRCA1 locus. A hypothetical breast cancer pedigree is used to calculate gene carrier probabilities under the CASH model, thereby illustrating some of our concerns regarding the use of this model to detect and exclude 17q linkage in breast cancer families.

Introduction

Breast cancer is the most common cancer among American women and has long been recognized to be, in part, a familial disease (Anderson 1972). Studies have shown that a woman's risk of developing breast cancer is increased if one or more first-degree relatives have had breast cancer. A recent analysis of the inherited susceptibility of breast cancer (Claus et al. 1991) in the large data set collected as part of the Cancer and Steroid Hormone Study (CASH) demonstrated a rare dominant susceptibility allele for breast cancer. This population-based study collected family history information on first-degree relatives in a series of 4,730 breast cancer probands diagnosed between the ages of 20 and 54 years. The increased risk due to this susceptibility allele ranged from almost 100-fold in women in their 20s to a modest, twofold increase in women in their 80s. Hall et al. (1990) have shown that breast cancer susceptibility in some kindreds with early age at onset is linked to chromosome 17q. In another study, three of five kindreds with both breast cancer and ovarian cancer confirm the reported linkage (Narod et al. 1991).

Our initial studies of breast cancer were statistical studies of the Utah Population Database, the genealogy of the descendants of the Utah pioneers, linked to the Utah Cancer Registry. This data set allowed us to ascertain large cancer-prone kindreds from a well-defined population. Starting with genealogical records kept in the Utah Genealogical Society, we constructed a genealogy of 1.5 million Utah descendants of approximately 10,000 Mormon pioneers (Skolnick 1980). The Utah cancer registry has been part of the National Cancer Institute's Surveillance, Epidemiology, and End Results system since its inception in 1973. It began statewide efforts in the 1960s, with fairly complete incidence data available since 1966; it currently has more than 113,000 cases. Links were created between the genealogy and the cancer registry whenever unique identity could be established. This resource allows us to obtain informa-

Received June 15, 1992; final revision received December 1, 1992.

Address for correspondence and reprints: David E. Goldgar, Ph.D., Genetic Epidemiology, 420 Chipeta Way, Suite #180, Salt Lake City, UT 84108.

0002-9297/93/5204-0012$02.00

tion on a proband's relatives before beginning a study, specifically their cancer status, family size, and location within the state. In this study we examine linkage to 17q markers for 18 kindreds ascertained for multiple cases of breast or ovarian cancer and discuss the implications of the CASH model for detecting or excluding linkage of breast cancer to genetic markers.

Methods

The 18 families which are reported in this paper were ascertained from a number of sources. Fourteen of these families were submitted to the Breast Cancer Linkage Consortium (BCLC) and are included in the analyses reported in the consortium summary paper (Easton et al. 1993), although it should be noted that, because of substantial work done on these families after the final submission of data to the BCLC, there may be differences between the results obtained in the analyses reported here and those in the consortium summary paper. Four additional families were ascertained too recently for BCLC submission. Moreover, genotypic data for the marker D17S579 (MFD188) included in this report were not available for the consortium analysis. These 18 families were ascertained as follows.

Ten families which had three or more cases with inferrable genotypes were selected for linkage studies to 17q markers, from a set of 29 families originally ascertained for a study of proliferative breast disease and breast cancer (Skolnick et al. 1990). Additionally, we included two families (K1001 and K9018) which we have studied since 1980, as part of our breast cancer linkage studies; five families (K2019, K2073, K2079, K2080, and K2082) ascertained from the Utah Population Database for the presence of clusters of breast and/or ovarian cancer; and a self-referred family (K2035) with early-onset breast cancer. These 18 families were investigated and expanded in our clinic and were typed for chromosome 17q markers. Table 1 displays the characteristics of the 18 families included in the present study.

Analysis of these families, as well as other families contributed to the consortium, localized the breast cancer locus between two polymorphic CA-repeat markers: 42D6 (D17S588), a CA repeat isolated in our laboratory; and Mfd15 (D17S250), a CA repeat provided by Jim Weber (Weber et al. 1990). LOD scores of each family, with each of these two markers and with a third marker, Mfd188 (D17S579; Hall et al. 1992), located roughly midway between these two markers, were calculated at a recombination fraction of .001, under two models for the breast cancer locus. The first model was derived by Claus et al. (1991) and has an estimated gene frequency of $q = .0033$, a lifetime risk in gene carriers of about .80, and age-specific risks for breast cancer in nongene carriers. In the second model (affected only [AO]), all cases are assumed to be gene carriers of a rare autosomal dominant gene ($q = .0001$), while all unaffected individuals are assumed to be of unknown phenotype, regardless of age. While the no-phenocopies assumption implicit in the AO model is certainly unrealistic for a common disease such as breast cancer, the model is useful for identifying potential recombinants and is better suited to evaluate linkage in families with later age at onset. For the three markers used for the LOD score calculations, allele frequencies were calculated from our own laboratory typings of unrelated individuals in the CEPH panel. In some of the families, additional polymorphic short tandem repeat markers in the region were typed in order to clarify haplotypes; these markers included 26C2 (D17S514; Oliphant et al. 1991), THRA1 (Futreal et al. 1992), and three polymorphisms—NM23, SCG40 (D17S181), and 6C1 (D17S293)—provided to us by Dr. Donald Black (Hall et al. 1992). The order of these loci which was assumed in constructing haplotypes is centromere–26C2–Mfd15–THRA1–Mfd188–SCG40–6C1–42D6–NM23–telomere; this order is based on our own CEPH analysis, as well as on data reported at the 1992 chromosome 17 workshop (Fain 1992).

Results

Table 2 shows the results of the pairwise linkage analysis of each family, with the three markers 42D6, Mfd188, and Mfd15. When a LOD score greater than 1.0 for at least one locus under the CASH model is used as the criterion for linkage to 17q, it would appear that, of these 18 families, 4 (K1901, K1925, K2035, and K2082) are linked to 17q. Three of these kindreds had an informative recombinant, and these will be detailed below.

Kindred 2082 is the largest 17q-linked breast cancer family reported, to date, by any group. The kindred contains 18 cases of breast cancer, 8 cases of ovarian cancer, and 2 cases with both ovarian cancer and breast cancer. The evidence of linkage to 17q in this family is overwhelming; the LOD score with the linked haplotype is greater than 6.0, despite the existence of three cases of breast cancer which appear to be sporadic, that is, they share no part of the linked haplotype from MFD15–42D6. These three sporadic cases were diagnosed with breast cancer at ages 46, 47, and 54 years.

Table 1

Description of the 18 Kindreds

Kindred[a]	No. of Individuals		Breast Cancer				Ovarian Cancer			
				Age at Diagnosis (years)				Age at Diagnosis (years)		
	Total	Typed	No. Affected	Minimum	Median	Maximum	No. Affected	Minimum	Median	Maximum
1910	15	10	4	27	34	49	...	...	...	...
1001	133	98	13	28	37	64	...	...	...	...
2035	42	25	8	28	37	45	1	...	60	...
2027	21	11	4	34	38	41	...	...	...	...
9018	54	17	9	30	40	72	2	46	48	50
1925	50	27	4	39	42	53	...	...	...	...
1927	49	29	5	32	42	51	...	...	...	...
1911	28	21	7	28	42	76	...	...	...	...
1929	16	11	4	34	43	73	...	...	...	...
1901	35	19	10	31	44	76	...	...	...	...
2082	180	105	20	27	47	67	10	45	52	66
2019[b]	42	19	10	42	53	79	...	...	...	...
1900	70	23	8	45	55	70	1	...	78	...
2080[b]	264	74	22[c]	27	55	92	4	45	53	71
2073[b]	57	29	9	35	57	80	...	...	...	...
1917	16	6	4	43	58	61	...	...	...	...
1920	22	14	3	62	63	68	...	...	...	...
2079[b]	136	18	14	38	66	84	4	52	59	65

NOTE.—Three women diagnosed with both breast cancer and ovarian cancer are counted in both categories.

[a] Kindreds are listed in ascending order of median age at diagnosis of breast cancer.

[b] Not included in collaborative data set.

[c] Includes one case of male breast cancer.

The key recombinant in this pedigree is a woman who developed ovarian cancer at age 45 years and whose mother and aunt had ovarian cancer at ages 58 and 66 years, respectively. She inherited the linked portion of the haplotype for all typed loci distal to Mfd15, including Mfd188, SCG40, 42D6, and NM23, while inheriting unlinked alleles at Mfd15; THRA1 was not informative in this sibship. This recombinant event places BRCA1 distal to Mfd15.

K1901 is typical of early-onset breast cancer families seen by other groups (Hall et al. 1990). The kindred contains 10 cases of breast cancer with a median age at diagnosis of 43.5 years; four cases were diagnosed under age 40 years. The LOD score of this kindred, with the marker 42D6, is 1.5; if the prior probability of 17q linkage for site-specific breast cancer (Easton et al. 1993) is assumed to be .45, this corresponds to a posterior probability of .96 for linkage to this marker. Examination of multilocus haplotypes in this pedigree identifies a recombinant haplotype in an obligate male carrier and his affected daughter, who was diagnosed with breast cancer at age 45 years. Their haplotype for markers Mfd15 and THRA1 differs from that found in all other cases in the family (except one case which could not be completely inferred from her children). The two haplotypes are identical for loci below THRA1, including Mfd188, SCG40, 6C1, and 42D6. Accordingly, kindred 1901 would place the BRCA1 locus distal to THRA1.

Kindred 2035 is similar, in phenotype, to K1901. The median age at diagnosis for the eight cases of breast cancer in this family is 37 years. One case also had ovarian cancer at age 60 years. The breast cancer cases in this family descend from two sisters who were both unaffected with breast cancer until their death in their 8th decade. Each branch contains four cases of breast cancer, with at least one case in each branch having markedly early onset. This family has a LOD score of 2.34, with Mfd15. The haplotypes segregating with breast cancer in the two branches share an identical allele at 26C2, Mfd15, and THRA1, but differ for loci distal to THRA1, i.e., Mfd188, SCG40, 6C1, and

Table 2

LOD Scores of Utah Families, with 42D6, Mfd188, and Mfd15, under Two Genetic Models of Breast Cancer: CASH and AO

	LOD Score[b]					
	42D6 (D17S588)		Mfd188 (D17S579)[c]		Mfd15 (D17S250)	
Kindred[a]	CASH	AO	CASH	AO	CASH	AO
1910	.06	.30	.06	.30	.06	.30
1001	−.30	−1.40	NT	NT	−.52	−3.93
2035	2.34	2.24	.94	−.80	2.34	2.29
2027	−1.22	−2.22	−1.20	−2.23	−1.16	−2.14
9018	−.54	−1.57	−.17	−.18	.11	.13
1925	1.08	.82	.55	.28	−.11	.17
1927	−.41	−1.50	−.35	−1.46	−.44	−1.58
1911	−.27	−1.69	−.43	−1.68	.49	.75
1929	−.49	−2.32	NT	NT	−.49	−2.32
1901	1.50	1.42	.78	.91	.65	−1.31
2082	4.25	3.83	6.07	2.88	2.00	−2.43
2019	−.10	−1.39	−.11	−.58	−.18	−4.06
1900	−.14	−2.69	NT	NT	−.12	−1.67
2080	−.16	−3.77	.76	−4.12	−1.25	−7.81
2073	−.41	−4.97	.63	−.39	−.23	−3.02
1917	−.02	−.30	NT	NT	−.01	.05
1920	−.03	−2.32	NT	NT	.00	.13
2079	.02	−1.89	−.01	−2.10	.01	−1.99

[a] Kindreds are listed in ascending order of median age at diagnosis of breast cancer.
[b] A recombination fraction of .001 between the breast cancer locus and each marker is assumed.
[c] NT = kindred not typed.

NM23. Although the two haplotypes are concordant for marker 42D6, it is likely that the alleles are shared identical by state rather than identical by descent, since the shared allele is the second most common allele at this locus. By contrast, the linked allele shared at Mfd15 has a frequency of .04. This is perhaps the key recombinant in the set, as it is the sole recombinant in which breast cancer segregated with the proximal portion of the haplotype, thus setting the distal boundary. For this event not to be a key recombinant requires that a second mutant BRCA1 gene be present in a married-in spouse who also shares the rare Mfd15 allele segregating with breast cancer in both branches of the family. The evidence from this family therefore places the BRCA1 gene proximal to Mfd188.

One additional family shows some evidence of linkage to 17q. Kindred 1911 gives a modest positive LOD score, with Mfd15, under both models but shows evidence of recombination with both Mfd188 and 42D6. Whether this reflects either a recombination event between BRCA1 and the distal markers or lack of informativeness in the key individual cannot be determined.

Among the families which appear not to be linked to 17q, K1001, K2027, and K2080 are of particular interest. K1001 is a large kindred with generally early onset, in which eight cases of breast cancer were typed or could be inferred for markers in the region. There are at least three unique haplotypes present among the breast cancer cases in this family, although the LOD score under the CASH model is essentially uninformative. Linkage to 17q in this family requires that three cases of breast cancer with ages at diagnosis of 49, 54, and 61 years be of sporadic origin. Kindred K2027 contains four cases of early-onset breast cancer (age at diagnosis 34–41 years) with a recombinant which excludes the breast cancer susceptibility locus from the Mfd15-42D6 region of 17q. It appears unlinked to all three markers, even under the CASH model. Kindred K2080 is a large kindred ascertained from the Utah Population Database for the presence of breast and ovarian cancer. Although this kindred has four cases of breast cancer diagnosed between the ages of 27 and 45 years, the kindred is characterized by largely postmenopausal breast cancer, a smaller degree of postmenopausal

ovarian cancer, and a case of male breast cancer. This family gives highly negative LOD scores for all three markers under the AO model but yields a LOD score of 0.76 for Mfd188 under the CASH model. The structure of this family is quite fragmented, and therefore its linkage to 17q is difficult to assess directly. However, of the 14 cases of breast cancer which could be typed or inferred, the allele most often shared among the cases was present seven times for each marker. For Mfd188, the most commonly shared allele is quite rare; this undoubtedly accounts for the observed positive LOD score. In this family, if the bulk of breast cancer cases are due to a single susceptibility locus, the inherited susceptibility does not appear to be due to a gene located on 17q, although the moderate positive LOD score obtained under the CASH model for Mfd188 may warrant further investigation. Although genotyping is in the early stages in family K2079, it shows a phenotypic and linkage pattern similar to that of kindred K2080.

Discussion

In our 18 families, we estimate that 17q-linked, 17q-unlinked, and families which are currently uninformative are present in roughly equal numbers. Surprisingly, of the six families which contain at least one case of ovarian cancer, two are almost certainly linked to 17q, while perhaps three are unlinked and one is largely uninformative. It is noteworthy that the two apparently linked breast-ovarian families have relatively early onset of breast cancer. Thus, the presence of early-onset breast cancer may be a more distinguishing feature of the 17q-linked susceptibility locus than is the presence of ovarian cancer. However, it is important to note that the CASH model under which the data are typically analyzed makes it difficult to assess linkage in later-onset breast cancer families. When the data are analyzed under a model which assumes all cases to be gene carriers, there is no readily apparent relationship between the median age at diagnosis and presence or absence of linkage to 17q. Of the 10 kindreds with median onset age less than 45 years, 5 appear to be unlinked to 17q; in the 8 with later age at diagnosis, there is at least 1 17q-linked family.

Several conclusions emerge from comparing the LOD scores obtained under the two models shown in table 2. First, in the families with both exclusively or primarily early-onset cancer and reasonably large positive LOD scores, there is very little difference in the LOD scores obtained under the two models, suggesting that, under the CASH model, data from old unaffected individuals are not utilized. Second, in families with primarily later onset, there is little linkage information available under the CASH model, even for families which are highly informative under the AO model. Last, and perhaps most significant, when families have a mixed pattern of age at onset (i.e., kindreds K1001, K1911, K1929, K2073, K9018, and K2082), there can be a large discrepancy between the LOD score obtained under the AO model and that obtained under the CASH model. To better separate the effects of not allowing for the phenocopies in the model versus ignoring information from unaffecteds, we recalculated the LOD scores under a model which assumed that all unaffected individuals were of unknown phenotype but otherwise used the CASH model. In our largest linked family (K2082), the LOD score for Mfd188 was reduced from 6.07 to 4.51. Interestingly, for K2035 the LOD score for Mfd188 was reduced, from 0.94 to 0.04, indicating that most of the positive evidence for linkage in this family (which we believe to contain a Mfd188 recombinant) came from the unaffected individuals, under the CASH model. Among the apparently unlinked families, K2080 had large negative LOD scores for all three markers under the AO model but a LOD of 0.76 for Mfd188, while the LOD scores ignoring information coming from unaffected individuals are 0.30, 0.39, and −2.14 for 42D6, Mfd188, and Mfd15, respectively. Thus evidence for linkage is decreased for two of the markers and is slightly increased for the third.

To further illustrate the consequences of the CASH model, consider the small hypothetical pedigree shown in figure 1. This family has a mixed age-at-onset distribution, with ages at diagnosis ranging from 32 to 62 years. We have calculated the genotypic probabilities under the CASH model as implemented in the LINKAGE package (Lathrop et al. 1985), for individuals 4 and 10 in this family, by assuming various ages and ages at diagnosis, respectively. When the age at diagnosis of individual 10 was varied, the age of individual 4 was set at 79 years; when individual 4 was considered, the age at diagnosis of individual 10 was fixed at 51 years. These probabilities are calculated solely on the basis of phenotypes and pedigree structure; no linkage information is incorporated. It appears counterintuitive that individual 10, who is assumed to have an age at diagnosis of 51 years and an affected mother and two early-onset affected cousins, has only a slightly better than 50% chance of carrying a breast cancer susceptibility allele. Note that, if individual 4 were unaffected at 69 years, she would still have a 17% chance of carrying a breast cancer gene. This demonstrates (*a*) why it is diffi-

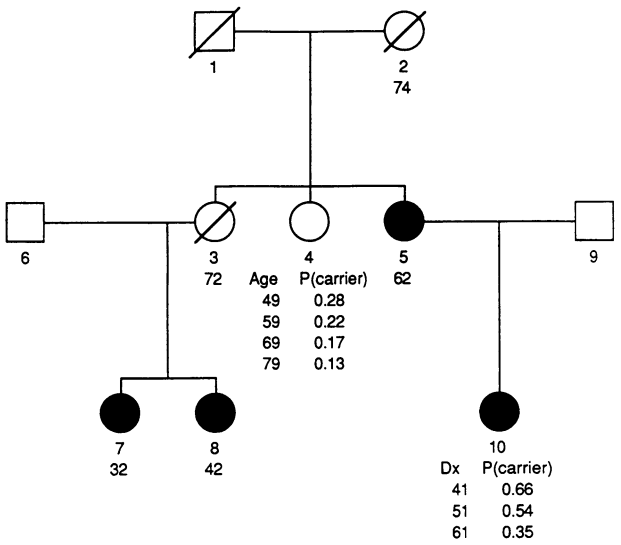

Figure 1 Hypothetical pedigree with familial breast cancer. Beneath each pedigree symbol is the individual number and current age (if unaffected and living), age at death (if deceased), or age at diagnosis (if affected). For individuals 4 and 10 in the pedigree, risks of carrying a breast cancer susceptibility allele under the CASH model are given, for various ages or ages at diagnosis.

cult to exclude linkage in families with either late or mixed age at onset under the CASH model and (*b*) why unaffected individuals contribute little linkage information under this model.

There are at least three distinct susceptibility loci responsible for inherited breast cancer: the p53 locus on 17p, the BRCA1 locus on 17q, and at least one locus responsible for the unlinked residual. It is unlikely that these loci would all have the same phenotypic effects. As the CASH data set represents, in some sense, an average of all of these loci, it is not surprising that the model may at times appear inappropriate for the data in a specific pedigree. We believe that the identification of large extended pedigrees in which linkage can be clearly established and in which key recombinant individuals can be identified will be of great value in refining the localization of the BRCA1 locus, a step which we hope will rapidly lead to its cloning and isolation. The analysis of the pedigrees reported in this paper, in which we have localized the BRCA1 locus between THRA1 and Mfd188, represents a first step in this direction. Once the susceptibility gene is cloned and the mutations are characterized, we will be able to examine more accurately both the phenotypic effects of the 17q-linked breast cancer locus and the interaction of this gene with other risk factors.

Acknowledgments

This work was supported by grants CA-48711, CN-05222, and RR-00064 from the National Institutes of Health.

References

Anderson DE (1972) A genetic study of human breast cancer. J Natl Cancer Inst 48:1029–1034

Claus EB, Risch N, Thompson WD (1991) Genetic analysis of breast cancer in the Cancer and Steroid Hormone Study. Am J Hum Genet 48:232–242

Easton DF, Bishop DT, Ford D, Crockford GP, the Breast Cancer Linkage Consortium (1993) Genetic linkage analysis in familial breast and ovarian cancer: results from 214 families. Am J Hum Genet 52:678–701

Fain PR (1992) Third International Workshop on Human Chromosome 17. Cytogenet Cell Genet 60:178–181

Futreal PA, Barrett JC, Wiseman RW (1992) Dinucleotide repeat polymorphism in the THRA1 gene. Hum Mol Genet 1:66

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, Newman B, Morrow JE, Anderson LA, Huey B, King M-C (1990) Linkage of early-onset familial breast cancer to chromosome 17q21. Science 250:1684–1689

Lathrop GM, Lalouel JM, Julier C, Ott J (1985) Multilocus linkage analysis in humans: detection of linkage and estimation of recombination. Am J Hum Genet 37:482–498

Narod SA, Feunteun J, Lynch HT, Watson P, Conway T, Lynch J, Lenoir GM (1991) Familial breast-ovarian cancer locus on chromosome 17q12-q23. Lancet 338:82–83

Oliphant AR, Wright EC, Swensen J, Gruis NA, Goldgar DE, Skolnick MH (1991) Dinucleotide repeat polymorphism at the D17S514 locus. Nucleic Acids Res 19:4794

Skolnick M (1980) The Utah genealogical data base: a resource for genetic epidemiology. In: Banbury rep no 4: Cancer incidence in defined populations. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp 285–297

Skolnick MH, Cannon-Albright LA, Ward JH, Marshall CJ, Schumann GB, Hogle H, McWhorter WP, et al (1990) Inheritance of proliferative breast disease in breast cancer kindreds. Science 250:1715–1720

Weber JL, Kwitek AE, May PE, Wallace MR, Collins FS, Ledbetter DH (1990) Dinucleotide repeat polymorphisms at the D17S250 and D17S261 loci. Nucleic Acids Res 18:4640