Am. J. Hum. Genet. 62:676–689, 1998

Genetic Heterogeneity and Penetrance Analysis of the BRCA1 and BRCA2 Genes in Breast Cancer Families

D. Ford,[1] D. F. Easton,[3] M. Stratton,[2] S. Narod,[5] D. Goldgar,[6] P. Devilee,[7] D. T. Bishop,[8] B. Weber,[9] G. Lenoir,[6] J. Chang-Claude,[10] H. Sobol,[11] M. D. Teare,[3] J. Struewing,[12] A. Arason,[13] S. Scherneck,[15] J. Peto,[1] T. R. Rebbeck,[9] P. Tonin,[16] S. Neuhausen,[17] R. Barkardottir,[13] J. Eyfjord,[14] H. Lynch,[18] B. A. J. Ponder,[4] S. A. Gayther,[4] J. M. Birch,[19] A. Lindblom,[20] D. Stoppa-Lyonnet,[21] Y. Bignon,[22] A. Borg,[23] U. Hamann,[10] N. Haites,[25] R. J. Scott,[26] C. M. Maugard,[27] H. Vasen,[28] S. Seitz,[15] L. A. Cannon-Albright,[17] A. Schofield,[25] M. Zelada-Hedman,[20] and the Breast Cancer Linkage Consortium

Sections of [1]Epidemiology and [2]Molecular Carcinogenesis, Institute of Cancer Research, Sutton, United Kingdom; [3]CRC Genetic Epidemiology Unit, Institute of Public Health, and [4]CRC Human Cancer Genetics Research Group, University of Cambridge, Cambridge; [5]Center for Research in Women's Health, Toronto; [6]International Agency for Research on Cancer, Lyon; [7]Departments of Pathology and Human Genetics, University of Leiden, Leiden; [8]ICRF Genetic Epidemiology Laboratory, Leeds; [9]Departments of Medicine and Genetics, University of Pennsylvania, Philadelphia; [10]Deutsches Krebsforschungszentrum, Heidelberg; [11]Départment d'Oncologie-Génétique/INSERM CRI9703, Institut Paoli Calmettes, Marseille; [12]Genetic Epidemiology Branch, National Cancer Insitute, Bethesda; [13]Laboratory of Cell Biology, University Hospital of Iceland, and [14]The Icelandic Cancer Society, Reykjavik; [15]Max-Delbrück-Centrum fur Moleculäre Medizin, Berlin; [16]Departments of Medicine and Human Genetics, McGill University, Montreal; [17]Genetic Epidemiology Group, University of Utah, Salt Lake City; [18]Department of Preventative Medicine and Public Health, Creighton University School of Medicine, Omaha; [19]CRC Paediatric and Familial Cancer Research Group, Christie Hospital and Holt Radium Institute, Manchester; [20]Department of Clinical Genetics, Karolinska Institute, Stockholm; [21]Unité de Génétique Oncologique, Institut Curie, Paris; [22]Laboratoire d'Oncologie Moleculaire, Centre Jean Perrin, Clermont-Ferrand, France; [23]Department of Oncology, University Hospital, Lund, Sweden; [24]Medical Genetics, University of Aberdeen, Aberdeen; [25]Humangenetik, Kantonsspital, Basel; [26]Centre Réné Gauducheau, Nantes; and [27]Foundation for the Detection of Hereditary Tumors Leiden

Summary

The contribution of BRCA1 and BRCA2 to inherited breast cancer was assessed by linkage and mutation analysis in 237 families, each with at least four cases of breast cancer, collected by the Breast Cancer Linkage Consortium. Families were included without regard to the occurrence of ovarian or other cancers. Overall, disease was linked to BRCA1 in an estimated 52% of families, to BRCA2 in 32% of families, and to neither gene in 16% (95% confidence interval [CI] 6%–28%), suggesting other predisposition genes. The majority (81%) of the breast-ovarian cancer families were due to BRCA1, with most others (14%) due to BRCA2. Conversely, the majority of families with male and female breast cancer were due to BRCA2 (76%). The largest proportion (67%) of families due to other genes was found in families with four or five cases of female breast cancer only. These estimates were not substantially affected either by changing the assumed penetrance model for BRCA1 or by including or excluding BRCA1 mutation data. Among those families with disease due to BRCA1 that were tested by one of the standard screening methods, mutations were detected in the coding sequence or splice sites in an estimated 63% (95% CI 51%–77%). The estimated sensitivity was identical for direct sequencing and other techniques. The penetrance of BRCA2 was estimated by maximizing the LOD score in BRCA2-mutation families, over all possible penetrance functions. The estimated cumulative risk of breast cancer reached 28% (95% CI 9%–44%) by age 50 years and 84% (95% CI 43%–95%) by age 70 years. The corresponding ovarian cancer risks were 0.4% (95% CI 0%–1%) by age 50 years and 27% (95% CI 0%–47%) by age 70 years. The lifetime risk of breast cancer appears similar to the risk in BRCA1 carriers, but there was some suggestion of a lower risk in BRCA2 carriers <50 years of age.

Introduction

Mutations in a number of genes are now known to cause susceptibility to breast and/or ovarian cancer. In the context of high-risk families, the most important genes are BRCA1 (MIM 113705 [http://www3.ncbi.nlm.nih.gov:80/htbin-post/Omim/dispmim?113705]) and BRCA2 (MIM 600185 [http://www3.ncbi.nlm.nih.gov:80/htbin-

Received June 30, 1997; accepted for publication December 4, 1997; electronically published March 4, 1998.

Address for correspondence and reprints: Dr. Douglas Easton, CRC Genetic Epidemiology Unit, Strangeways Research Laboratories, Worts Causeway, Cambridge CB1 4RN, United Kingdom. E-mail: douglas.easton@srl.cam.ac.uk

0002-9297/98/6202-0023$02.00

post/Omim/dispmim?600185]). BRCA1 was localized to chromosome 17q by genetic linkage in 1990 (Hall et al. 1990) and subsequently was cloned in 1994 (Miki et al. 1994). Studies to date suggest that BRCA1 accounts for the majority of families containing multiple cases of breast and ovarian cancer (Easton et al. 1993; Narod et al. 1995*b*), for less than half the families containing breast cancer only (Easton et al. 1993), and for few families that include male breast cancer cases (Stratton et al. 1994). The risks conferred by BRCA1 have been estimated both from linkage data (Easton et al. 1995) and from the study of cancer incidence in carriers within families with disease believed to be linked to the gene (Ford et al. 1994). The cumulative risk of breast or ovarian cancer in female carriers is estimated to be >80% by age 70 years, although there is evidence that the ovarian:breast cancer risk ratio varies between families, with mutations toward the 5′ end of the gene conferring a relatively higher ovarian cancer risk (Easton et al. 1995; Gayther et al. 1995; Holt et al. 1996).

BRCA2 was localized to chromosome 13q in 1994 (Wooster et al. 1994) and was cloned in 1995 (Wooster et al. 1995; Tavtigian et al. 1996). In their study of 15 families with evidence against linkage to BRCA1, Wooster et al. (1994) estimated that disease was linked to BRCA2 in 74% of families, providing preliminary evidence that BRCA1 and BRCA2 together might account for most high-risk breast cancer families. There is particularly clear evidence that BRCA1 and BRCA2 together are likely to account for the majority of high-risk breast-ovarian cancer families. In a study of 145 breast-ovarian cancer families, there were 10 families with strong evidence against linkage to BRCA1 (multipoint LOD scores <-1.0); of these, 3 now have an identified BRCA1 mutation, and 7 have an identified BRCA2 mutation (Narod et al. 1995*a,* 1995*b*). A number of families with evidence of linkage to 13q and/or with identified BRCA2 mutations contain male cases. Easton et al. (1997) have estimated the risks of cancer in BRCA2 mutation carriers in the two largest families showing linkage in the original study. They found that the breast cancer risk in females was similar to that conferred by a BRCA1 mutation and estimated the risk in males to be 6% by age 70 years. The increased risk of ovarian cancer in mutation carriers appears to be lower in most families than it is in families with a BRCA1 mutation. Analyses of BRCA2 mutation data have provided evidence that the risks of breast and ovarian cancer are related to the position of the mutation: truncating mutations in families with the highest risk of ovarian cancer relative to breast cancer are clustered in a region in the middle of exon 11 (Gayther et al. 1997).

The aims of this collaborative study were twofold: first, to estimate the respective proportions of different types of high-risk cancer families in which the disease is due to BRCA1 and BRCA2 and to determine what proportion of families might be due to unidentified genes; and, second, to estimate the penetrance of BRCA2 in a large data set. Groups in the Breast Cancer Linkage Consortium (BCLC) were asked to submit linkage data on markers flanking BRCA1 and BRCA2, for all families containing at least four cases of either female breast cancer diagnosed at age <60 years or male breast cancer diagnosed at any age. Information on BRCA1 mutation testing, including method and outcome, was collected on all families, and mutation details for families with positive mutation tests for BRCA2 were provided. We report here on the results of analyses of 237 families.

Families and Methods

Families

Families included in the study were all those that had been typed, by the Breast Cancer Linkage Consortium (BCLC), for linkage to BRCA1 and that contained at least four cases of either female breast cancer diagnosed at age <60 years or male breast cancer diagnosed at any age, irrespective of any ovarian cancers in the family. Two hundred thirty-seven families were included, 121 (51%) of which had been included in previous BCLC studies. Sixty-seven of the 94 families with at least one case of ovarian cancer but no male breast cancer (referred to subsequently as "breast-ovarian families") had been included in a BCLC study reported by Narod et al. (1995*b*). Twelve of the 26 families with at least one case of male breast cancer were part of a study by Stratton et al. (1994). Only 42 of the 117 families with no ovarian or male breast cancer were in the first BCLC study, which examined linkage to BRCA1 (Easton et al. 1993).

One hundred eighty of the families had been tested for a BRCA1 mutation, including 67 that had been tested by direct sequencing. Mutations within the BRCA1 coding sequence had been reported in 64 families. Thirty-six families were known to have a BRCA2 mutation.

Genetic Markers

All families included in this study had been typed for markers flanking BRCA1. For the majority of families, D17S579 (Hall et al. 1992) and D17S250 (Weber et al. 1990) typings were available. For a small number of families, data were unavailable for these two markers, and data for other markers in the region—in particular, THRA1 (Bowcock et al. 1993) and D17S855 (Gyapay et al. 1994)—were used. The 17q marker typings had not been scored consistently across the families, so, in the analysis, each marker was coded as a system of equally frequent alleles (with the number of alleles chosen to reflect the observed polymorphism of the marker).

One hundred seventy-seven families had, in addition, been typed for markers D13S260 and D13S267, which flank BRCA2 (Wooster et al. 1994). Thirty-five families that had not been typed for linkage to BRCA2 had a mutation in the BRCA1 gene. An additional six families showed strong evidence for linkage to BRCA1 (LOD score >1.0). The remaining 19 families had not been studied for linkage to BRCA2, either because the investigating group concerned had no remaining DNA from critical family members or because it was not feasible because of other reasons. 13q marker typings were scored consistently across families, and the allele frequencies used for analysis were taken from the Genome Database, in which frequencies are based on 56 CEPH chromosomes.

Statistical Methods

Estimating the proportions of families in which the disease is due to BRCA1 and BRCA2.—The proportions of families with disease due to BRCA1 and BRCA2 were estimated by computation of heterogeneity LOD scores. In our first analysis, we assumed that the same breast cancer risks were conferred by mutations in BRCA1, BRCA2, and other susceptibility genes. These risks were based on a "standard" genetic model for breast cancer, derived by Claus et al. (1991), from the Cancer and Steroid Hormone Study (the "CASH" model). Under this model, susceptibility to breast cancer in females is conferred by an autosomal dominant gene with population frequency .003, and the risk of breast cancer at age <70 years is 67% in carriers. The penetrance for ovarian cancer in carriers of the susceptibility gene was derived by assuming a constant relative risk with age and an estimated cumulative risk of 10% by age 60 years (Claus et al. 1993).

In our second analysis, BRCA1 was assumed to confer a penetrance higher than that in the CASH model, as had been suggested by previous studies: a cumulative breast cancer risk in females of 49% by age 50 years and 71% by age 70 years and a cumulative ovarian cancer risk of 16% by age 50 years and 42% by age 70 years (Easton et al. 1993; Narod et al. 1995*b*). BRCA2 and any other genes were still assumed to confer the CASH risks. In the present analysis, we estimate the relative frequency of mutations in BRCA1, BRCA2, and other genes and use these estimates to derive the proportions of families in which the disease is linked to each gene. Computational details are given in Appendix A.

A small number of families ($n = 37$) were not typed for BRCA2 markers, despite having LOD scores <1.0 at BRCA1, and, in some cases ($n = 18$), this was related to having a BRCA1 mutation. This could lead to some bias in the estimates derived by the methods described above, which do not incorporate BRCA1 mutation status. We therefore performed some further analyses, in which the BRCA1 mutation status, as well as the linkage data, were used to estimate the proportion of families in which the disease is due to BRCA1 and BRCA2. In order to do this, an extra parameter, δ, was introduced, to model the probability of identifying a BRCA1 mutation when one existed (see Appendix B). The sensitivity parameter δ is likely to depend on the method used for detection of mutations. The method of detection here refers to the method used initially to screen the coding sequence and splice sites for alterations; all investigating groups used direct sequencing to identify the precise mutation in individuals with screening abnormalities. All groups except two (CRC and ICRF, where the testing was split between more than one laboratory) used one technique consistently. We first estimated the sensitivity of completely sequencing the coding sequence of the BRCA1 gene (Miki et al. 1994) of at least one affected individual (δ_1) and the sensitivity of the other genomic screening techniques taken together (δ_2)—namely, constant-denaturating gel electrophoresis (CDGE; Stoppa-Lyonnet et al. 1997), confirmation-sensitive gel electrophoresis (CSGE; Ganguly et al. 1993) , direct screening for deletions and inversions, in combination with protein-truncation analysis (DSDI/PTT; Hogervorst et al. 1995; Peelen et al. 1997), SSCP-heteroduplex analysis (SSCP-HA; Gayther et al. 1996), and SSCP alone. Since δ_1 and δ_2 were estimated identically, subsequent analyses were based on a single estimate for δ, to represent the sensitivity of mutational analysis by any of these methods.

Estimating the penetrance of BRCA2.—The basic method used for estimation of the age-specific risks conferred by BRCA2 was to maximize the LOD score over different values of the penetrance function (Easton et al. 1993). This method is equivalent to maximizing the likelihood conditional on all phenotypic data and hence is free from the ascertainment bias caused by family selection on the basis of multiple affected individuals. In the first analysis, we used the families with an identified mutation in BRCA2. We then repeated the analyses, using all families with six or more cases of either female breast cancer at age 60 years or male breast cancer at any age, assuming that all these families were due to either BRCA1 or BRCA2. The BRCA1 risks were held fixed at the values estimated elsewhere (Easton et al. 1993; Narod et al. 1995*b*). In addition to the BRCA2 risks, an extra parameter was estimated, corresponding to the proportion of all high-risk mutations that are BRCA1 mutations. Further details are given in Appendix C.

Table 1

Heterogeneity Analysis under the CASH Model

Family Group	No. of Families	Proportion (95% CI) of Families with Disease Linked to: BRCA1	BRCA2	Other	LOD Score
All families:					
Four or five breast cancers, male or female	154	.55 (.40–.70)	.12 (.00–.29)	.33 (.14–.53)	19.28
Six or more breast cancers	83	.46 (.32–.60)	.50 (.35–.65)	.04 (.00–.17)	67.44
Overall	237	.52 (.42–.63)	.35 (.24–.46)	.13 (.03–.25)	84.08
All families, no males:					
Four or five breast cancers	140	.61 (.44–.76)	.08 (.00–.25)	.31 (.12–.53)	19.90
Six or more breast cancers	71	.50 (.34–.65)	.44 (.28–.61)	.06 (.00–.21)	50.40
Overall	211	.57 (.45–.68)	.28 (.18–.40)	.15 (.04–.28)	68.13
Male breast cancer	26	.19 (.01–.47)	.77 (.43–.97)	.04 (.00–.42)	18.03
Breast-ovarian cancer, no males:					
One ovarian cancer	42	.69 (.46–.89)	.19 (.04–.41)	.12 (.00–.33)	15.11
Two or more ovarian cancers	52	.88 (.69–.98)	.12 (.02–.31)	.00 (.00–.15)	44.07
Overall	94	.80 (.66–.92)	.15 (.05–.28)	.05 (.00–.17)	58.57
Female breast cancer only:					
Four or five breast cancers	83	.32 (.10–.55)	.09 (.00–.35)	.59 (.26–.89)	2.29
Six or more breast cancers	34	.19 (.04–.41)	.66 (.39–.88)	.15 (.00–.44)	14.18
Overall	117	.28 (.13–.45)	.37 (.20–.56)	.35 (.14–.57)	14.47

Results

Proportions of Families with Disease Due to BRCA1 and BRCA2

Tables 1 and 2 summarize the results of the heterogeneity analyses using BRCA 1 and BRCA2 linkage data only. In table 1 the estimated proportions of families linked to each gene are given, with the CASH model being assumed for all genes. Overall, an estimated 52% of families have disease due to BRCA1, and 35% have disease due to BRCA2. There is significant evidence of families with disease not linked to either gene ($\chi_1^2 = 8.57$; $P = .003$), although the estimated proportion of families with disease due to other genes is only 13%. In families with fewer than six cases of breast cancer, the proportion of families with disease due to other genes (33%) is higher than it is in families with six or more cases (4%), in which the evidence for other genes is, in fact, not significant. In the majority (77%) of families with one or more cases of male breast cancer, the disease is estimated to be due to BRCA2, with a smaller proportion (19%) estimated as being due to BRCA1. The evidence for other susceptibility genes is not significant, although the upper confidence limit on the remaining proportion of families is 42%. Of the families with breast and ovarian cancer but with no male case, the large majority of families are estimated to have disease due to BRCA1 (80%), with a smaller proportion (15%) estimated to be due to BRCA2. The evidence for other genes is weak in this group of families, the upper confidence limit for the proportion of disease due to other genes being 17%. Among the 117 families with female breast cancer only, the proportions of families estimated to have disease due to BRCA1, BRCA2, and other genes were similar (28%, 37%, and 35%, respectively). The estimated proportion of families with disease due to other genes was higher in families with fewer than six cases (59%) than in families with six or more cases (15%, with a lower confidence limit of 0%). Interestingly, in families with six or more cases of breast cancer and no ovarian cancer, the largest proportion of disease was due to BRCA2 (66%, compared with 19% due to BRCA1).

In table 2 the estimated proportions of high- risk mutations, as well as the corresponding estimated proportions of families, attributable to each gene are given, under the assumption that BRCA1 confers the risks estimated in previous consortium analyses whereas BRCA2 and other genes confer risks given by the CASH model. Families containing male breast cancers have been excluded from table 2, since (1) the penetrance in males is uncertain for all genes and (2) estimation of the proportion of high-risk mutations attributable to each gene is strongly dependent on the assumed penetrances. In general, the estimated proportions of families with disease due to each of the genes is very similar in the two analyses, strengthening the support for these estimates. The largest difference is in the families with six or more cases of female breast cancer only, where the estimated proportion due to BRCA1 is higher (29% vs. 19%) and the proportion due to other genes is, correspondingly, lower (5% vs. 15%), when a higher penetrance for BRCA1 is assumed. There is stronger evidence for other predisposition genes in the breast-ovarian can-

Table 2

Heterogeneity Analysis Allowing BRCA1 to Confer Higher Penetrance

Family Group	No. of Families	Proportion (95% CI) of High-Risk Mutations and Families[a]			LOD Score
		BRCA1	BRCA2	Other	
All families, no males:					
Four or five breast cancers	140	.52 (.31–.73)	.06 (.00–.24)	.42 (.19–.67)	23.04
		.56 (.42–.70)	.05 (.00–.19)	.39 (.21–.56)	
Six or more breast cancers	71	.39 (.22–.58)	.51 (.32–.70)	.10 (.00–.28)	55.65
		.50 (.45–.55)	.42 (.34–.49)	.08 (.00–.17)	
Overall	211	.47 (.33–.62)	.31 (.18–.45)	.22 (.09–.38)	76.11
		.54 (.46–.61)	.27 (.19–.35)	.19 (.10–.30)	
Breast-ovarian cancer, no males:					
One ovarian cancer	42	.60 (.31–.86)	.20 (.04–.47)	.20 (.03–.47)	13.58
		.67 (.52–.81)	.16 (.07–.30)	.17 (.05–.30)	
Two or more ovarian cancers	52	.42 (.19–.77)	.52 (.10–.81)	.06 (.00–.52)	46.73
		.80 (.72–.89)	.18 (.05–.27)	.02 (.00–.15)	
Overall	94	.52 (.31–.76)	.29 (.10–.52)	.19 (.03–.42)	59.99
		.74 (.66–.83)	.16 (.08–.24)	.10 (.03–.19)	
Female breast cancer only:					
Four or five breast cancers	83	.43 (.15–.69)	.07 (.00–.30)	.50 (.20–.84)	2.75
		.36 (.16–.55)	.08 (.00–.28)	.56 (.28–.83)	
Six or more breast cancers	34	.35 (.11–.63)	.61 (.33–.85)	.04 (.00–.32)	15.66
		.29 (.19–.38)	.66 (.52–.76)	.05 (.00–.25)	
Overall	117	.42 (.23–.62)	.32 (.16–.49)	.26 (.07–.48)	16.21
		.35 (.24–.47)	.36 (.22–.48)	.29 (.10–.47)	

[a] Within each table cell, the underlined entry is the proportion of high-risk mutations, and the entry below it is the proportion of families.

cer families in this analysis, although the estimated proportion is still small (10%), and, in families with at least two cases of ovarian cancer, there is still no significant evidence of other predisposition genes.

The proportions of high-risk mutations that occur in each gene should not depend on family type, so the estimates should be consistent across subgroups of families, providing that the penetrances of the genes are well estimated. If the families with male cases are excluded, an estimated 47% of high-risk mutations occur in BRCA1 and 31% occur in BRCA2, leaving 22% of mutations in unidentified genes. The estimates do not differ significantly between the breast-ovarian cancer families and the families with female breast cancer only (χ^2_2 =; $P = .81$) or between families with only one ovarian cancer and families with two or more ovarian cancers ($\chi^2_2 = 1.47$; $P = .48$). Estimates do differ, however, between families with fewer than six cases of female breast cancer and families with six or more cases ($\chi^2_2 = 11.88$; $P = .003$), reflecting the difference seen in the families without ovarian cancer, in which 7% of mutations are attributed to BRCA2 and 50% are attributed to other genes if there are fewer than six cases in the family, compared with 61% and 4%, respectively, if there are six or more cases ($\chi^2_2 = 10.13$; $P = .006$).

Analysis Including BRCA1 Mutation Data

One hundred eighty of the families were tested for mutations in BRCA1, the majority by direct sequencing (67 families) or by one or more of the following: CDGE, CSGE, DSDI/PTT, SSCP-HA, and SSCP alone (106 families). One family had been tested by PTT only, and six families had been analyzed by SSCP across a portion of the gene.

In the initial heterogeneity analyses incorporating BRCA1 testing information, the estimated proportion of mutations identified by direct sequencing was allowed to differ from the proportion identified by CDGE, CSGE, DSDI/PTT, SSCP-HA, or SSCP. Both the family tested by PTT and the six families only partially analyzed were assumed not to have been tested. Across all families, the estimated sensitivity of sequencing under the CASH model was .64 (95% CI .45–.84), which was the same as the estimated sensitivity for the other methods (.64 [95% CI .49-.80]). The estimated proportions of families with disease due to BRCA1 (52% [95% CI 42%–62%]), BRCA2 (32% [95% CI 22%-43%]), and other genes (16% [95% CI 6%-28%]) were almost identical to those obtained by use of linkage data only (see table 1). We subsequently reanalyzed the data, assuming a single sensitivity parameter for all the standard screening methods.

Table 3

Heterogeneity Analysis under the CASH Model Incorporating BRCA1 Mutation Data

Family Group	No. of Families			Proportion (95% CI) of Linked Families			Proportion (95% CI) of Mutations Identified[a]
	Total	Tested for BRCA1	Mutation Positive	BRCA1	BRCA2	Other	
All families:							
Four or five breast cancers, male or female	154	121	40	.50 (.36–.65)	.13 (.0–.28)	.37 (.19–.56)	.64 (.46–.84)
Six or more breast cancers, male or female	83	59	24	.50 (.37–.63)	.45 (.31–.58)	.05 (.00–.17)	.69 (.50–.86)
Overall	237	180	64	.52 (.42–.62)	.32 (.22–.43)	.16 (.06–.28)	.63 (.51–.77)
All families, no males:							
Four or five breast cancers	140	114	39	.54 (.40–.69)	.08 (.00–.23)	.38 (.19–.58)	.63 (.45–.82)
Six or more breast cancers	71	52	23	.55 (.41–.69)	.38 (.24–.53)	.07 (.00–.21)	.70 (.51–.87)
Overall	211	166	62	.55 (.45–.66)	.26 (.16–.37)	.19 (.08–.31)	.64 (.51–.77)
Male breast cancer	26	14	2	.16 (.02–.48)	.76 (.43–.97)	.08 (.00–.44)	.67 (.09–1.0)
Breast-ovarian cancer, no males:							
One ovarian cancer	42	39	20	.69 (.50–.86)	.21 (.06–.40)	.10 (.00–.28)	.79 (.56–.96)
Two or more ovarian cancers	52	45	26	.91 (.76–.99)	.09 (.01–.24)	.00 (.00–.11)	.62 (.46–.77)
Overall	94	84	46	.81 (.68–.91)	.14 (.05–.26)	.05 (.00–.16)	.68 (.55–.81)
Female breast cancer only:							
Four or five breast cancers	83	63	11	.28 (.11–.50)	.05 (.00–.29)	.67 (.35–.89)	.61 (.29–1.0)
Six or more breast cancers	34	19	5	.21 (.08–.42)	.60 (.34–.83)	.19 (.01–.45)	1.0 (.41–1.0)
Overall	117	82	16	.26 (.13–.42)	.32 (.17–.50)	.42 (.21–.62)	.68 (.38–1.0)

[a] By sequencing, CDGE, CSGE, DSDI/PTT, SSCP-HA, SSCP, or PTT.

Table 3 describes the heterogeneity results under the CASH model with the incorporation of the BRCA1 mutation data. The proportion of mutations identified refers here to a single estimate for all methods. PTT alone (which had been used in one family) was assumed to be as sensitive as sequencing, CDGE, CSGE, DSDI/PTT, SSCP-HA, or SSCP, and the partial screening by SSCP, which was used in six families, was assumed to be 25% as efficient as the other methods. The results in table 3 show that, when mutation data are incorporated, the estimated proportions of families with linkage to each gene are almost identical to those estimated from linkage data alone. The estimated proportions of mutations being identified is consistent, in all family types, with the overall estimate of 63% (95% CI 51%–77%). We also repeated the analyses under the alternative model, allowing the risks conferred by BRCA1 to be higher than those conferred by BRCA2 or the other genes; again, incorporating the mutation data had no influence on the heterogeneity estimates (data not shown).

Penetrance of BRCA2

The overall penetrance of BRCA2 was estimated in the 32 families in which a BRCA2 mutation was found and in which D13S267 typing was available and informative. Table 4 gives the risks of breast or ovarian cancer in these families, obtained by maximizing the LOD score over the age-specific incidence rates of disease. The estimated overall penetrance is 29% (95% CI 9%–44%) by age 50 years and 88% (95% CI 48%–97%) by age 70 years. Most of this risk is due to breast cancer; among mutation carriers, there are no cases of ovarian cancer at age <40 years and only two cases of ovarian cancer at age <50 years. The estimated breast cancer risk in this group is 28% (95% CI 9%–44%) by age 50 years and 84% (95% CI 43%–95%) by age 70 years. The corresponding ovarian cancer risks are 0.4% (95% CI 0%–1%) by age 50 years and 27% (95% CI 0%–47%) by age 70 years.

Table 5 shows the risks obtained for BRCA2 mutation carriers if families with at least six cases of breast cancer (females at age <60 years or males at any age) are attributed to BRCA1 or BRCA2. (One family was excluded from this analysis, because of a consanguineous relationship that could not be ignored.) In the age group 40–59 years, the estimated overall risks and breast cancer risks are higher than the risks in the mutation families, although the differences do not reach statistical significance (difference in overall cancer risks by age 50 years [$P = .11$]; difference by age 60 years [$P = .50$]), and the risks by age 70 years are very similar in the two analyses (86% overall vs. 88% overall). The risks previously estimated for BRCA1 are also given, for comparison. The estimated risks for BRCA2 are slightly lower by ages 40 years and 50 years and are slightly higher by age 70 years, but they are not significantly different at any age. When only the mutation families are considered, the estimated BRCA2 risk, up to age 50 years, is more markedly lower than is the BRCA1 risk, although, again, the risks by age 70 years are similar.

Table 4

Incidence and Cumulative Risks of Breast and Ovarian Cancer in BRCA2 Mutation Carriers

Age Group (years)	Breast and Ovarian Cancer		Breast Cancer Only			Ovarian Cancer Only		
	Incidence	Cumulative Risk (95% CI)	No.	Incidence	Cumulative Risk (95% CI)	No.	Incidence	Cumulative Risk (95% CI)
20–29	.000633	.006 (0–.019)	7	.000633	.006 (0–.019)	0	.00	.00
30–39	.0118	.12 (0–.24)	61	.0118	.12 (0–.24)	0	.00	.00
40–49	.0215	.29 (.090–.44)	99	.0210	.28 (.090–.44)	2	.000425	.004 (0–.011)
50–59	.0390	.52 (.24–.70)	44	.0318	.48 (.22–.65)	10	.00722	.074 (0–.15)
60–69	.142	.88 (.48–.97)	20	.118	.84 (.43–.95)	4	.0236	.27 (0–.47)

Note.—All data are for first cancers; the number exclude second primaries and known noncarriers of mutation or "linked"-haplotype.

Table 6 provides the risks for the BRCA2 mutation families, according to mutation location. The gene has been divided into three regions, the middle region being the ovarian cancer–cluster region (OCCR) suggested by Gayther et al. (1997). One family with a missense mutation was excluded from this analysis. The estimated breast cancer risks appear to be similar in all three groups. The estimated ovarian cancer risk is, however, higher in the OCCR, in agreement with the results of Gayther et al. (1997), although the confidence limits are very wide.

Discussion

This report describes the analysis of the largest collection of breast cancer families in the world, for linkage to BRCA1 and BRCA2. Two hundred thirty-seven families with four or more cases of either female breast cancer diagnosed at age <60 years or male breast cancer at any age were contributed by 21 investigating groups from nine countries. All families were typed with genetic markers flanking BRCA1, and 177 families were typed with genetic markers flanking BRCA2.

As had been suggested by other BCLC studies, almost all breast-ovarian cancer families appear to be compatible with linkage to BRCA1 or BRCA2 (Narod et al. 1995*a,* 1995*b*). Only 2 families (CJP3 and MDC60) of the 94 breast-ovarian cancer families had multipoint LOD scores of <-1.0 at both loci; both these families had been screened for germ-line BRCA1 mutations, although neither was sequenced, and neither had been tested for a mutation in BRCA2. Each of these two families had only one case of ovarian cancer. One of these families (MDC60) recently has been shown, tentatively, to have disease linked to markers on chromosome 8p (Seitz et al. 1997), a location for a breast cancer–susceptibility gene first suggested by Kerangueven et al. (1995). There was very little evidence for another predisposition gene in the 52 families containing two or more cases of ovarian cancer; under the CASH model the point estimate for the proportion of remaining families was 0.0, and under the model allowing BRCA1 to confer a higher penetrance it was 2%.

This study has confirmed that the large majority of families containing both male and female breast cancer have disease that is due to BRCA2. In slight contrast to previous linkage studies (Stratton et al. 1994), there is

Table 5

Estimated Penetrance of BRCA2 and Comparison with Penetrance of BRCA1

Age (years)	Penetrance (95% CI)					
	BRCA2					
	32 Families with BRCA2 Mutation		82 Families with Six or More Breast Cancers[b]		BRCA1[a]	
	Breast and Ovarian —Cancer	Breast Cancer —Only	Breast and Ovarian —Cancer	Breast Cancer —Only	Breast and Ovarian —Cancer	Breast Cancer —Only
30	.006 (0–.019)	.006 (0–.19)	.046 (.004–.086)	.046 (.004–.086)	.036 (0–.14)	.036 (.0–.14)
40	.12 (0–.24)	.12 (0–.24)	.12 (.048–.18)	.12 (.048–.18)	.18 (0–.36)	.18 (.0–.35)
50	.29 (.090–.44)	.28 (.090–.44)	.46 (.31–.57)	.45 (.31–.57)	.57 (.33–.73)	.49 (.28–.64)
60	.52 (.24–.70)	.18 (.22–.65)	.61 (.39–.76)	.59 (.37–.73)	.75 (.53–.87)	.64 (.43–.77)
70	.88 (.48–.97)	.84 (.43–.95)	.86 (.25–.98)	.83 (.27–.96)	.83 (.65–.92)	.71 (.53–.82)

[a] Source: Narod et al. (1995*b*).

[b] Proportion of BRCA2 mutations is .84 (95% CI .56–1.0) when it is assumed that BRCA1 and BRCA2 account for cancers in all families. Corresponding proportion of families with BRCA2 mutations is .50 (95% CI .39–1.0).

Table 6

Cumulative Risks of Breast and Ovarian Cancer in BRCA2 Mutation Carriers, by Mutation Location

	Cumulative Risk (95%CI)					
	Exons 1–10 and 11, Nucleotides 1–1034		Exon 11, Nucleotides 1035–6629		Exon 11, Nucleotides 6630–end and Exons 12–23	
Age	Breast Cancer	Ovarian Cancer	Breast Cancer	Ovarian Cancer	Breast Cancer	Ovarian Cancer
30	.0011 (0–.0042)	0	.016 (0–.10)	0	.017 (0–.071)	0
40	.14 (0–.34)	0	.090 (0–.30)	0	.10 (0–.25)	0
50	.26 (0–.47)	.0036 (0–.12)	.27 (0–.62)	0	.32 (0–.56)	.0063 (0–.021)
60	.41 (.053–.63)	.053 (0–.13)	.47 (0–.82)	.24 (0–.64)	.60 (.12–.82)	.052 (0–.13)
70	.82 (0–.97)	.26 (0–.53)	.83 (0–.99)	.76 (0–.98)	.78 (0–.97)	.052 (0–13)

now clear evidence, both from linkage data and from mutation data (two families had identified mutations), that a proportion of these families are the result of BRCA1 mutations; 16% (95% CI 2%–48%) of families with a male case were estimated to be due to BRCA1. Of the seven families with two or more cases of male breast cancer, four had BRCA2 mutations, and data for one additional family were consistent with linkage to BRCA2 (LOD score 0.8), whereas data for the remaining two families were consistent with linkage to BRCA1 but not to BRCA2. The most likely interpretation of these findings is that BRCA1 mutations do confer an increased risk of male breast cancer but that the risk is lower than the risk conferred by BRCA2.

Heterogeneity results for the families containing only female breast cancer are of particular interest because the majority of families in the population are of this type, and the BCLC data on these families have expanded enormously since our first study (Easton et al. 1993). In the current study there were 117 families with no ovarian or male breast cancers. We estimate that in ~60% of these families the disease is likely to be due to a BRCA1 mutation or a BRCA2 mutation, regardless of the analysis model used. Of the families with six or more cases of female breast cancer, 21% are estimated to have disease that is due to BRCA1, and 60% are estimated to have disease that is due to BRCA2, leaving 19% due to other genes, when the CASH model is assumed for all genes and the BRCA1 mutation data are incorporated, with similar estimates in the other analyses. There were no large families with only female breast cancer that had LOD scores <-1.0 at both BRCA1 and BRCA2. In the group of families with four or five cases of female breast cancer, the proportions of families with disease attributable to BRCA1, BRCA2, and other genes are 28%, 5%, and 67%, respectively. The low proportion of families with four or five cases of female breast cancer that is due to BRCA2, as compared with the proportion of families with six or more cases due to it, may be due partly to chance, particularly since the latter group contains only 34 families. Both the low estimate of the proportion of families with only breast cancer due to BRCA1 and the slightly lower estimate for the proportion of the larger families with only breast cancer due to it are likely to be due to the significant ovarian cancer risk associated with BRCA1, although it is worth noting that a BRCA1 mutation was identified in 5 of the 34 families with six or more breast cancers and no ovarian cancer. The most important conclusion from this analysis is that a large proportion, perhaps the majority, of families with five or fewer cases of breast cancer and no ovarian or male breast cancer cases are not due to either BRCA1 or BRCA2. This is also indicated by the recent study by Serova et al. (1997), who performed mutation screening of BRCA1 and BRCA2 in 31 site-specific breast cancer families (7 of which have been included in the present study) and found mutations in only 8 of them. Similar results also were observed by Håkansson et al. (1997) and Schubert et al. (1997). A particular value of the current study, however, in addition to its much larger size, is that its results are based primarily on linkage and hence are not dependent on the sensitivity of the mutation testing. The fact that the proportion of families without linkage is much larger among families with fewer than six cases is consistent with the hypothesis that susceptibility alleles in other breast cancer genes confer risks lower than those conferred by BRCA1 or BRCA2 but are, correspondingly, more common in the population. The other known susceptibility genes—such as TP53, ATM, the mismatch-repair genes, the newly identified PTEN gene involved in Cowden syndrome (Liaw et al. 1997), and the gene(s) responsible for Peutz-Jeghers syndrome (Hemminki et al. 1997)—are unlikely to explain the high frequency of disease in an important fraction of these families. We conclude that other susceptibility genes responsible for a large fraction of familial breast cancer remain to be identified. Note that the present study did not examine linkage or mutation data in families with fewer than four cancer cases, so we cannot make any direct estimate of the contribution of BRCA1 and BRCA2 to families with two or three cases (in practice, by far the largest group), but, by extrapolation from the results in this

study, one would predict that their contribution would be relatively minor.

Overall, the heterogeneity analyses suggest that, in the western European and North American populations studied, BRCA1 is a slightly more common cause of cancer in high-risk families than is BRCA2. There may, however, be important local variations due to founder effects, most notably in Iceland, where, in six of the eight high-risk families, the disease is due to a single BRCA2 mutation. Moreover, the results do not imply that BRCA1 mutations are more common in lower-risk families, breast cancer cases, or the general population. In fact, if the penetrance of BRCA2 were markedly lower at younger ages, BRCA2 mutations could be more common in the population.

Although this was primarily a study based on linkage rather than on mutation data, we also performed some heterogeneity analyses incorporating BRCA1 mutation data. The main aim of these additional analyses was to ensure that no serious biases had been introduced by selective typing of BRCA2 markers on the basis of BRCA1 mutation status. Intuitively, such bias seem unlikely, since only a small proportion of data could have been influenced, and the analyses incorporating the mutation data obtained essentially the same estimates as were obtained by the analyses without mutation data. As a by-product of these analyses, we were able to obtain an estimate of the sensitivity of the mutation testing used in this study. Inevitably, this is a quite crude analysis, since different investigating groups used different screening techniques. Assuming that all methods (other than screening less than the whole of the gene) were equally sensitive, we estimated an identification probability of 63% (95% Cl 51%–77%). In this data set there was no evidence that sensitivity differed between sequencing and other methods. The obvious explanation for this relatively low sensitivity estimate, even for direct sequencing, is that a substantial fraction of BRCA1 alterations occur outside the coding sequence and splice sites. This explanation is supported by the observation that no mutation has been identified for a relatively high proportion of families where the disease is clearly linked to BRCA1. Of the 33 families with a LOD score >1.0 for linkage to BRCA1, 30 had been screened for a BRCA1 mutation. On the basis of the posterior probabilities in these families, one would predict that 29.3 (98%) of cases should be due to BRCA1. In fact, 21 mutations (70%) were found (3 of 6 tested by direct sequencing, 12 of 17 tested by CDGE, CSGE, DSDI/PTT, or SSCP-HA, and 6 of 7 tested by SSCP alone). Of the nine families without mutations, two (UTAH 2035 and IARC 2090) were shown to have loss of the entire BRCA1 transcript in cDNA, and one (IARC 1816) was found to have loss of exon 18 in transcript, but with no splice-site alteration. These families are therefore false negatives, according to the usual DNA-based testing methods. The true proportion of families with inferred regulatory mutations may be much higher, since cDNA testing is not always possible or informative. Other possible explanations for the "missing" mutations would include large deletions, promoter-sequence alterations, or simply a coding-sequence alteration that was missed.

We were able to estimate the penetrance of BRCA2 in families with a known mutation and to provide evidence for a high lifetime risk of breast or ovarian cancer: our overall estimates were 29% (95% Cl 9%–44%) by age 50 years and 88% (95% Cl 48%–97%) by age 70 years, the majority of which is attributable to breast cancer. Although not significantly different, the estimated breast cancer risk at younger ages was somewhat lower in BRCA2 mutation carriers compared with BRCA1 mutation carriers; 30% (76 of 250) of breast cancers at age <70 years in the 36 families with a germline BRCA2 mutation were diagnosed at age <40 years and 73% (182 of 250) were diagnosed at age <50 years, compared with 50% (210 of 416) diagnosed at age <40 years and 80% (333 of 416) diagnosed at age <50 years in the 64 families with a germ-line BRCA1 mutation. The overall risks of ovarian cancer (7% by age 60 years and 27% by age 70 years) are lower than the corresponding estimates for BRCA1 that had been obtained by previous consortium analyses (31% by age 60 and 42% by age 70 years [Easton et al. 1993; Narod et al. 1995*b*]; and 30% by age 60 years and 63% by age 70 years [Easton et al. 1995]), although they still are substantially elevated above the risks in the general population. This is entirely consistent with the heterogeneity results showing that BRCA2 is responsible for a smaller—but still significant—fraction of breast-ovarian cancer families than is due to BRCA1.

Our penetrance estimates assumed that all mutations conferred the same risks, which is likely to be an oversimplification. Gayther et al. (1997) found evidence for an increased risk of ovarian cancer, relative to the risk of breast cancer, associated with mutations in the central portion of BRCA2 (the OCCR), compared with mutations outside this region. Our study provides some support for this genotype-phenotype correlation, in that higher risks of ovarian cancer—but similar risks of breast cancer—were estimated for BRCA2 mutations in the OCCR, compared with mutations outside it.

Recently Struewing et al. (1997) estimated the penetrance of BRCA1 and BRCA2 mutations on the basis of the family histories of 120 Ashkenazi Jewish volunteers found to have any of the three founder mutations common in this population. In carriers of the 6174delT mutation, which lies within the OCCR, they estimated a breast cancer risk of ~50% by age 70 years and an ovarian cancer risk of ~10% by age 60 years and 18% by age 70 years. These estimates are somewhat lower

than the estimates derived in the present study, but they are not inconsistent, given the wide confidence limits.

In this study, we have not attempted to estimate the risk of male breast cancer in BRCA2 families. The maximum-LOD-score approach would not provide a precise estimate, because the penetrance is clearly quite low. Using the two largest BRCA2 families with linkage (families UTAH 107 and CRC 186) Easton et al. (1997) estimated the cumulative risk of male breast cancer in BRCA2 carriers to be 6% by age 70 years, but this was based on only four observed cases and hence is very imprecise. There is also circumstantial evidence of variation in male breast cancer risk; whereas the large majority (19 of 26) of "male breast cancer" families in this study contain only one male case, two families contained four cases each. In both of these families, BRCA2 mutations have been identified. This variation in risk is likely to be due, at least in part, to modifying factors, since both families have mutations identical to those in large breast cancer families that do not have male breast cancer cases.

In addition to the risks of breast and ovarian cancer, previous studies have suggested an increased risk of a number of other cancers in BRCA2 carriers, including prostate cancer (Tulinius et al. 1992), pancreatic (Phelan et al. 1996), and ocular melanoma (Easton et al. 1997). The risks of other cancers in the collaborative BCLC data set are currently being evaluated.

The penetrance estimates provided by this study, together with the comparable estimates for BRCA1 carriers, should be useful for genetic counseling of mutation carriers. Such estimates must always be used with caution, however. They are appropriate for counseling in multiple-case families but may not apply to mutation carriers with little or no family history. The risks to such individuals could be lower than those estimated here, either if there is risk variation between mutations, or if modifying genes or other familial risk factors strongly influence penetrance.

Acknowledgments

The analysis of these data was funded by the Cancer Research Campaign. Meetings of the Breast Cancer Linkage Consortium, as well as data management, were supported by EC Concerted Action. Data collection and laboratory analysis were supported by the Cancer Research Campaign, the Imperial Cancer Research Fund, Aberdeen Royal Hospitals Endowment, NIH grant CA55914, U.S. Army grants DAMD-17-94-J-4260 and DAMD-17-94-J-4340, the ARC, Ligue Contre le Cancer, Fédération Nationale des Centres de Lutte Contre le Cancer, Fédération des Groupements d'Entreprises dans la Lutte Contre le Cancer, INSERM CR 4U00 3C, Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie grant 01ZZ9509, Deutsche Krebshilfe, and Swiss Cancer League grants AKT 332 and AKT 463. We thank Lesley McGuffog and Amy Storfer-Isser for their help with the statistical analysis, and we thank François Eisinger and Daniel Birnbaum for help with data collection and laboratory work.

Appendix A

Heterogeneity Analysis

Linkage analyses were performed by means of the FASTLINK version of the LINKAGE program (Lathrop et al. 1984; Cottingham et al. 1993). The CASH model was modified to assume that the incidence of breast cancer in gene carriers age $\geqslant 70$ years was the same as in carriers age 60–69 years. The risks in noncarrier females were assumed to follow population rates for breast and ovarian cancer for England and Wales. (All families were from western European or North American populations that have similar breast and ovarian cancer incidence rates.) For implementation in LINKAGE, 21 liability classes were constructed, allowing for seven age groups and three phenotypes (unaffected, affected with breast cancer, and affected with ovarian cancer). Individuals with both breast and ovarian primary cancers were treated as if they were affected with breast cancer at age <30 years, in order to maximize their probability of being gene carriers. Males with breast cancer were also treated as if they were females affected at age <30 years, because of the rarity of male breast cancer. All unaffected males were treated as they were unaffected females of age <30 years, which is approximately equivalent to being of unknown disease status.

For most families, multipoint LOD scores for linkage to BRCA1 and BRCA2 were computed by use of two flanking markers and the disease. A small minority of families contained consanguineous or marriage loops, and these were broken if this resulted in no loss of information, but, when a loop had to be retained, the most informative marker on each chromosome was used in a two-point analysis. BRCA1 was assumed to lie either midway between THRA1 and D17S579 (1.3 cM from THRA1 on the male genetic map, with a 2:1 female: male genetic-distance ratio) or at D17S855. BRCA2 was assumed to lie 1 cM distal to D13S260 and 2 cM proximal to D13S267.

Estimates of proportions of families with disease due to BRCA1 and BRCA2 were obtained by maximization of the heterogeneity LOD score,

$$\begin{aligned} &\mathrm{LOD}(\alpha_1, \alpha_1) \\ &\quad = \Sigma \log_{10}[\alpha_1 10^{\mathrm{LOD}_1}(\underline{\theta}_1) + \alpha_2 10^{\mathrm{LOD}_2}(\underline{\theta}_2), \\ &\qquad + (1 - \alpha_1 - \alpha_2)] \end{aligned} \tag{A1}$$

with respect to α_1 and α_2, where α_1 and α_2 are the respective proportions of families with disease linked to

BRCA1 and BRCA2. The proportion of families with disease due to other high-penetrance genes is estimated by $1 - \alpha_1 - \alpha_2$. θ_1 is the vector of recombination fractions between the 17q markers and BRCA1, and θ_2 is the vector of recombination fractions between the 13q markers and BRCA2. $LOD_1(\underline{\theta_1})$ and $LOD_2(\underline{\theta_2})$ are the multipoint LOD scores for linkage to BRCA1 and BRCA2, respectively. Summation is over all families.

Maximizing this heterogeneity LOD score provides consistent estimates for α_1 and α_2. This is true despite the fact that families showing strong evidence for linkage to BRCA1 were sometimes not typed for BRCA2 markers. The reasoning behind this is as follows. Equation (A1) may be rewritten as

$$\begin{aligned}
&\mathrm{LOD}(\alpha_1, \alpha_2) \\
&\quad = \Sigma \log_{10} \mathrm{L}(\mathrm{D}, \mathrm{M}_1, \mathrm{M}_2 / \underline{\theta_1}, \underline{\theta_2}, \alpha_1, \alpha_2) \\
&\qquad - \Sigma \log_{10} \mathrm{L}(\mathrm{D}, \mathrm{M}_1, \mathrm{M}_2 / \underline{1/2}, \underline{1/2}) \\
&\quad = \Sigma \log_{10} \mathrm{L}(\mathrm{D}, \mathrm{M}_1, \mathrm{M}_2 / \mathrm{D}, \underline{\theta_1}, \underline{\theta_2}, \alpha_1, \alpha_2) \\
&\qquad - \Sigma \log_{10} \mathrm{L}(\mathrm{M}_1, \mathrm{M}_2) \ ,
\end{aligned}$$

where D represents the disease phenotypes and M_1 and M_2 represent the observed marker genotypes. Thus maximizing the LOD score is equivalent to maximizing the likelihood of all marker and disease phenotypes, conditional on the disease phenotypes (Elston 1995). This, in turn, may be written as

$$\begin{aligned}
&\mathrm{LOD}(\alpha_1, \alpha_2) \\
&\quad = \Sigma \log_{10} \mathrm{L}(\mathrm{D}, \mathrm{M}_1 / \mathrm{D}, \underline{\theta_1}, \alpha_1) \\
&\qquad + \Sigma \log_{10} \mathrm{L}(\mathrm{D}, \mathrm{M}_1, \mathrm{M}_2 / \mathrm{D}, \mathrm{M}_1 \underline{\theta_1}, \underline{\theta_2} \alpha_1, \alpha_2) \\
&\qquad - \Sigma \log_{10} \mathrm{L}(\mathrm{M}_1, \mathrm{M}_2) \ .
\end{aligned}$$

The first term is a conditional log likelihood for the markers at the BRCA1 locus, and it therefore gives consistent estimates of α_1. The second term is a conditional log likelihood for the BRCA2 markers, conditional on the disease status and the BRCA1 marker genotypes, which is therefore a valid log likelihood leading to consistent estimates of α_1 and α_2, even though there is some selection based on M_1. The third term is simply a constant. Maximizing LOD (α_1,α_2) will therefore give consistent estimates for α_1 and α_2 (by the usual argument that, since the expected derivative of each term on the right-hand side of the equation is zero at the true values of α_1 and α_2, since each is a valid log likelihood, this must also be true of the sum).

Analyses allowing BRCA1 to confer a higher penetrance (the "BRCA1 model") were based on maximization of the LOD score,

$$\begin{aligned}
&\mathrm{LOD}(\alpha_1, \alpha_2) \\
&\quad = \Sigma \log_{10} [\alpha_1 \mathrm{L}_1(\mathrm{D}, \mathrm{M}_1, \mathrm{M}_2 / \underline{\theta_1}, \underline{1/2}) \\
&\qquad + \alpha_2 \mathrm{L}_2(\mathrm{D}, \mathrm{M}_1, \mathrm{M}_2 / \underline{1/2}, \underline{\theta_2}) \\
&\qquad + (1 - \alpha_1 - \alpha_2) \mathrm{L}_2(\mathrm{D}, \mathrm{M}_1, \mathrm{M}_2 / \underline{1/2}, \underline{1/2})] \\
&\qquad - \Sigma \log_{10} [\alpha_1 \mathrm{L}_1(\mathrm{D}, \mathrm{M}_1, \mathrm{M}_2 / \underline{1/2}, \underline{1/2}) \\
&\qquad + \alpha_2 \mathrm{L}_2(\mathrm{D}, \mathrm{M}_1, \mathrm{M}_2 / \underline{1/2}, \underline{1/2}) \\
&\qquad + (1 - \alpha_1 - \alpha_2) \mathrm{L}_2(\mathrm{D}, \mathrm{M}_1, \mathrm{M}_2 / \underline{1/2}, \underline{1/2})] \ ,
\end{aligned}$$

with respect to α_1, the proportion of BRCA1 mutations as a proportion of all high-risk mutations, and α_2, the proportion of BRCA2 mutations as a proportion of all high-risk mutations, L_1 represents the likelihood for a family under the BRCA1 model, and L_2 represents the likelihood under the CASH model. The relative frequency of mutations in high-risk genes other than BRCA1 and BRCA2 is estimated by $1 - \alpha_1 - \alpha_2$. Summation is over all families.

Dividing the numerator and the denominator by L_2(D,M_1,M_2/1/2, 1/2) allows simplification to

$$\begin{aligned}
&\mathrm{LOD}(\alpha_1, \alpha_2) \\
&\quad = \Sigma \log_{10} [\alpha_1 10^{\mathrm{LOD}_1(\underline{\theta_1}) - G} \\
&\qquad + \alpha_2 10^{\mathrm{LOD}_2(\underline{\theta_2})} + (1 - \alpha_1 - \alpha_2)] \\
&\qquad - \Sigma \log_{10} [\alpha_1 10^{\mathrm{G}} + \alpha_2 + (1 - \alpha_1 - \alpha_2)] \ . \qquad \text{(A2)}
\end{aligned}$$

$LOD_1(\theta)_1$ is the multipoint LOD score for linkage to BRCA1, computed under the BRCA1 model, and $LOD_2(\theta)_2$ is the multipoint LOD score for linkage to BRCA2, computed under the CASH model. $G = log_{10}[L_1(D)/L_2(D)]$; that is, G is the log of the ratio of the likelihood of the disease phenotypes in the family under the BRCA1 model to the likelihood of the disease phenotypes under the CASH model. (For practical purposes, $L_1(D)/L_2(D)$ is the same as L_1(D,M/1/2)/L_2(D,M/1/2); hence, both numerator and denominator may be computed in LINKAGE, with the same marker data but with different models; we used the 17q marker typings for which we had already computed the numerator and denominator.)

α_1, α_2, and $1 - \alpha_1 - \alpha_2$ no longer estimate the proportions of families with disease due to BRCA1, BRCA2, and other genes, since these depend on the genes' penetrances as well as on their frequencies. For each family, the posterior probabilities p_1 (i.e., of segregating a BRCA1 mutation), p_2 (i.e., of segregating a BRCA2 mutation), and p_3 (i.e., of segregating a mutation in another high-risk gene) were therefore computed, where

$$p_1 = \alpha_1 10^{\mathrm{LOD}_1(\underline{\theta}_1)+G}/[\alpha_1 10^{\mathrm{LOD}_1(\underline{\theta}_1)+G} + \alpha_2 10^{\mathrm{LOD}_2(\underline{\theta}_2)} + (1-\alpha_1-\alpha_2)] ,$$

$$p_2 = \alpha_2 10^{\mathrm{LOD}_2(\underline{\theta}_2)}/[\alpha_1 10^{\mathrm{LOD}_1(\underline{\theta}_1)+G} + \alpha_2 10^{\mathrm{LOD}_2(\underline{\theta}_2)} + (1-\alpha_1-\alpha_2)] ,$$

and $p_3 = 1 - p_1 - p_2$. The proportion of families due to each gene was then estimated as the average of the corresponding posterior probabilities.

Appendix B

Heterogeneity Analysis Incorporating BRCA1 Mutation Data

Parameters α_1, α_2, and δ were estimated by joint maximization of a likelihood incorporating both linkage and mutation data. Under the assumption that the CASH model applies to all genes, a family with an identified mutation contributes $\log_{10}(\alpha_1\delta 10^{\mathrm{LOD}_1(\underline{\theta}_1)})$ to the total likelihood expression, a family in which the mutation has been looked for but was not found contributes

$$\log_{10}[\alpha_1(1-\delta)10^{\mathrm{LOD}_1(\underline{\theta}_1)} + \alpha_2 10^{\mathrm{LOD}_2(\underline{\theta}_2)} + (1-\alpha_1-\alpha_2)] ,$$

and a family in which mutation testing has not been performed contributes as previously described (see eq. [A1].

The parameter δ can similarly be introduced into the analysis, allowing BRCA1 to confer a higher penetrance. A family with an identified mutation then contributes

$$\log_{10}(\alpha_1\delta 10^{\mathrm{LOD}_1(\underline{\theta}_1)+G}) - \log_{10}[\alpha_1 10^{G} + \alpha_2 + (1-\alpha_1-\alpha_2)]$$

to the total likelihood expression. A family in which the mutation has been looked for but has not been found contributes

$$\log_{10}[\alpha_1(1-\delta)10^{\mathrm{LOD}_1(\underline{\theta}_1)+G} + \alpha_2 10^{\mathrm{LOD}_2(\underline{\theta}_2)} + (1-\alpha_1-\alpha_2)] - \log_{10}[\alpha_1 10^{G} + \alpha_2 + (1-\alpha_1-\alpha_2)]$$

to the total likelihood, and a family in which mutation testing has not been performed contributes as described previously (see eq. [A2]).

Confidence intervals for the parameters α_1, α_2, $(1-\alpha_1-\alpha_2)$, and δ were computed by comparing the difference in $\log_e$ likelihoods for different values of α_1, α_2, $(1-\alpha_1-\alpha_2)$, and δ, respectively, and maximizing over the other parameters, to a χ^2 distribution on 1 df. To compute the upper and lower confidence limits on the proportion of families with disease attributable to a gene (e.g., BRCA1), we used the upper and lower confidence limits, respectively, on α_1, with the maximum-likelihood estimate for α_2, given α_1, and with computed posterior probabilities of linkage to each gene, as described above.

Appendix C

Estimation of the Penetrance of BRCA2

Computations were performed with a modified version of the ILINK program (Lathrop et al. 1984); D13S267 typings were used, since the computer time that would be required for maximization over multipoint data was prohibitive. Incidence rates in noncarriers were fixed at the population rates for England and Wales. Rates in carriers were modeled under the assumption of a separate parameter for the ratio of the incidences in gene carriers versus the incidences for carriers in the CASH model, for each of five age groups (20–29 years, 30-39 years, 40–49 years, 50–59 years, and 60–69 years). Follow-up of all females was censored at age 70 years. Men were included in the liability class for the 20–29-years-old age group. Male breast cancer cases were therefore assumed to be likely to be gene carriers, regardless of age at diagnosis, and no attempt was made to estimate the penetrance for male breast cancer. The overall penetrance, defined as risk of breast or ovarian cancer, was modeled by defining age at onset as the age at diagnosis of the first cancer (either breast or ovarian). Estimates of breast and ovarian cancer incidence were then obtained by dividing the overall incidence rate in each age group by the observed age-specific proportions of breast and ovarian cancer.

The penetrance of BRCA2, as based on families with six or more breast cancers, was estimated by maximization of the LOD score

$$\mathrm{LOD}(\gamma) = \Sigma\log_{10}\left[\gamma 10^{\mathrm{LOD}_1(\underline{\theta}_1)} + (1-\gamma)10^{\mathrm{LOD}_2(\underline{\theta}_2)+F}\right] - \Sigma\log_{10}[\gamma + (1-\gamma)10^{F}] ,$$

where γ is the proportion of high-risk mutations that are BRCA1 mutations. The proportion of high-risk mutations that are BRCA2 mutations is then forced to be $(1-\gamma)$. $\mathrm{LOD}_1(\theta_1)$ is the LOD score at BRCA1, under the BRCA1 model (and is fixed), and $\mathrm{LOD}_2(\theta_2)$ is the LOD score at BRCA2, under the model estimated for BRCA2 (which is maximized over different penetrance functions). $F = \log_{10}[L_1(D)/L_2(D)]$, where $L_1(D)$ is the likelihood of the disease phenotypes under the BRCA1

model (and is fixed) and $L_2(D)$ is the likelihood of the disease phenotypes under the BRCA2 model (and has to be recomputed at each iteration, as the model changes). (In practice we computed $F = \log_{10}[L_1(D,M/1/2)/L_2(D,M/1/2)]$ by using MLINK with D13S267 typings [when these were available] or, otherwise, a dummy marker with complete typing [for speed].)

References

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

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

Claus EB, Schildkraut JM, Thompson WD, Risch NJ (1993) Analysis of the genetic relationship between breast and ovarian cancer. Am J Hum Genet Suppl 53:787

Cottingham RW Jr, Idury RM, Schäffer AA (1993) Faster sequential genetic linkage computations. Am J Hum Genet 53: 252–263

Easton DF, Bishop DT, Ford D, Crockford GP, 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

Easton DF, Ford D, Bishop DT, Breast Cancer Linkage Consortium (1995) Breast and ovarian cancer incidence in BRCA1-mutation carriers. Am J Hum Genet 56:265–271

Easton DF, Steele L, Fields P, Ormiston W, Averill D, Daly PA, McManus R, et al (1997) Cancer risks in two large breast cancer families linked to BRCA2 on chromosome 13q12-13. Am J Hum Genet 61:120–128

Elston RC (1995) 'Twixt cup and lip: how intractable is the ascertainment problem? Am J Hum Genet 56:15–17

Ford D, Easton DF, Bishop DT, Narod SA, Goldgar DE, Breast Cancer Linkage Consortium (1994) Risks of cancer in BRCA1 mutation carriers. Lancet 343:692–695

Ganguly A, Rock MJ, Prockop DJ (1993) Confirmation sensitive gel electrophoresis for rapid detection of single base differences in double stranded PCR products and DNA fragments—evidence for solvent induced bends in DNA heteroduplexes. Proc Natl Acad Sci USA 90:10325–10329

Gayther SA, Harrington P, Russell P, Kharkevich G, Garkavtseva RF, Ponder BAJ (1996) Rpaid detection of regionally clustered germ-line BRCA1 mutations by multiple heteroduplex analysis. Am J Hum Genet 58:451–456

Gayther SA, Mangion J, Russell P, Seal S, Barfoot R, Ponder BAJ, Stratton MR, et al (1997) Variation of risks of breast and ovarian cancer associated with different germline mutations of the BRCA2 gene. Nat Genet 15:103–105

Gayther SA, Warren W, Mazoyer S, Russell PA, Harrington PA, Chiano M, Seal S, et al (1995) Germline mutations of the BRCA1 gene in breast and ovarian cancer families provide evidence for a genotype-phenotype correlation. Nat Genet 11:428–433

Gyapay G, Morissette J, Vignal A, Dib C, Fizomes C, Millasseau P, Marc S, et al (1994) The 1993–94 Généthon human genetic linkage map. Nat Genet 7:246–339

Håkansson S, Johannsson O, Johansson U, Sellberg G, Loman N, Gerdes A-M, Holmberg E, et al (1997) Moderate frequency of BRCA1 and BRCA2 germ-line mutations in Scandinavian familial breast cancer. Am J Hum Genet 60: 1068–1078

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 MC (1990) Linkage of early onset familial breast cancer to chromosome 17q21. Science 250: 1684–1689

Hemmenki A, Tomlinson I, Markie D, Jarvinen H, Sistonen P, Bjorkqvist AM, Knuutila S, et al (1997) Localization of a susceptibility locus for Peutz-Jeghers syndrome to 19p using comparative genomic hybridization and targeted linkage analysis. Nat Genet 15:87–90

Hogervorst FBL, Cornelis RS, Baut M, van Vliet M, Oosterwijk JC, Olmer R, Bakker B, et al (1995) Rapid detection of BRCA1 mutations by the protein truncation test. Nat Genet 10:208–212

Holt JT, Thompson ME, Szabo C, Robinson-Benion C, Arteaga CL, King M-C, Jenson RA (1996) Growth retardation and tumour inhibition. Nat Genet 12:298–302

Kerangueven F, Essioux L, Dib A, Noguchi T, Allione F, Geneix J, Longy M, et al (1995) Loss of heterozygosity and linkage analysis in breast-carcinoma—indication for a putative 3rd susceptibility gene on the short arm of chromosome 8. Oncogene 10:1023–1026

Lathrop GM, Lalouel JM, Julier C, Ott J (1984) Strategies for multilocus linkage analysis in humans. Proc Natl Acad Sci USA 81:3443–3446

Liaw D, Marsh D, Li J, Dahia PLM, Wang SI, Zheng Z, Bose S, et al (1997) Germline mutations of the PTEN gene in Cowden disease, an inherited breast and thyroid cancer syndrome. Nat Genet 16:64–67

Miki Y, Swensen J, Schattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, Liu QY, et al (1994) Isolation of BRCA1, the 17q-linked breast and ovarian cancer susceptibility gene. Science 266:66–71

Narod SA, Ford D, Devilee P, Barkardottir RB, Eyfjord J, Lenoir G, Serova O, et al (1995*a*) Genetic heterogeneity of breast-ovarian cancer revisited. Am J Hum Genet 57:957

Narod S, Ford D, Devilee P, Barkardottir RB, Lynch HT, Smith SA, Ponder BAJ, et al (1995*b*) An evaluation of genetic heterogeneity in 145 breast-ovarian cancer families. Am J Hum Genet 56:254–264

Peelen T, van Vliet M, Petrij-Bosch A, Mieremet R, Szabo C, van den Ouweland AMW (1997) A high proportion of novel mutations in BRCA1 with strong founder effects among Dutch and Belgian hereditary breast and ovarian cancer families. Am J Hum Genet 60:1041–1049

Phelan CM, Lancaster JM, Tonin P, Gumbs C, Cochran C, Carter R, Ghadirlan P, et al (1996) Mutation analysis of the BRCA2 gene in 49 site-specific breast cancer families. Nat Genet 13:120–122

Schubert EL, Lee MK, Mefford HC, Argonza RH, Morrow JE, Hull J, Dann JL, et al (1997) BRCA2 in American fam-

ilies with four or more cases of breast or ovarian cancer: recurrent and novel mutations, variable expression, penetrance, and the possibility of families whose cancer is not attributable to BRCA1 or BRCA2. Am J Hum Genet 60: 1031–1040

Seitz S, Rohde K, Bender E, Nothnagel A, Kolble K, Schlag PM, Scherneck S (1997) Strong indication for a breast cancer susceptibility gene on chromosome 8p12-p22: linkage analysis in German breast cancer families. Oncogene 14: 741–743

Serova OM, Mazoyer S, Puget N, Dubois V, Tonin P, Shugart YY, Goldgar D, et al (1997) Mutations in *BRCA1* and *BRCA2* in breast cancer families: are there more breast cancer–susceptibility genes? Am J Hum Genet 60:486–495

Stoppa-Lyonnet D, Laurent-Puig P, Essioux L, Pagès S, Ithier G, Ligot L, Fourquet A, et al (1997) BRCA1 sequence variations in 160 individuals referred to a breast/ovarian family cancer clinic. Am J Hum Genet 60:1021–1030

Stratton MR, Ford D, Neuhasen S, Seal S, Wooster R, Friedman LS, King MC, et al (1994) Familial male breast cancer is not linked to BRCA1. Nat Genet 7:103–107

Struewing JP, Hartge P, Wacholder S, Baker SM, Berlin M, McAdams M, Timmerman MM, et al (1997) The risk of cancer associated with specific mutations of BRCA1 and BRCA2 among Ashkenazi Jews. N Engl J Med 336: 1401–1408

Tavtigian SV, Simard J, Rommens J, Couch F, Shattuck-Eidens D, Neuhausen S, Merajver S, et al (1996) The complete BRCA2 gene and mutations in 13q-linked kindreds. Nat Genet 12:333–337

Tulinius H, Egilsson V, Olafsdottir GH, Sigvaldson H (1992) Risk of prostate, ovarian and endometrial cancer among relatives of women with breast cancer. Br Med J 305: 855–857

Weber JM, Kwitek AE, May PE, Wallace MS, Collins FS, Ledbetter DH (1990) Dinucleotide repeat polymorphisms at the D17S250 and D17S261 loci. Nucleic Acids Res 18:464

Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, Collins N, et al (1995) Identification of the breast cancer susceptibility gene BRCA2. Nature 378:789–792

Wooster R, Neuhausen S, Manigion J, Quirk Y, Ford D, Collins N, Nguyen K, et al (1994) Localisation of a breast cancer susceptibility gene (BRCA2) to chromosome 13q by genetic linkage analysis. Science 265: 2088–2090