CGH-targeted linkage analysis reveals a possible BRCA1 modifier locus on chromosome 5q

Katherine L. Nathanson1,2, Yin Y. Shugart5, Romaica Omaruddin1, Csilla Szabo6, David Goldgar6, Timothy R. Rebbeck3 and Barbara L. Weber1,2,4,*

1Department of Medicine, 2Abramson Family Cancer Research Institute, 3Department of Biostatistics and Epidemiology, 4Department of Genetics, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA, 5Epidemiology Department, Bloomberg of School of Public Health, Johns Hopkins University, Baltimore, MD 21205, USA and 6International Agency for Research on Cancer, 69372 Lyon Cedex 08, France

Received February 2, 2002; Revised and Accepted March 15, 2002

Women with germline mutations in BRCA1 have a greatly elevated risk of breast and ovarian cancer. However, considerable variation in the degree of breast cancer risk associated with a BRCA1 mutation has been observed, suggesting that modifiers of BRCA1 penetrance may exist. We hypothesized that the modifier genes might be located in regions of allelic imbalance in the tumors of BRCA1 mutation carriers, as have been reported on chromosomes 4p, 4q and 5q. In order to determine whether novel genetic modifiers of BRCA1-associated breast cancer penetrance in these regions exist, we used non-parametric linkage analysis methods to determine whether allele sharing of chromosomes 4p, 4q and 5q was observed preferentially within BRCA1 mutation families in women with BRCA1 mutations and breast cancer. No significant linkage on chromosome 4p or 4q was observed associated with breast cancer risk in BRCA1 mutation carriers. However, we observed a significant linkage signal at D5S1471 on chromosome 5q (P = 0.009) in all the families analyzed together. The significance of this observation increased in the subset of families with an average of breast cancer diagnosis less than 45 years (P = 0.003). These results suggest the presence of one or more genes on chromosome 5q33–34 that modify breast cancer risk in BRCA1 mutation carriers. The approach described here may be utilized to identify penetrance modifiers in other autosomal dominant syndromes.

INTRODUCTION

The observation of substantial variability in penetrance of breast cancer due to germline mutations in BRCA1 has led to speculation that inheritance of a BRCA1 mutation alone may not completely describe the observed cancer phenotype. The initial estimate of BRCA1-associated breast cancer risk of 85% was calculated using the original families ascertained for BRCA1 linkage analysis. However, population- and hospital-based studies have estimated a lifetime breast cancer risk in BRCA1 mutation carriers of 36–56% (1–6). In a recent hospital-based study of Ashkenazi Jewish women, the risk of breast cancer due to the two Ashkenazi Jewish BRCA1 founder mutations was estimated to be 46%, substantially the lower breast cancer risk seen in previous studies (7). Thus breast cancer risk due to BRCA1 mutation differs based on the population of ascertainment, suggesting that other factors – either genetic or environmental – modify breast cancer risk.

While germline mutations in BRCA1 and BRCA2 account for the majority of families with both breast and ovarian cancers, they account for only a minority of families with multiple cases of breast cancer only. Studies in populations of early-onset breast cancer cases have demonstrated that approximately 85% of the breast cancer risk to mothers and sisters is not attributable to mutations in either gene, suggesting the presence of variants in additional genes that confer breast cancer risk (8–10). While mutations or variants in susceptibility genes other than BRCA1 or BRCA2 could presumably act alone to affect breast cancer risk in the general population, in BRCA1 mutation carriers they could act in concert with the high-penetrance mutation to modify age at breast cancer diagnosis. In fact, using cases of breast cancer under the age of 55 years from the East Anglia Cancer Registry (UK), after accounting for mutations in BRCA1 and BRCA2 and adjusting for parity, a polygenic model was the best explanation for additional breast cancer susceptibility genes. In addition, these putative common susceptibility alleles have an effect in both BRCA1 and BRCA2 carriers.

*To whom correspondence should be addressed at: 514 BRB 2/3, 421 Curie Blvd, University of Pennsylvania, Philadelphia, PA 19104, USA. Tel: +1 215 898 0247; Fax: +1 215 573 2486; Email: weberb@mail.med.upenn.edu
mutation carriers and those without mutations (11). Therefore, the evidence suggests that genetic variation, other than that due to mutations in BRCA1 and BRCA2, contributes to breast cancer susceptibility.

Several approaches have been taken to identify additional breast cancer susceptibility genes. These approaches have included linkage studies, population-based case-control association studies, and case-control association studies in BRCA1 and BRCA2 mutation carriers using those affected with cancer as cases and those without cancer as controls. This third group of studies mainly has examined candidate modifier genes, selected on the basis of biological plausibility (12–14). The first such study demonstrated an association of rare alleles of HRAS with ovarian cancer (a 2.11-fold risk increase) but not of breast cancer (15). Additional studies have demonstrated that increased length of CAG repeats within both the androgen receptor (AR, Xq11.2–12) and AIB1 (20q12) are significantly associated with increased breast cancer risk in BRCA1 mutation carriers (16,17). However, the AR genotypes associated with increased risk are infrequent in the population and unlikely to account for all of the phenotypic variability. In addition, replication studies have not shown consistent results (18,19). Association studies using variants in candidate genes support the evidence that genetic modifiers of penetrance in BRCA1 and BRCA2 mutation carriers exist.

While variants in candidate genes have been associated with altered penetrance in BRCA1 mutation carriers, these studies are intrinsically limited to known genes selected on the basis of biological plausibility. Therefore, we sought to determine whether chromosomal regions containing yet-uncharacterized genes that alter breast cancer risk in BRCA1 mutation carriers could be identified. In order to prioritize chromosomal regions likely to contain genes affecting breast cancer penetrance, we used data from comparative genome hybridization (CGH) studies in BRCA1 mutation carriers. CGH is sensitive to both chromosomal gains and losses, indicative of the presence of oncogenes and tumor suppressors, respectively. The most consistently observed alterations in breast tumors from BRCA1 mutation carriers were loss of the chromosomal arms 5q (86%), 4q22–qter (81%) and 4p (64%). In contrast, at 5q, 4q and 4p, sporadic breast tumors show losses of 11% (20,21). While many modifiers of penetrance are presumably not tumor suppressor genes that select for deletions, consistent loss of 4p, 4q and 5q provides the strongest available evidence that there are genes in these regions related to BRCA1-associated tumorigenesis.

RESULTS

For the initial evaluation of chromosomes 4 and 5q, microsatellite markers spaced at an average of 12 cM across these regions were analyzed in 18 families (Fig. 1). Twelve families had deleterious BRCA1 mutations. Six families had early-onset breast cancer co-segregating with polymorphic markers on 17q21 surrounding BRCA1, but no detectable mutations. In order to maximize the potential of finding a modifier locus in this initial evaluation, we were as inclusive of families as possible. The initial analysis was performed using identity-by-descent methods without assuming a model of inheritance for a modifier gene effect. Using SimIBD, we did not identify any significant candidate loci associated with breast cancer in BRCA1 mutation carriers on chromosome 4. However, we identified 3 markers on chromosome 5q with P-values less than 0.05: D5S495 (P = 0.016), D5S2115 (P = 0.009) and D5S400 (P = 0.02).

In order to further explore the possibility that one or more modifier genes exist at these loci, we increased both the stringency of the family selection for BRCA1 mutation carriers and the marker density at the loci with P < 0.05. We added four newly identified families with known deleterious mutations in BRCA1 and removed four of the six families with evidence of linkage to 17q and posterior probabilities of carrying a mutation in BRCA1 less than 90%, for 18 families in total. The remaining two families without detectable mutations in BRCA1 had posterior probabilities of carrying a mutation in BRCA1 greater than 90% (96% and 98%). The families with posterior probabilities less than 90% were eliminated so that the population studied was as homogenous as possible in the second stage of analysis. Additional markers were targeted to the locations of the initial positive SimIBD scores, so that fine mapping could be performed (Fig. 2). In order to perform multiple analyses with all the families, including those with extended pedigrees, SimWalk2 was used, and the statistic E (equivalent to SIII in Genehunter) was calculated (22).

After the families were analyzed in total, they were stratified in two different groups. In the first stratified analysis, the families were classified by average age of breast cancer diagnosis. Women diagnosed with ovarian cancer only or with ovarian cancer prior to their breast cancer diagnosis were not included either in the calculations or in the analysis, since their therapy for ovarian cancer may have affected their risk of breast cancer. In addition, the families were grouped by the number of breast cancer cases per family (< 6, 7 families or ≥ 6, 11 families).

When all the families were analyzed together, the strongest linkage signal was seen at D5S1471 (P = 0.009) (Fig. 2).
This effect was even more pronounced in families with an average age of diagnosis of 45 years or less ($n = 15$) ($P = 0.003$). No difference in $P$-values was seen when the families were stratified by number of breast cancer cases per family (data not shown).

**DISCUSSION**

These results suggest that gene or genes located on chromosome 5q33–34 may modify breast cancer penetrance in kindreds with germline BRCA1 mutations. Given that the most significant results were seen in families with an average age of diagnosis less than 45 years, we postulate that a variant in a gene or genes within that region increases the age-adjusted penetrance of BRCA1 mutations. The region within which the gene(s) is most likely located in the interval defined by an 10-fold increase in the $P$-value from the peak of 0.003 to 0.03. This interval encompasses from D5S422 to approximately halfway between D5S2075 and D5S408, 10 Mb centered around D5S1471.

In order to make a preliminary list of genes within this region that may be responsible for the effects observed here, we compared the set of known and predicted genes (hypothetical proteins) that were localized by sequence to this region in three databases: the Human Genome sequence (NCBI Build 27), Ensembl (v1.2.0) and the Celera database (v3.3). For this analysis, we included known genes and novel transcripts of more than one exon when there was additional supporting evidence for their existence from other sources such as Unigene or EMBL. We then compared the lists of genes from each source, and eliminated known or predicted genes that were only localized to this region by one of the three data sources (Table 1). Within this region, the Ensembl and Celera databases were very consistent. However, the NCBI region contained a contig (NT 006907.7) that had been within this region on earlier builds, but was now located approximately 10 cM proximal to the region. As the markers within the contig are consistent with its location within the 10 Mb region of interest, the genes on the contig are considered in Table 1. Although the genome annotation is unfinished in this region, this list of genes and transcripts provides a starting point that can be expanded as the

---

**Figure 2.** The results of the second stage of analysis looking for modifiers of penetrance in BRCA1 mutation carriers using SimWalk2 are shown as negative logarithm of the $P$-values obtained, with the markers used shown on the x-axis. The marker distance is shown in centimorgans across 5q, with D5S647 (centromeric marker) at 0 cM and D5S408 (telomeric marker) at 121 cM. Results are demonstrated as a negative $log P$-values for greater discrimination. All pedigrees: The SimWalk2 analysis results for all families together, showing the peak negative logarithm of the $P$-value 2.06 at D5S1471, corresponding to a $P$-value of 0.009. AAD < 45: The SimWalk2 analysis results in families with an average age of breast cancer diagnosis 45 years or less, showing the peak negative logarithm of the $P$-value at 2.48, corresponding to a $P$-value of 0.003.
annotation is completed. Of the genes within the region, cyclin G1 (CCNG1) is of particular interest, since it is a transcriptional target of p53, which is known to interact with BRCA1, localizes to nuclear foci after DNA damage, and is upregulated in breast involution (23,24).

While there are interesting candidate genes in the region identified, further studies are necessary to confirm our finding. The data suggest that a genome-wide search to identify modifiers using this approach in a larger number of families with BRCA1 mutations is feasible. In addition to being able to replicate these findings in a larger sample set, we shall be able to subset families into more genetically homogenous groups, based on mutation status and ethnicity, such as the Ashkenazi Jewish population. Genetically homogenous study groups also based on mutation status and ethnicity, such as the Ashkenazi to subset families into more genetically homogenous groups, with replicate these findings.

The data suggest that a genome-wide search to identify candidate low-penetrance cancer susceptibility genes in the general population and then evaluated in larger studies.

### MATERIALS AND METHODS

#### Patients

BRCA1 mutation carriers were ascertained through families recruited for research studies or the Cancer Risk Evaluation Program at the University of Pennsylvania. This study was reviewed and approved by the Institutional Review Board of the University of Pennsylvania.

For the initial screen, 18 families were analyzed: 12 with BRCA1 mutations and 6 with evidence of linkage to BRCA1 but no detectable mutations, with early-onset breast cancer co-segregating with polymorphic markers on 17q21. The posterior probabilities for the latter set of carrying a BRCA1 mutation ranged from 20% to 98%. Women were considered as affected for linkage purposes if they were affected with breast cancer and carriers of the deleterious BRCA1 mutation or the familial linked haplotype. Seventy-three women with breast cancer were genotyped: 44 carried a deleterious mutation in BRCA1 and 29 carried the 17q disease-associated haplotype in their family.

In the latter stage of this study (fine mapping on chromosome 5q and more stringent familial selection), 16 families with deleterious mutations and 2 families with a posterior probability greater than 90% of being due to mutations in BRCA1, but no detectable mutation, were analyzed. The 16 families with deleterious BRCA1 mutations contained the 12 families in the initial analysis and 4 additional newly identified families. Fifty-five female BRCA1 mutation carriers and 9 women who shared 17q haplotypes were typed in the

---

**Table 1. Genes located in the 10 Mb region around D5S1471: all known or predicted genes in two of the three databases examined are included**

<table>
<thead>
<tr>
<th>Gene</th>
<th>NCBI</th>
<th>Celera</th>
<th>Ensembl</th>
<th>Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCNG1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Cyclin G1</td>
</tr>
<tr>
<td>DOCX2</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Dedicator of cytokinesis 2</td>
</tr>
<tr>
<td>FBXW1</td>
<td>O</td>
<td>X</td>
<td>X</td>
<td>f-box and WD-40 domain protein 1B</td>
</tr>
<tr>
<td>FGFI3</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Fibroblast growth factor 18</td>
</tr>
<tr>
<td>FLj12899</td>
<td>O</td>
<td>X</td>
<td>—</td>
<td>Hypothetical protein similar to panthothenate kinase 1β</td>
</tr>
<tr>
<td>FLj20364</td>
<td>O</td>
<td>—</td>
<td>—</td>
<td>Hypothetical protein/arsenite-related gene 1</td>
</tr>
<tr>
<td>FOX1</td>
<td>X</td>
<td>X</td>
<td>—</td>
<td>Forkhead box 1</td>
</tr>
<tr>
<td>GABRP</td>
<td>X</td>
<td>—</td>
<td>—</td>
<td>GABA receptor π</td>
</tr>
<tr>
<td>HMMP</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>GABA receptor γ</td>
</tr>
<tr>
<td>KCNN1</td>
<td>O</td>
<td>X</td>
<td>X</td>
<td>K+ channel-interacting protein</td>
</tr>
<tr>
<td>KCNM1</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>K+ channel-interacting protein</td>
</tr>
<tr>
<td>LCP2</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Large conductance Ca2+-activated channel, Mβ1</td>
</tr>
<tr>
<td>KIAA0869</td>
<td>O</td>
<td>X</td>
<td>X</td>
<td>Lymphocyte cytosolic protein 2</td>
</tr>
<tr>
<td>LOC86683</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Methylthioltransferase protein 1</td>
</tr>
<tr>
<td>MAT2A</td>
<td>O</td>
<td>X</td>
<td>X</td>
<td>Nucleoside diphosphate kinase A</td>
</tr>
<tr>
<td>NPM1</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Odd Oz/ten-m homolog 2</td>
</tr>
<tr>
<td>ODZ2</td>
<td>X</td>
<td>X</td>
<td>—</td>
<td>Odd Oz/ten-m homolog 2</td>
</tr>
<tr>
<td>RANBP17</td>
<td>X</td>
<td>X</td>
<td>—</td>
<td>RAN-binding protein 17</td>
</tr>
<tr>
<td>RAS</td>
<td>O</td>
<td>X</td>
<td>X</td>
<td>Arginyl-tRNA synthetase</td>
</tr>
<tr>
<td>SLIT3</td>
<td>O</td>
<td>—</td>
<td>—</td>
<td>Slit homolog 3</td>
</tr>
<tr>
<td>STK10</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Serine/threonine kinase 10</td>
</tr>
<tr>
<td>TLX3</td>
<td>X</td>
<td>X</td>
<td>—</td>
<td>Homeobox 11-like2</td>
</tr>
</tbody>
</table>

X, known or predicted genes within the region represented on 2 or more of the databases; O, the genes on NT_006907.7, which was in the region on earlier NCBI builds and contains markers consistent with its location within the region; —, the known or predicted gene is not represented in database examined.

*NPM1 is located on chromosome 8 in Ensembl.*

---

Human Molecular Genetics, 2002, Vol. 11, No. 11

1330

12 families in the initial analysis and 4 additional newly identified families. Fifty-five female BRCA1 mutation carriers and 9 women who shared 17q haplotypes were typed in the
18 families. Within these families, the women with breast cancer and BRCA1 mutations or linkage to the shared 17q haplotype were contained within 22 sib pairs, 20 second-degree relative pairs, 22 third-degree relative pairs and 18 fourth-degree or greater relative pairs, for 82 relative pairs in total. So that the identity-by-descent methods could be utilized to their full advantage, unaffected parents and siblings, female and male, were genotyped where available. In total, 70 relatives without breast cancer from 15 families, those with relatives available, were typed, ranging from 2 to 8 per family. These included 38 males (13 with and 25 without BRCA1 mutations) and 32 females (5 with and 27 without BRCA1 mutations). The 5 females with BRCA1 mutations were either unaffected and under age 25 years or affected with ovarian cancer. The 16 families with disease-associated mutations in BRCA1 contained a range of mutations, including missense (C64Y), nonsense (E143X, Y1563X), frameshift (185delAG) and splice site (IVS8 + 2 T > A) mutations, as well as genomic rearrangements (intron 7-9 deletion, exon 13 duplication).

Posterior probability of linkage to BRCA1

Posterior probabilities of disease being due to a mutation in BRCA1 were calculated for the seven families that showed linkage to 17q21 without a deleterious mutation. The prior probability of breast cancer susceptibility being due to BRCA1 was assumed to be 0.8 for breast-ovarian cancer families and 0.3 for site-specific breast cancer families (27). A sequential application of Bayes’ rule incorporating lod scores and subsequently the mutation screening data, with detection sensitivities of 0.7 for CSGE and 0.85 for CSGE/Southern blotting, was used to calculate the posterior probabilities.

Markers

Genomic DNA was purified from peripheral lymphocytes and 60 ng per reaction was amplified by PCR in 15 μl containing 0.33 μM primer, 25 μM each dNTP, 1 × PCR buffer II (Perkin-Elmer) and 0.6 U AmpliTaq Gold (Perkin-Elmer). PCR cycles were as follows: 95°C for 12 min, then 10 cycles at 94°C for 15 s, annealing at 55°C for 15 s and extension at 72°C for 15 s, followed by 20 cycles at 89°C for 15 s, annealing at 55°C, and extension at 72°C with a final extension at 72°C for 10 min. Primers were fluorescently labeled and products were separated on 4% polyacrylamide gels, using an ABI Prism 377 automated sequencer, and genotypes were identified by use of ABI Genescan/Genotyper software (PE Biosystems). All genotypes were checked for Mendelian inheritance, both by inspection and by use of PEDCHECK (28). Attempts were made to resolve all inconsistent genotypes by repeating the PCR amplification and rerunning the laboratory analyses. If a genotype remained inconsistent after these checks, it was removed from the analysis. In total, 0.7% of the genotypes were removed from the analysis.

Linkage methods

In the absence of a clear understanding of the mode of inheritance of the trait of interest, we used ‘model-free’ methods for both single- and multiple-marker analyses. Single-marker analysis was performed with the SimIBD developed by Davis et al. (29), which is ideal for analyzing data sets that contain multiple pairs of affected relatives. SimIBD is based on a simulation-based IBD statistic, and uses a recursive algorithm to determine the probability of two affecteds sharing a specific allele IBD. Therefore the method is suitable for analyzing the multiplex pedigrees that we have collected. For fine mapping on chromosome 5q, SimWalk2 (v2.6) was used, and the statistic E (equivalent to the S_all in Genehunter) was calculated in order to perform multiple analyses with all the families, including those with extended pedigrees (22), since the current version of SimIBD cannot be used for multipoint analysis and the Genehunter program has limitation on pedigree size (30). Marker allele frequencies were calculated using data from one individual randomly chosen from each pedigree.

ACKNOWLEDGEMENTS

We should like to thank Betty Doan, who helped with an earlier version of these analyses, and Dr Andrew Collins for running his software ‘Beta’ to test whether there is deviation from HWE in the marker sets used in our study. This study was supported by US National Institutes of Health Grant K08 CA 84030 (K.L.N.), by the Breast Cancer Research Foundation (B.L.W.) and by the Maryland Cigarette Restitution fund (Y.Y.S.).

REFERENCES

and BRCA2 gene mutations in patients with early-onset breast cancer.


