Breast cancer accounts for 20% of all female cancers. Many risk factors have been identified but a positive family history remains one of the most important risk factors, with first-degree relatives of patients having a 2-fold elevated risk. Known breast cancer susceptibility genes such as BRCA1 and BRCA2 explain only 20-25% of this risk, suggesting the existence of other breast cancer susceptibility genes.

Here we report the results of a genome-wide linkage scan in 55 high-risk Dutch breast cancer families with no mutations in BRCA1 and BRCA2. In addition we performed CGH-analyses in 61 tumors of these families and 31 sporadic tumors. Twenty-two of these cancer families were also included in the previous linkage study by the Breast Cancer Linkage Consortium. Three regions were identified with parametric HLOD scores >1, and three with non-parametric LOD scores >1.5. Upon
further marker genotyping for the candidate loci, and the addition of another 30 families to the analysis, only the locus on chromosome 9 (9q21-22, marker D9S167) remained significant, with a non-parametric multipoint LOD score of 3.96 (parametric hLOD 0.56, α=0.18). With CGH-analyses we observed preferential copynumber loss at BAC RP11-276H19, containing D9S167 in familial tumors as compared to sporadic tumors (p<0.001). Five candidate genes were selected from the region around D9S167 and their coding regions subjected to direct sequence analysis in 16 probands. No clear pathogenic mutations were found in any of these genes.

INTRODUCTION

Breast cancer is the most commonly occurring cancer among women, accounting for 22% of all female cancers and the cumulative lifetime risk for a woman to develop breast cancer is approximately 1 in 10. Many risk factors have been identified but a positive family history remains among the most important ones established for breast cancer, with first-degree relatives of patients having an approximately 2-fold elevated risk. This risk increases with the number of affected relatives and is greater for women with relatives affected at a young age, bilateral disease or a history of benign breast disease. It is currently estimated that approximately 20-25% of this risk is explained by known breast cancer susceptibility genes, mostly those conferring high risks, such as BRCA1 and BRCA2. This suggests that other susceptibility genes remain to be found, although it is not entirely clear which genetic model explains the remainder of familial risk best. Depending on the population investigated, some of the risk could still be due to rare, moderately penetrant autosomal dominant effects, a common recessive effect, or a polygenic model. Recently the Breast Cancer Linkage Consortium (BCLC) published the results of a genome-wide linkage search for new breast cancer susceptibility genes in 149 high risk breast cancer families. The highest LOD score obtained was 1.80 under the dominant model, for a region on chromosome 4. A maximum heterogeneity-LOD (hLOD) score of 2.40 was found on chromosome arm 2p in a subset of families with four or more cases of breast cancer diagnosed below age 50 years. Other studies scanning for linkage were also unable to detect significant lod scores, but were much smaller in terms of number of families included.

The failure to detect strong linkage signals might reflect extensive locus heterogeneity, whereby multiple susceptibility loci each explain only a small proportion of families. Greater statistical linkage power might be achieved by considering subsets
families from more homogeneous populations in which the number of such loci might be reduced. We have here performed a search for linkage in a set of 55 breast cancer families of Dutch origin that are unlikely to be segregating BRCA1 or BRCA2 mutations. The cumulative lifetime risk of developing breast cancer in the Netherlands is about 1 in 9 women, which ranks among the highest worldwide. Founder effects at several major breast cancer loci have been detected in the Dutch population,\textsuperscript{12-14} as well as for many other disease genes. The assumption of reduced genetic heterogeneity for breast cancer susceptibility in the Netherlands is therefore not unrealistic. No significant lod scores were obtained in parametric analyses under a dominant or recessive model. Non-parametric (allele-sharing) analysis identified a locus on chromosome 9q21 with a multipoint npl-score of 3.96 (marker D9S167), but no clearly pathogenic mutations were detected in 5 candidate genes flanking this marker in 16 probands from families putatively linked to chromosome 9.

**MATERIALS AND METHODS**

**Family collection**

The families were ascertained through the Clinical Genetic Centers in Leiden, Rotterdam, and Nijmegen, as well as through the Netherlands Foundation for the Detection of Hereditary Tumors (stoet). The families were eligible for inclusion if there were at least three cases diagnosed with breast cancer before the age of 60, no cases of ovarian cancer, and no cases of male breast cancer.\textsuperscript{1} Polymorphic marker information had to be retrievable for at least three cases under 60, either by direct genotyping of blood samples, or by inferring from genotyped spouses and children. The resulting 55 families constituted our 'linkage search group' (208 genotyped breast cancer cases). Twenty-two of these families were also included in the genome-wide linkage search conducted by the BCLC.\textsuperscript{1} Another 30 families (119 breast cancer cases), were designated 'linkage conformation group', because they were selected on the same cancer phenotype, but differed slightly from the search group in that they did not meet the genotype or age of onset criteria. Thus, there were 4 families with two genotyped cases diagnosed before the age 60, and one diagnosed at the age 60, 18 families with three or more cases under 60, of which only two were genotyped, 3 families with three or more cases under 60, of which only one was genotyped, and 5 families with two genotyped cases diagnosed before 60, one of whom is a bilateral case (with both primaries diagnosed before age 60). To meet the ‘three cases’ sampling criteria in these families, we also genotyped cases diagnosed above 60 if they had donated a blood sample (67 breast cancer cases).
Paraffin-embedded tumor samples and pathological reports or medical reports were retrieved where available. Blood samples were collected after obtaining written informed consent. The institutional ethical committees of all of the hospitals involved approved this study.

**BRCA1 and BRCA2 mutation testing**

In each family, the youngest breast cancer patient from whom a blood sample was available was tested for mutations in the **BRCA1** and **BRCA2** genes (and for many families the next youngest as well). The joint Clinical Genetic Centers applied a variety of methodologies. The largest central exons (exon 11 in **BRCA1** and **BRCA2**, exon 10 of **BRCA2**) were scanned by protein truncation tests. The small exons were scanned for mutations by denaturing gradient gel electrophoreses (DGGE) or direct sequencing. All of the laboratories specifically assayed the presence of large founder deletions in **BRCA1** by deletion junction-PCR. The entire coding sequences of **BRCA1** and **BRCA2** were investigated by conformation-sensitive gel electrophoresis (CSGE) in families that were incompletely scanned at the time of ascertainment. Since 2002, each center offers full sequence analysis and DGGE covering the entire coding regions of both genes, and Multiple Ligation-dependent Probe Amplification (MLPA) to detect large deletions/duplications in **BRCA1**.

**Linkage analysis**

For the genome-wide linkage search, the Applied Biosystems Linkage Mapping Set MD10, consisting of 416 microsatellite markers at ~10 cM average spacing, was analyzed on an ABI3700 DNA sequencer. Additional markers were used to investigate the region of interest on chromosome 9. Genotypes were called automatically using Genemapper software and were then checked manually by two individuals. DNA from CEPH 1347-02 was typed as reference to ensure consistency of allele sizing. Allele frequencies for parametric linkage analyses were calculated based on one randomly chosen individual from each family. Multipoint linkage analyses were carried out using the program GENEHUNTER version 2.1-B. We used a model in which susceptibility to breast cancer is conferred by a dominant allele with a reduced penetrance and a population frequency of 0.003. The risk of breast cancer by age 80 was assumed to be 0.85 in carriers and 0.096 in non-carriers. Risks are modeled in seven age categories (<30, 30-39, 40-49, 50-59, 60-69, 70-79, and 80+) as described. Under the recessive model, the risk to carriers and noncarriers were identical to those under the dominant model, but the disease allele frequency was assumed to
be 0.08. We used the multipoint LOD-scores for each family to compute heterogeneity LOD scores, using the standard admixture model, and hence estimated the proportion of families (α) linked to the putative 'BRCAX' locus by maximizing the heterogeneity LOD score. Non-parametric linkage analyses were carried out by the program GENEHUNTER version 2.1-b and MERLIN version 0.9.12b. Both the singlepoint and multipoint settings were used, as well as both the 'pairs' and 'all' setting.

**Snp-genotyping**

Four single nucleotide polymorphisms were initially selected from an approximately 65-kb region surrounding D9S167. More recently, however, these snp's were repositioned 1.74 Mb distal of this marker by the human genome sequencing effort. We then selected 4 other snp's, immediately adjacent to D9S167 and covering about 17 kb within the 48-kb linkage disequilibrium-block around marker D9S167 (www.hapmap.org). These were rs12335588 (hapmap position 82,996,423), rs10867942 (83,002,124), rs11139937 (83,011,568), and rs11139938 (83,011,664). Marker D9S167 is at hapmap position 83,013,562. Primers were designed in such a way that the polymorphism would create or destroy a restriction site. PCR-products were digested by the appropriate restriction enzyme, and analysed on a 2.5% agarose gel. Results were scored by two observers independently. Data from all 8 snp’s were used to reconstruct haplotypes around D9S167.

**Chromosome 9 copy number analysis**

We performed array-CGH-analysis of 61 paraffin-embedded tumor samples from 58 patients from 27 families, using a method described previously. Similar material from 31 sporadic cases served as control. These arrays contain approximately 3,500 BAC clones, of which 13 derive from an 8-cM region of interest on chromosome 9. The BAC’s were considered to report copy number gain if the ratio of tumor derived genomic DNA compared to normal DNA exceeded 0.2 on a 2-log-scale, and copy number loss if the signal was below −0.2. The full dataset describing gains and losses on all chromosomes in this patient material will be described elsewhere (Van Beers et al., manuscript in preparation).

**Sequence analysis of candidate genes**

All known genes in an 14-cM interval D9S175-D9S167-D9S283 were retrieved from Ensemble (release 42). The cellular functions of these genes – in as much as they were known – were retrieved from OMIM. A literature search was then performed by
a computer-program dubbed ‘Anni’, which can find functional associations between large numbers of genes and other biomedical concepts (in this case ‘breast cancer’) from free-text literature. For each gene, a profile of related concepts was constructed that summarizes the context in which the gene is mentioned in literature. In addition, all genes from the region were analyzed by software termed ‘Prioritizer’. On this basis, 5 genes (of the 14 annotated genes with a known function in an approx 5-cM region around D9S167) were selected for direct sequence analysis in a set of 16 DNA samples from breast cancer patients from 16 different families. These families were selected because analysis of genotype data with the program ‘Haploviev’ had indicated that all patients share a haplotype in this region. Candidate genes were analyzed by DNA sequence analysis on the ABI3730 Prism Genetic Analyzer (Applied Biosystems, Foster City, CA)and the Mutation Surveyor software package.

**TABLE 1**

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Position</th>
<th>Model</th>
<th>HLOD or NPL LOD score</th>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>226</td>
<td>Dominant</td>
<td>1.40</td>
<td>0.12</td>
</tr>
<tr>
<td>4</td>
<td>64</td>
<td>NPL</td>
<td>1.26</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>164</td>
<td>NPL</td>
<td>1.94</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>164</td>
<td>Recessive</td>
<td>1.52</td>
<td>0.53</td>
</tr>
<tr>
<td>7</td>
<td>90</td>
<td>NPL</td>
<td>1.25</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>30</td>
<td>NPL</td>
<td>2.22</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>86</td>
<td>NPL</td>
<td>2.34</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>88</td>
<td>Recessive</td>
<td>1.18</td>
<td>0.47</td>
</tr>
<tr>
<td>9</td>
<td>90</td>
<td>Dominant</td>
<td>1.24</td>
<td>0.43</td>
</tr>
<tr>
<td>15</td>
<td>114</td>
<td>Dominant</td>
<td>1.19</td>
<td>0.29</td>
</tr>
<tr>
<td>15</td>
<td>114</td>
<td>NPL</td>
<td>1.12</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>22</td>
<td>NPL</td>
<td>1.72</td>
<td></td>
</tr>
</tbody>
</table>

Generated with the Genehunter software package. Position (cM) based on deCode map.

HLOD, heterogeneity LOD score (dominant and recessive models);
NPL, nonparametric (allele sharing) LOD score.

Alpha is the proportion of linked families in the admixture model.
We performed a genome-wide linkage analysis with 416 microsatellite markers, with an average spacing of approximately 10 cM in the group of 55 linkage search families. The highest heterogeneity LOD-score generated by GENEHUNTER under a parametric dominant model was 1.40 on chromosome 1 at position 226 cM (Table 1, Figure 1). Two other regions with hLOD-scores greater than 1.0 were identified on chromosome 9 (hLOD=1.23 at position 90 cM) and 15 (hLOD=1.19 at position 114 cM). Under a recessive model hLOD-scores >1.0 were found on chromosome 6 (164 cM) and 9 (88 cM). With non-parametric linkage analysis (nPL) we identified seven regions with a NPL-score higher than 1 (chromosome 4, 6, 7, 9, 15 and 21). The highest NPL-score found was on chromosome 9 (NPL=2.34, 86 cM, p=0.015). A second, distinct region on chromosome 9 had an NPL-score of 2.23 (30 cM, p=0.019). This second region also showed a hLOD-score >1 under the dominant model. To

### TABLE 2

Haplotypes around D9S167

<table>
<thead>
<tr>
<th>D9S167 allele (bp)</th>
<th>All families</th>
<th>Complete sharing</th>
<th>Near complete sharing†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of families</td>
<td>Number of haplotypes</td>
<td>Number of families</td>
</tr>
<tr>
<td>313</td>
<td>11</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>317</td>
<td>25</td>
<td>19</td>
<td>3</td>
</tr>
<tr>
<td>319</td>
<td>20</td>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td>321</td>
<td>29</td>
<td>22</td>
<td>6</td>
</tr>
<tr>
<td>323</td>
<td>9</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>325</td>
<td>15</td>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>327</td>
<td>5</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>329</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>331</td>
<td>9</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>333</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>335</td>
<td>13</td>
<td>14</td>
<td>4</td>
</tr>
<tr>
<td>337</td>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Totals</td>
<td>25</td>
<td>21</td>
<td>12</td>
</tr>
</tbody>
</table>

† Total number of families (sharing and non-sharing)

Total number of different haplotypes in the complete set of families

Families with >80% sharing, excluding families with complete sharing

Total number of different haplotypes in families with >80% sharing

**RESULTS**

**Genome-wide linkage scan**

We performed a genome-wide linkage analysis with 416 microsatellite markers, with an average spacing of approximately 10 cM in the group of 55 linkage search families. The highest heterogeneity LOD-score generated by GENEHUNTER under a parametric dominant model was 1.40 on chromosome 1 at position 226 cM (Table 1, Figure 1). Two other regions with hLOD-scores greater than 1.0 were identified on chromosome 9 (hLOD=1.23 at position 90 cM) and 15 (hLOD=1.19 at position 114 cM). Under a recessive model hLOD-scores >1.0 were found on chromosome 6 (164 cM) and 9 (88 cM). With non-parametric linkage analysis (nPL) we identified seven regions with a NPL-score higher than 1 (chromosome 4, 6, 7, 9, 15 and 21). The highest NPL-score found was on chromosome 9 (NPL=2.34, 86 cM, p=0.015). A second, distinct region on chromosome 9 had an NPL-score of 2.23 (30 cM, p=0.019). This second region also showed a hLOD-score >1 under the dominant model. To
evaluate these linkage signals further we genotyped an additional 30 families (confirmation group) for the microsatellite markers on chromosome 1, 6, 9, 15, and 21 at which the peak LOD scores were observed. For all these loci the evidence for linkage decreased, except for the locus on chromosome 1, for which the HLOD increased to 1.46 (α=0.13, p=0.39). The locus at position 86 cM on chromosome 9 decreased only slightly (NPL= 1.98, p=0.028) and the HLOD was 0.56 (α=0.18).

We then also generated LOD-scores using the method of Kong and Cox with the Merlin software package, because this method is less conservative when marker information is not complete (Figure 1). The multipoint LOD-score at position 86 cM on chromosome 9 was 3.06 (p=9x10⁻⁵) at marker D9S167 in the 55 families, and increased to 3.96 (p=10⁻⁵) when the other 30 families were added to the analysis. The single-point LOD-score over all 85 families for D9S167 was 4.63 (p=10⁻⁶). To evaluate this region on chromosome 9 further we genotyped 4 additional microsatellite markers, i.e., D9S1843 and D9S1674 proximal of D9S167, and D9S1865 and D9S1812 distal of it, defining a 9.4 cM-region. With these additional markers the multipoint LOD-score calculated by Merlin at D9S167 in the 85 families declined to 3.02, while those at D9S1843 and D9S287 were below 1.5 (Figure 2). This defined the linked region to be between the markers D9S1674 and D9S287 (~15.7 cM).

Haplotype analysis
To aid haplotyping around D9S167, we genotyped 8 SNP’s, 4 of which immediately proximal of D9S167 in a ~48-kb LD-block, and 4 covering a 65-kb region about 1.7 Mb distal of D9S167. We analyzed haplotype-sharing in each family with the program ‘Haploview’. In 32 families all genotyped patients shared an allele at D9S167, but in 5 families this allele was on a different haplotype, indicating that the shared alleles were not identical by descent (IBD). In another four families not all patients were successfully genotyped at D9S167, but in two of those the patients shared a haplotype from D9S1674 to D9S1812, suggesting allele-sharing at D9S167. Thus, in 25 families all patients shared an allele at D9S167 IBD (Table 2), in total comprising 10 different alleles (range: 1 – 6 families per allele) on 21 different haplotypes. In 50 families there was no sharing of an allele among genotyped patients (in 3 families the marker data did not allow phasing of the haplotypes). In 12 of these 50 families, more than 80% of the patients shared the same haplotype (4 out of 5 patients in 5 families, 5 out of 6 in 4 families, and 6 out of 7, 7 out of 8, 8 out of 10 in 1 family each). Again, 8 different D9S167-alleles were shared on a total of 12 different haplotypes (Table 2). Overall, the 12 different D9S167-alleles found to be shared either
Figure 1. HLOD scores by chromosome for the dominant model (blue line), as computed by GENEHUNTER, and nonparametric LOD scores (red line), as computed by MERLIN, in 55 breast cancer families.

Fig. 2. Nonparametric LOD scores as computed by Merlin, in all families (blue line), and with additional markers (red line). The dashed lines represent the -1 LOD interval.
completely or almost completely, did so on 29 different haplotypes. Although, depending on the number of markers considered around D9S167, a suggestive core haplotype could sometimes be discerned between two or more haplotypes (data not shown). These results support the NPL scores for D9S167, but also indicate extensive allelic heterogeneity for this sharing, as well as genetic heterogeneity across families because not all families contribute to the NPL score.

Candidate gene analysis
We selected five genes from the region between markers D9S1843 and D9S283, on the basis of their presumed cellular function (see Materials & Methods). These included \textit{ubqln1}, \textit{rasef}, \textit{dapk1}, \textit{tle1}, and \textit{gadd45γ}. The entire coding regions of these genes were sequenced in 16 patients from 16 families displaying complete haplotype sharing at D9S167. Nineteen variants were found in one or more patients (Table 3), 11 of which were known SNP’s. For several variants we detected homozygotes for both alleles, making them unlikely candidates for susceptibility alleles. Of all the exonic variants found, there was only one missense change, in \textit{tle1} in one family. This variant did not co-segregate with disease. Three of the 6 intronic variants were known SNP’s, and none were predicted to affect the nearest splice-site. The latter was also found for all the neutral exonic changes. We conclude that no clear disease-related changes were detected in this screen.

cgh-analysis
Copy number changes of the linked region on chromosome 9 were investigated in 61 tumors from 27 families by examining the intensity ratios of the 13 BAC clones representing this region on the array (Table 4). We were able to compare the results of 22 tumors from 10 families that displayed complete sharing of a 8-cM haplotype around D9S167 in all patients (‘linked tumors’), with 39 tumors from 17 families without such haplotype sharing (‘unlinked tumors’). A high proportion of linked tumors (55%) showed copy number loss at a BAC RP11-276H19 containing the \textit{gas1} gene and D9S167, and none showed gain. However, this was not statistically different from the unlinked tumors in which 31% showed copy loss at this BAC. We did observe a significant difference in the percentage of \textit{brcax} tumors with loss of this BAC as compared to sporadic tumors (average \textit{t-log} ratio of \(-0.208\) as compared with a \textit{t-log} ratio of \(0.088\) for control tumors). This difference has a two-tailed unpaired t-test p-value of \(0.00039\).
### TABLE 3

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Gene Change</th>
<th>Exon / intron</th>
<th>Times found heterozygous</th>
<th>Known SNP</th>
<th>Splice-site prediction</th>
<th>Co-segregation</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAPK1</td>
<td>c.393C&gt;T, p.His131His</td>
<td>Exon 4</td>
<td>1</td>
<td>No</td>
<td>No change</td>
<td>NI</td>
</tr>
<tr>
<td></td>
<td>TTCA(G/A)GAT, 143,481A&gt;AG, p.267Gln-&gt;Gln</td>
<td>Exon 9</td>
<td>1 (and 2 minor homozygotes)</td>
<td>No</td>
<td>No change</td>
<td>NI</td>
</tr>
<tr>
<td></td>
<td>g.144573A&gt;AG</td>
<td>Intron 9</td>
<td>7 (and 3 minor homozygotes)</td>
<td>rs3118846</td>
<td>No change</td>
<td>NI</td>
</tr>
<tr>
<td></td>
<td>g.150139G&gt;AG</td>
<td>Intron 14</td>
<td>4</td>
<td>rs12274607</td>
<td>No change</td>
<td>NI</td>
</tr>
<tr>
<td></td>
<td>c.1008G&gt;T, p.Asp337Asp</td>
<td>Exon 16</td>
<td>13</td>
<td>rs3181854</td>
<td>No change</td>
<td>NI</td>
</tr>
<tr>
<td></td>
<td>c.1830G&gt;A, p.Gly610Gly</td>
<td>Exon 18</td>
<td>2</td>
<td>No</td>
<td>No change</td>
<td>NI</td>
</tr>
<tr>
<td></td>
<td>g.201485G&gt;T</td>
<td>Intron 23</td>
<td>3</td>
<td>rs3128495</td>
<td>No change</td>
<td>NI</td>
</tr>
<tr>
<td></td>
<td>c.3597C&gt;T, p.Arg1200Arg</td>
<td>Exon 26</td>
<td>10</td>
<td>rs3118863</td>
<td>No change</td>
<td>NI</td>
</tr>
<tr>
<td>TLE1</td>
<td>c.983C&gt;T, p.Thr328Met</td>
<td>Exon 12</td>
<td>1</td>
<td>No</td>
<td>No change</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>c.1101A&gt;G, p.Pro367Pro</td>
<td>Exon 13</td>
<td>1</td>
<td>No</td>
<td>No change</td>
<td>No</td>
</tr>
<tr>
<td>GADD45G</td>
<td>c.102+65G&gt;C</td>
<td>Intron 1</td>
<td>4</td>
<td>rs3135852</td>
<td>No change</td>
<td>NI</td>
</tr>
<tr>
<td></td>
<td>c.157-18_19delCC-18_19insCTAG</td>
<td>Intron 2</td>
<td>2</td>
<td>No</td>
<td>No change</td>
<td>No</td>
</tr>
<tr>
<td>RASEF</td>
<td>c.785C&gt;T, p.Arg262Cys</td>
<td>Exon 5</td>
<td>10</td>
<td>rs4146960</td>
<td>No change</td>
<td>NI</td>
</tr>
<tr>
<td></td>
<td>c.1202+57_61delGTAAA</td>
<td>Intron 9</td>
<td>6</td>
<td>No</td>
<td>No change</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>c.1731T&gt;G, p.Asp577Asp</td>
<td>Exon 13</td>
<td>1</td>
<td>rs34303676</td>
<td>No change</td>
<td>NI</td>
</tr>
<tr>
<td></td>
<td>c.2223+18A&gt;T</td>
<td>Intron 17</td>
<td>1</td>
<td>No</td>
<td>No change</td>
<td>No</td>
</tr>
<tr>
<td>UBQLN1</td>
<td>c.1949C&gt;T, p.Ser499Ser</td>
<td>Exon 10</td>
<td>5 (and 1 minor homozygote)</td>
<td>rs2781104</td>
<td>No change</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>c.1590C&gt;A, p.Leu530Leu</td>
<td>Exon 10</td>
<td>7 (and 1 minor homozygote)</td>
<td>rs7866234</td>
<td>No change</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>c.1617+157G&gt;A</td>
<td>Intron 10</td>
<td>8</td>
<td>rs9314722</td>
<td>No change</td>
<td>No</td>
</tr>
</tbody>
</table>

1 Out of 16 patients tested
2 Using splice-prediction programs NNSPLICE version 0.9 by Neural Network (www.fruitfly.org/seq_tools/splice.html), NetGene 2 Server version 2.42 by CBS Software Package Manager (www.cbs.dtu.dk/services/NetGene2) and Alex Dong Li’s SpliceSiteFinder (http://violin.genet.sickkids.on.ca/~ali/splicessitefinder.html)
3 NI = Not investigated
DISCUSSION

The analysis reported here represents the largest single-center genome-wide linkage search for new susceptibility loci in non-\textit{BRCA1/2} breast cancer families to date. The rationale for this study was that there exist further breast cancer genes which confer moderate to high risks \( (6,28) \). The patterns of familial clustering in the families that we selected for our study suggest that such alleles are likely to be dominant. Initial suggestive linkage peaks observed in a ‘linkage search’ group of 55 families were subsequently confirmed and confined to a locus on chromosome 9 in a total set of 85 families. Linkage evidence was most apparent using allele-sharing analyses with the Merlin package, with a single point non-parametric lodscore of 4.63 and a multipoint score of 3.96 at marker D9S167. These \textit{npl} scores represent the highest for any single locus in a linkage search after the identification of \textit{BRCA1} and \textit{BRCA2}, but they are dif-
ficult to compare with previous studies because these have mainly analysed marker
data with the more conservative GENEHUNTER software. Under a parametric domi-
nant model, however, the hlod at D9S167 was 0.56 (a=0.18), indicating that even
though the allele-sharing at D9S167 was significant in the total set of families, most
families did not support linkage to this locus. Indeed, we noted complete allele-
sharing among patients among 25 of the 85 families (29%) at D9S167, and suggestive
incomplete sharing in 12 other families (14%). We observed extensive haplotype
heterogeneity around shared alleles at D9S167. One explanation for this is that there
is a gene (or genes) near D9S167 in which multiple rare variants confer substantially
increased risks to breast cancer. The low hlod score at this marker is probably due to
the small number of families demonstrating complete haplotype sharing among the
patients in conjunction with the fact that for most families and patients the (founder)
parents were unavailable for genotyping. In the merlin analysis, the overall informa-
tion content at D9S167 in the 85 families was 42%, which is in agreement with simu-
lation studies with microsatellite maps of ~10 cM and incomplete parental geno-
types.29 Further genotyping of the region at much higher resolutions could therefore
help to identify regions with more consistent allele-sharing.

Recently the Breast Cancer Linkage Consortium published the results of a genome
wide linkage search for breast cancer susceptibility genes,1 which included 149 mul-
tiple case non-brca1/2 breast cancer families. The highest lod score under the domi-
nant model was 1.80, for a region on chromosome 4. Although several other sug-
gestive lod scores were reported, the number of linkage peaks did not differ from
the number expected by chance and therefore these peaks probably reflect the play
of chance rather than true susceptibility loci. In agreement with the bclc-study we
found no evidence for linkage to markers on 2q32,10 3p26,30 8p12-22,31 10q23.32-
q25.3,11 11q23,33 13q21-22 and 22q13.1,30,34 which were all previously suggested to har-
bour susceptibility loci. But we also did not find any evidence for linkage on the re-
gions reported by the bclc-publication. The inability to detect strong linkage signals
may be a reflection of extensive locus heterogeneity.

The bclc analysis1 included 22 Dutch families that were also part of the 55 families
investigated here. The npl score near the D9S167 locus in that study was 0.74, and
although this was the second highest score for the Dutch families, there was no evi-
dence for allele-sharing at this locus in the other 127 families collected in that study,
derived from Australia, United Kingdom, USA, Canada and France. Of the 22 Dutch
families, 10 showed allele-sharing at D9S167, but 2 of these on different haplotypes.
It is possible that our linkage study might have achieved greater statistical power
because the families derive from a more homogeneous population (i.e., with reduced genetic heterogeneity). The Dutch population exhibits distinct founder mutations for several known cancer susceptibility genes, and therefore could be considered, to an extent, to be genetically distinct. Although a strong founder effect at the 9q-locus seems less likely, given the diversity of shared haplotypes, such an effect has also been observed at BRCA1 in the presence of extensive allelic heterogeneity.

In a genome-wide scan for linkage in 14 Finnish breast cancer families, the second highest HLOD peak was found at D9S283, just 5 cM distal of D9S167. D9S167 was also shown to be linked (with a multipoint parametric LOD score of 3.02) to ocular melanoma in three Danish families with multiple cases of ocular malignant melanoma, cutaneous malignant melanoma and other malignancies, including breast cancer. In addition, D9S167 was in the center of a small chromosomal deletion in a case of acute myeloid leukemia. These results suggest there is a gene in this region that can be linked to cancer susceptibility. Further evidence that the 9q21 region may be involved in a subset of the familial form of breast cancer comes from our observation that over 50% of BRCAx tumors putatively linked to 9q21 show copy-number loss at this locus, as opposed to 5-25% in sporadic breast tumors. Others did not observe excess copy-number losses of 9q in familial non-BRCA1/2 breast tumors using classical CGH, although these patients were selected under different criteria than our cases. We previously reported that ~30% of the same set of BRCAx tumors showed loss of heterozygosity (or allelic imbalance) at a marker for 9q34, which is not significantly higher than found in sporadic tumors. Because that marker is a long distance away from D9S167, it is possible that some of the copy-number losses in the BRCAx tumors are tightly localized around 9q21.

The number of genes between D9S1843 and D9S283 presently annotated is 49. We performed sequence analysis of 5 of these to search for possible susceptibility alleles. No clear pathogenic changes were found in any of them. For all genes an apparent link with tumorigenesis could be made, such as a Ras GTPase motif in the RAFF gene (closest to D9S167), transcription regulation (TLE1), or involvement in apoptosis (DAPK1) or stress response (GADD45g). However, a direct link with breast cancer has not yet been established for most of these candidates. In sporadic breast cancer, the expression of RASEF at mRNA-level is apparently not reduced. TLE1 has been suggested to play a role during epithelial differentiation and tumor progression through inhibition of the Wnt–CTNNB1 signaling pathway. DAPK1 and GADD45G are frequently targeted by inactivation through promotor hypermethylation in leukemias, lymphomas and a number of epithelial cancers.
In conclusion, through linkage analysis we have identified a region on 9q21 which shows significant haplotype sharing among patients belonging to non-BRCA1/2 families with at least three cases of breast cancer diagnosed before age 60. However, we observed extensive haplotype diversity at the shared locus, but have not yet identified sequence variants in candidate genes that could explain these results. There was some suggestion that the somatic genetic changes at this locus differ from that seen in sporadic breast tumors, which will have to be confirmed in larger series.

ACKNOWLEDGEMENTS

We want to thank Barbara Veldhuizen for the assistance with the Linkage analyses and Ben Oostra, Eline Slagboom and Mike Stratton for providing their genotype platforms. This research project was financially supported by the Dutch Cancer Society (KWF, RUL1999-2021).

REFERENCE LIST


