
Universiteit
Leiden
The Netherlands

# Search for new breast cancer susceptibility genes

Oldenburg, R.A.

**Citation**
Oldenburg, R. A. (2008, May 29). *Search for new breast cancer susceptibility genes*. Retrieved from https://hdl.handle.net/1887/12871

| | |
|---|---|
| Version: | Corrected Publisher's Version |
| License: | Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden |
| Downloaded from: | https://hdl.handle.net/1887/12871 |

**Note:** To cite this publication please use the final published version (if applicable).

## 5.2. GENOME-WIDE LINKAGE SCAN IN DUTCH HEREDITARY NON-BRCA1/2 BREAST CANCER FAMILIES IDENTIFIES 9q21-22 AS A PUTATIVE BREAST CANCER SUSCEPTIBILITY LOCUS

Rogier A. Oldenburg,[1,2,] Karin H.G. Kroeze-Jansema,[1] Jeanine J. Houwing-Duistermaat,[6] Jean-Pierre Bayley,[1] Cheryl Dambrot,[1] Christi J. van Asperen,[1] Ans M.W. van den Ouweland,[2] Bert Bakker,[1] Erik H. van Beers,[7] Petra M. Nederlof,[8] Hans Vasen,[4] Nicoline Hoogerbrugge,[3] Cees J. Cornelisse,[5] Hanne Meijers-Heijboer[8] and Peter Devilee[1,5]

**Submitted**

[1] *Center for Human and Clinical Genetics, Leiden University Medical Center, Leiden*
[2] *Dept. of Clinical Genetics, Erasmus Medical center, Rotterdam*
[3] *Dept. of Human Genetics, University Medical Center Nijmegen*
[4] *Netherlands Foundation for the Detection of Hereditary Tumors, Leiden*
[5] *Dept. of Pathology, Leiden University Medical Center, Leiden*
[6] *Dept. of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden*
[7] *Division of Experimental Therapy, Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam*
[8] *Department of Pathology, Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands*

**ABSTRACT**

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.[1] 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.[2] 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.[3] 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.[4,5] 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*.[6] 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.[7-9] 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.[1] 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.[1] Other studies scanning for linkage were also unable to detect significant lod scores, but were much smaller in terms of number of families included.[10,11]

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 of

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,[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.[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.[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.[15] The small exons were scanned for mutations by denaturating 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.[13] 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.[16] 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*.[17]

### 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.[1] 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.[18] 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.[19,20] 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.[20] 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.[21] 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.22 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.[23,24] 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

**TABLE 1**

**Maximum LOD scores in 55 breast cancer families by chromosome under the dominant, nonparametric and recessive model.**

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

Generated with the Genehunter software package. Position (cM) based on deCode map.
HLOD, heterogeneity LOD score (dominant and recessive models);
NPL, nonparametic (allele sharing) LOD score.
Alpha is the proportion of linked families in the admixture model.

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.[25] 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'.[26] 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 'Haploview'[27] 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 2**

**Haplotypes around D9S167**

| | All families | | Complete sharing | | Near complete sharing[3] | | |
|---|---|---|---|---|---|---|---|
| D9S167 allele (bp) | Number of families[1] | Number of haplotypes[2] | Number of families | Number of haplotypes | Number of families | Number of haplotypes | Total nr. of haplotypes[4] |
| 313 | 11 | 7 | 2 | 1 | 2 | 2 | 2 |
| 317 | 25 | 19 | 3 | 2 | 3 | 3 | 4 |
| 319 | 20 | 11 | 2 | 2 | 1 | 1 | 3 |
| 321 | 29 | 22 | 6 | 6 | 2 | 2 | 7 |
| 323 | 9 | 8 | 2 | 1 | 0 | 0 | 1 |
| 325 | 15 | 12 | 3 | 2 | 0 | 0 | 2 |
| 327 | 5 | 7 | 1 | 1 | 1 | 1 | 1 |
| 329 | 2 | 2 | 0 | 0 | 1 | 1 | 1 |
| 331 | 9 | 9 | 1 | 1 | 0 | 0 | 1 |
| 333 | 3 | 3 | 0 | 0 | 1 | 1 | 1 |
| 335 | 13 | 14 | 4 | 4 | 1 | 1 | 5 |
| 337 | 3 | 3 | 1 | 1 | 0 | 0 | 1 |
| Totals | | | 25 | 21 | 12 | 12 | 29 |

[1] Total number of families (sharing and non-sharing)

[2] Total number of different haplotypes in the complete set of families

[3] Families with > 80% sharing, excluding families with complete sharing

[4] 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'.[27] 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 succesfully 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

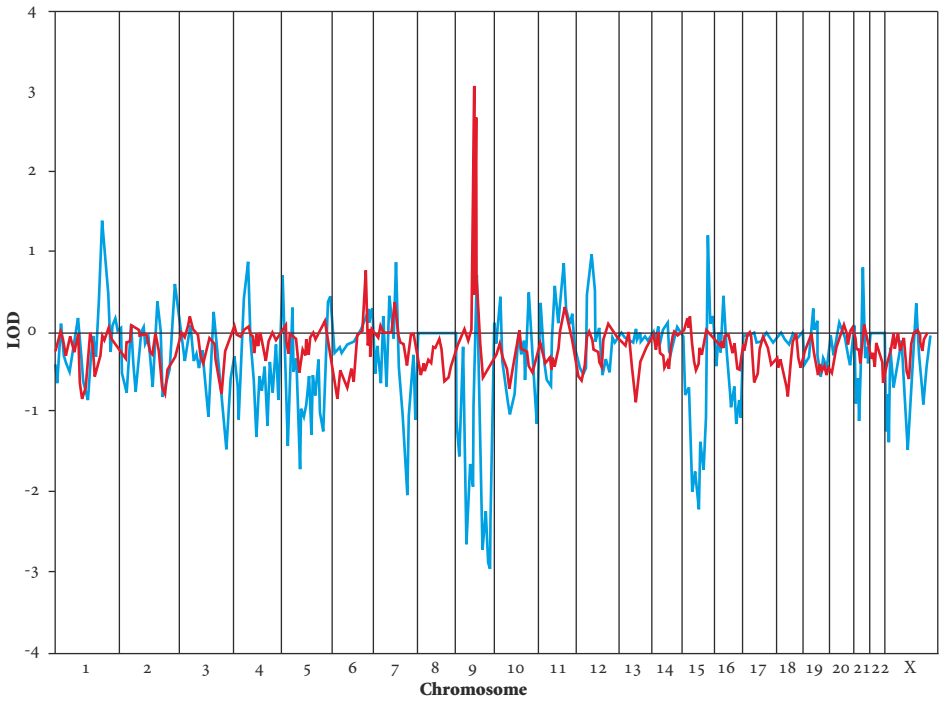


**Figure 1.** HLOD scores by chromosome for the dominant model (blue line), as computed by GENEHUNTER, and nonparametric LOD scores (red line), as computed by MERLIN, in 55 breast cancer families.

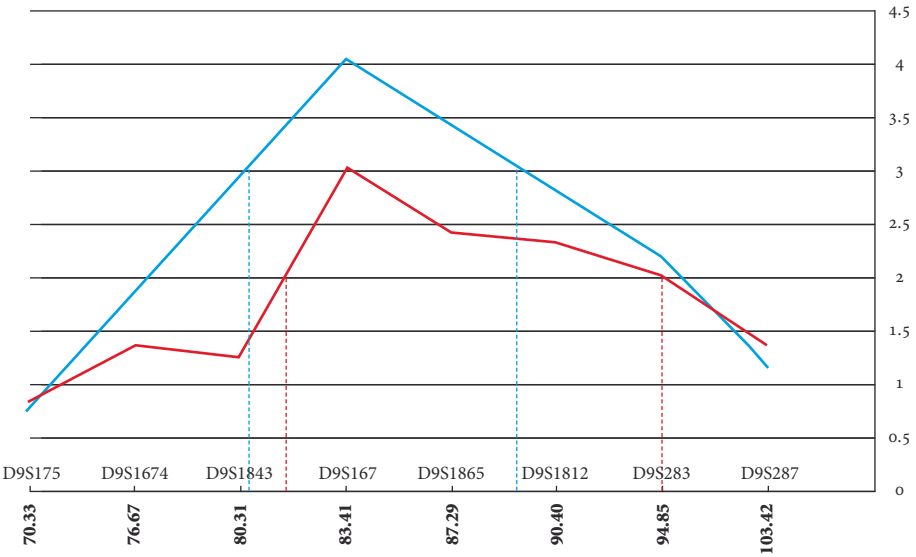


**Fig. 2.** Nonparametric LOD scores as computed by Merlin, in all families (blue line), and with additional markers (red line). The dashed lines represent the -1 LOD interval.

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 *UBQLN1*, *RASEF*, *DAPK1*, *TLE1*, and *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 *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 examing 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 *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 *BRCAX* tumors with loss of this BAC as compared to sporadic tumors (average $^2$log ratio of −0.208 as compared with a $^2$log ratio of 0.088 for control tumors). This difference has a two-tailed unpaired t-test p-value of 0.00039.

**TABLE 3**

## Gene changes detected in sequence analysis of 5 candidate genes

| Gene name | Gene Change | Exon / intron | Times found heterozygous[1] | Known SNP | Splice-site prediction[2] | Co-segregation[3] |
|---|---|---|---|---|---|---|
| DAPK1 | c.393C>T, p.His131His | Exon 4 | 1 | No | No change | NI |
| | TTCA(G/A)GAT, 143481A>AG, p.267Gln>Gln | Exon 9 | 1 (and 2 minor homozygotes) | No | No change | NI |
| | g.144573A>AG | Intron 9 | 7 (and 3 minor homozygotes) | rs3118846 | No change | NI |
| | g.150139G>AG | Intron 14 | 4 | rs2274607 | No change | NI |
| | c.1608C>T, p.Asp537Asp | Exon 16 | 13 | rs3818584 | No change | NI |
| | c.1830G>A, p.Gly610Gly | Exon 18 | 2 | No | No change | NI |
| | g.201463G>T | Intron 23 | 3 | rs3128495 | No change | NI |
| | c.3597C>T, p.Arg1200Arg | Exon 26 | 10 | rs3118863 | No change | NI |
| TLE1 | c.983C>T, p.Thr328Met | Exon 12 | 1 | No | No change | No |
| | c.1101A>G, p.Pro367Pro | Exon 13 | 1 | No | No change | NI |
| GADD45G | c.102+65G>C | Intron 1 | 4 | rs3138502 | No change | NI |
| | c.157-18_19delCC;-18_19insCTAG | Intron 2 | 2 | No | No change | No |
| RASEF | c.785C>T, p.Arg262Cys | Exon 5 | 10 | rs4146960 | No change | NI |
| | c.1202+57_61delGTAAA | Intron 9 | 6 | No | No change | No |
| | c.1731 T>G, p.Asp577Asp | Exon 13 | 1 | rs34303676 | No change | NI |
| | c.2223+18A>T | Intron 17 | 1 | No | No change | No |
| UBQLN1 | c.1494C>T, p.Ser498Ser | Exon 10 | 5 (and 1 minor homozygote) | rs2781004 | No change | No |
| | c.1590C>A, p.Leu530Leu | Exon 10 | 7 (and 1 minor homozygote) | rs7866234 | No change | No |
| | c.1617+17G>A | Intron 10 | 8 | rs9314722 | No change | No |

[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/splicesitefinder.html)

[3] NI = Not investigated

**TABLE 4**

**Results of array-CGH analysis for 13 BACs in the region 9q21-q22**

| BAC | Mb | gene | 9q-linked tumors[1] | | | 9q-nonlinked tumors[2] | | |
|---|---|---|---|---|---|---|---|---|
| | | | gain | loss | inc | gain | loss | inc |
| RP11-66D1 | 83 | TLE1 | 5% | 14% | 81% | 13% | 23% | 64% |
| RP11-432M2 | 84 | | 9% | 0% | 91% | 26% | 3% | 72% |
| RP11-541F16 | 84,7 | RASEF | 5% | 10% | 86% | 0% | 5% | 95% |
| RP11-439A18 | 85,3 | UBQLN1/HNRPK | 32% | 7% | 62% | 13% | 13% | 74% |
| RP11-59M22 | 86,3 | | 18% | 0% | 82% | 23% | 5% | 72% |
| RP11-172F7 | 87,1 | | 64% | 23% | 13% | 59% | 18% | 23% |
| RP11-280P22 | 87,9 | | 41% | 7% | 52% | 31% | 5% | 64% |
| RP11-276H19 | 88,8 | GAS1 | 0% | 55% | 45% | 8% | 31% | 62% |
| RP11-423O13 | 88,9 | | 0% | 18% | 82% | 0% | 8% | 92% |
| RP11-40C6 | 89,3 | | 9% | 10% | 81% | 15% | 13% | 72% |
| RP11-249H20 | 89,4 | DAPK1 | 9% | 15% | 76% | 13% | 3% | 85% |
| RP11-65B23 | 89,6 | CCRK | 18% | 17% | 65% | 26% | 10% | 64% |
| RP11-8B23 | 91,7 | GADD45 | 5% | 29% | 67% | 0% | 28% | 72 |

[1] Group of 22 breast tumors from 10 families in which all genotyped patients shared a haplotype around D9S167;

[2] Group of 39 breast tumors from 17 families without sharing of a haplotype around D9S167.
gain, ratio tumor/normal > 0.2 on a 2log-scale; loss, ratio tumor/normal less than –0.2 on a 2log-scale; inc, inconclusive (ratios between –0.2 and 0.2)

## DISCUSSION

The analysis reported here represents the largest single-center genome-wide linkage search for new susceptibility loci in non-*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 NPL scores represent the highest for any single locus in a linkage search after the identification of *BRCA1* and *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 dominant model, however, the HLOD at D9S167 was 0.56 (α=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 information content at D9S167 in the 85 families was 42%, which is in agreement with simulation studies with microsatellite maps of ~10 cM and incomplete parental genotypes.[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 multiple case non-*BRCA1/2* breast cancer families. The highest LOD score under the dominant model was 1.80, for a region on chromosome 4. Although several other suggestive 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,[31,32] 10q23.32-q25.3,[11] 11q23,[30] 13q21[33] and 22q13.1,[30,34] which were all previously suggested to harbour susceptibility loci. But we also did not find any evidence for linkage on the regions reported by the BCLC-publication. The inability to detect strong linkage signals may be a reflection of extensive locus heterogeneity.

The BCLC analysis[1] 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 evidence 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,[12,14,35,36] and therefore could be considered, to an extent, to be genetically distinct.[37] 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.[12,13]

In a genome-wide scan for linkage in 14 Finnish breast cancer families,[10] 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[38] 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.[39] 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.[40,41] Others did not observe excess copy-number losses of 9q in familial non-*BRCA1/2* breast tumors using classical CGH,[33,42] although these patients were selected under different criteria than our cases. We previously reported[43] 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.[44] 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 *RASEF*-gene (closest to D9S167), transcription regulation (*TLE1*), or involvement in apoptosis (*DAPK1*) or stress response (*GADD45γ*).[45-47] 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.[38] *TLE1* has been suggested to play a role during epithelial differentiation[48] and tumor progression through inhibition of the *Wnt–CTNNB1* signaling pathway.[45] *DAPK1* and *GADD45G* are frequently targeted by inactivation through promotor hypermethylation in leukemias, lymphomas and a number of epithelial cancers.[47,49]

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 financialy supported by the Dutch Cancer Society (KWF, RUL1999-2021).

**REFERENCE LIST**

1. **Smith P, McGuffog L, Easton DF, et al.** Genome wide linkage search for breast cancer susceptibility genes. *Genes Chrom Cancer; 45:646-55 (2006)*
2. **Parkin DM, Pisani P, and Ferlay J.** Global cancer statistics. *CA Cancer J Clin; 49: 33-64 (1999)*
3. **Collaborative Group on Hormonal Factors in Breast Cancer Familial breast cancer** Collaborative reanalysis of individual data from 52 epidemiological studies including 58209 women with breast cancer and 101986 women without the disease. *Lancet 358:1389-99 (2001)*
4. **McPherson K, Steel CM, and Dixon JM.** ABC of breast diseases. Breast cancer-epidemiology, risk factors, and genetics. *Br Med J; 321:624-8 (2000)*
5. **Thompson D and Easton D.** The genetic epidemiology of breast cancer genes. *Journal of Mammary Gland Biology and Neoplasia; 9:221-36 (2004)*
6. **Peto J, Collins N, Barfoot R, et al.** Prevalence of BRCA1 and BRCA2 gene mutations in patients with early- onset breast cancer. *J Natl Cancer Inst; 91:943-9 (1999)*
7. **Antoniou AC, Pharoah PDP, McMullan G, et al.** A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. *Br J Cancer; 86: 76-83 (2002)*
8. **Antoniou AC and Easton DF.** Polygenic inheritance of breast cancer: Implications for design of association studies. *Genet Epidemiol; 25:190-202 (2003)*
9. **Cui J, Antoniou AC, Dite GS, et al.** After BRCA1 and BRCA2-What Next? Multifactorial Segregation Analyses of Three-Generation, Population-Based Australian Families Affected by Female Breast Cancer. *Am J Hum Genet; 68:420-31 (2001)*
10. **Huusko P, Juo SH, Gillanders E, et al.** Genome-wide scanning for linkage in Finnish breast cancer families. *Eur J Hum Genet; 12:98-104 (2004)*
11. **Bergman A, Karlsson P, Berggren J, et al.** Genome-wide linkage scan for breast cancer susceptibility loci in Swedish hereditary non-BRCA1/2 families: suggestive linkage to 10q23.32-q25.3. *Genes Chrom Cancer; 46:302-9 (2007)*

12. **Peelen T, Van Vliet M, Petrij-Bosch A, et al.** 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; 60:1041-49 (1997)*
13. **Petrij-Bosch A, Peelen T, Van Vliet M, et al.** BRCA1 genomic deletions are major founder mutations in Dutch breast cancer patients. *Nature Genet; 17:341-5 (1997)*
14. **Broeks A, Urbanus JHM, De Knijff P, et al.** IVS10-6T > G, an ancient ATM germline mutation linked with breast cancer. *Hum Mutat; 21:521-8 (2003)*
15. **Hogervorst F, Cornelis R, Bout M, et al.** Rapid detection of BRCA1 mutations by the Protein Truncation Test. *Nature Genet; 10:208-12 (1995)*
16. **Korkko J, Annunen S, Pihlajamaa T, Prockop DJ, and Ala-Kokko L.** Conformation sensitive gel electrophoresis for simple and accurate detection of mutations: comparison with denaturing gradient gel electrophoresis and nucleotide sequencing. *Proc Natl Acad Sci USA; 95:1681-5 (1998)*
17. **Hogervorst FBL, Nederlof PM, Gille JJP, et al.** Large genomic deletions and duplications in the BRCA1 gene identified by a novel quantitative method. *Cancer Res; 63:1449-53 (2003)*
18. **Kruglyak L, Daly MJ, Reeve-Daly MP, and Lander ES.** Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet; 58:1347-63 (1996)*
19. **Claus EB, Risch N, and Douglas W.** Genetic analysis of breast cancer in the cancer and steroid hormone study. *Am J Hum Genet; 48:232-42 (1991)*
20. **Easton D, Bishop D, Ford D, Crockford G, and Breast Cancer Linkage Consortium** Genetic linkage analysis in familial breast and ovarian cancer: Results from 214 families. *Am J Hum Genet; 52:678-701 (1993)*
21. **Abecasis GR, Cherny SS, Cookson WO, and Cardon LR.** Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet; 30:97-101 (2002)*
22. **Ke X, Collins A, and Ye S.** PIRA PCR designer for restriction analysis of single nucleotide polymorphisms. *Bioinformatics; 17:838-9 (2001)*
23. **van Beers EH, Joosse SA, Ligtenberg MJ, et al.** A multiplex PCR predictor for aCGH success of FFPE samples. *Br J Cancer; 94:333-7 (2006)*
24. **Joosse SA, van Beers EH, and Nederlof PM.** Automated array-CGH optimized for archival formalin-fixed, paraffin-embedded tumor material. *Bmc Cancer; 7:43 (2007)*
25. **Jelier R, Jenster G, Dorssers LC, et al.** Text-derived concept profiles support assessment of DNA microarray data for acute myeloid leukemia and for androgen receptor stimulation. *BMC Bioinformatics; 8:e14 (2007)*
26. **Franke L, Bakel H, Fokkens L, de Jong ED, Egmont-Petersen M, and Wijmenga C.** Reconstruction of a functional human gene network, with an application for prioritizing positional candidate genes. *Am J Hum Genet; 78:1011-25 (2006)*
27. **Barrett JC, Fry B, Maller J, and Daly MJ.** Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics; 21:263-5 (2005)*
28. **Antoniou AC and Easton DF.** Models of genetic susceptibility to breast cancer. *Oncogene; 25:5898-905 (2006)*
29. **Evans DM and Cardon LR.** Guidelines for genotyping in genomewide linkage studies: single-nucleotide-polymorphism maps versus microsatellite maps. *Am J Hum Genet; 75:687-92 (2004)*
30. **Hartikainen JM, Tuhkanen H, Kataja V, et al.** An autosome-wide scan for linkage disequilibrium-based association in sporadic breast cancer cases in eastern Finland: three candidate regions found. *Cancer Epidemiol Biomarkers Prev; 14:75-80 (2005)*

31. **Kerangueven F, Essioux L, Dib A, et al.** Loss of heterozygosity and linkage analysis in breast carcinoma: Indication for a putative third susceptibility gene on the short arm of chromosome 8. *Oncogene; 10:1023-26 (1995)*
32. **Seitz S, Rohde K, Bender E, et al.** Strong indication for a breast cancer susceptibility gene on chromosome 8p12-p22: linkage analysis in German breast cancer families. *Oncogene; 14:741-3 (1997)*
33. **Kainu T, Juo SH, Desper R, et al.** Somatic deletions in hereditary breast cancers implicate 13q21 as a putative novel breast cancer susceptibility locus. *Proc Natl Acad Sci USA; 97:9603-8 (2000)*
34. **Hartikainen JM, Tuhkanen H, Kataja V, et al.** Refinement of the 22q12-q13 breast cancer-associated region: Evidence of TMPRSS6 as a candidate gene in an eastern Finnish population. *Clinical Cancer Research;12:1454-62 (2006)*
35. **Gruis NA, Van der Velden PA, Sandkuijl LA, et al.** Homozygotes for CDKN2 (p16) germline mutation in Dutch familial melanoma kindreds. *Nat Genet; 10:351-3 (1995)*
36. **Taschner PEM, Jansen JC, Baysal BE, et al.** Nearly all hereditary paragangliomas in the Netherlands are caused by two founder mutations in the SDHD gene. *Genes Chrom Cancer; 31:274-81 (2001)*
37. **Zeegers MP, van Poppel F, Vlietinck R, Spruijt L, and Ostrer H.** Founder mutations among the Dutch. *Eur J Hum Genet; 12:591-600 (2004)*
38. **Jonsson G, Bendahl PO, Sandberg T, et al.** Mapping of a novel ocular and cutaneous malignant melanoma susceptibility locus to chromosome 9q21.32. *J Natl Cancer Inst; 97:1377-82 (2005)*
39. **Sweetser DA, Peniket AJ, Haaland C, et al.** Delineation of the minimal commonly deleted segment and identification of candidate tumor-suppressor genes in del(9q) acute myeloid leukemia. *Genes Chromosomes Cancer; 44:279-91 (2005)*
40. **Rennstam K, Ahlstedt-Soini M, Baldetorp B, et al.** Patterns of chromosomal imbalances defines subgroups of breast cancer with distinct clinical features and prognosis. A study of 305 tumors by comparative genomic hybridization. *Cancer Res; 63:8861-8 (2003)*
41. **Chin K, DeVries S, Fridlyand J, et al.** Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell; 10:529-41 (2006)*
42. **Gronwald J, Jauch A, Cybulski C, et al.** Comparison of genomic abnormalities between BRCAX and sporadic breast cancers studied by comparative genomic hybridization. *Int J Cancer; 114:230-6 (2005)*
43. **Oldenburg RA, Kroeze-Jansema K, Meijers-Heijboer EJ, et al.** Characterization of familial non-BRCA1/2 breast tumors by loss of heterozygosity and immunophenotyping. *Clin Cancer Res; 12:1693-1700 (2006)*
44. **Wang Z, Lin M, Wei LJ, et al.** Loss of Heterozygosity and its correlation with expression profiles in subclasses of invasive breast cancers. *Cancer Res; 64:64-71 (2004)*
45. **Levanon D, Goldstein RE, Bernstein Y, et al.** Transcriptional repression by AML1 and LEF-1 is mediated by the TLE/Groucho corepressors. *Proc Natl Acad Sci U.S.A; 95:11590-5 (1998)*
46. **Inbal B, Cohen O, Polak-Charcon S, et al.** DAP kinase links the control of apoptosis to metastasis. *Nature; 390:180-4 (1997)*
47. **Ying J, Srivastava G, Hsieh WS, et al.** The stress-responsive gene GADD45G is a functional tumor suppressor, with its response to environmental stresses frequently disrupted epigenetically in multiple tumors. *Clin Cancer Res; 11:6442-9 (2005)*
48. **Liu Y, Dehni G, Purcell KJ, et al.** Epithelial expression and chromosomal location of human TLE genes: implications for notch signaling and neoplasia. *Genomics; 31: 58-64 (1996)*
49. **Raval A, Tanner SM, Byrd JC, et al.** Downregulation of death-associated protein kinase 1 (DAPK1) in chronic lymphocytic leukemia. *Cell; 129:879-90 (2007)*