Whole-genome Linkage Analysis and Sequence Analysis of Candidate Loci in Familial Breast Cancer

RAJESHWARI MARIKKANNU[1], CHRISTOS ARAVIDIS[1,2], JOHANNA RANTALA[1],
SIMONE PICELLI[1], TATJANA ADAMOVIC[1], MARKKU KEIHAS[1], TAO LIU[1],
VINAYKUMAR KONTHAM[1], DANIEL NILSSON[1] and ANNIKA LINDBLOM[1]

[1]Center for Molecular Medicine and Surgery, Karolinska University Hospital, Stockholm, Sweden;
[2]Department of Clinical Genetics, Rudbecklaboratoriet, Akademiska University Hospital, Uppsala, Sweden

Abstract. *Background: Known breast cancer-predisposing genes account for fewer than 25% of all familial breast cancer cases and further studies are required to find the remaining high- and moderate-risk genes. We set-out to couple linkage analysis using microsatellite marker data and sequence analysis of linked regions in 13 non-BRCA1/2 families in order to find novel susceptibility loci and high-penetrant genes. Materials and Methods: Genotyping with 540 fluorescently-labeled microsatellite markers located on the 23 chromosomes at 7.25 cM resolution was used for primary linkage analysis and an additional 40 markers were used for fine-mapping of loci with a logarithm of odds (LOD) or heterogeneity LOD (HLOD) score greater than one. Whole-exome sequencing data of 28 members from all 13 families were used for the bioinformatics sequence analysis on the linked regions of these families. Results: Linkage analysis identified three loci on chromosome 18q as a putative region of interest (overall LOD=1, HLOD=1.2). Sequencing analysis of the three linked regions on 18q and mutation prediction algorithms did reveal three probable damaging variants. Conclusion: Overall, our study identified three weakly linked loci on 18q and three probable damaging variants of interest in the 13 families with breast cancer.*

Breast cancer is the most common type of cancer among females and the most important risk factor is a family history of breast cancer. Epidemiological studies have shown that

Correspondence to: Annika Lindblom, L8:02; Center for Molecular Medicine and Surgery (CMM), Karolinska University Hospital, Stockholm-17176, Sweden. Tel: +46 851775248, Fax: +46 851773620, e-mail: annika.lindblom@ki.se

first-degree relatives of patients with breast cancer have a two-fold higher risk of breast cancer compared to the general population (1). Even though this could be, at least in part, due to shared environmental or lifestyle factors, twin studies found that up to 30% of all breast cancer cases may be due to genetic factors (2). Nonetheless, only approximately 5% of all breast cancer cases are attributed to the segregation of a germline mutation of a highly penetrant gene within the family (3). The two major high-risk genes are breast cancer 1 and 2 genes *BRCA1* and *BRCA2*, which together account for ~16% of the familial risk of breast cancer, while mutations in other high-risk genes such as: phosphatase and tensin homolog PTEN, serine/threonine kinase 11 *STK11*, cadherin 1 *CDH1* and tumor protein p53 *P53* or in the moderate-risk genes: ATM serine/threonine kinase *ATM*, checkpoint kinase 2 *CHEK2*, partner and localizer of BRCA2 *PALB2* and BRCA1 interacting protein C-terminal helicase 1 *BRIP1* explain a further 5% of familial cases (4). Thus, the great majority of families in whom the genetic association remains unexplained have been referred to as non-BRCA1/2 families.

Linkage analysis has been the method-of-choice for finding genes responsible for monogenic diseases and led to the identification of *BRCA1* and *BRCA2* (5, 6). However, many additional linkage studies have been performed in non-BRCA1/2 families using either short tandem repeat/microsatellite or single nucleotide polymorphism (SNP) markers without identifying any additional susceptibility genes (7-14). The lack of success could be attributed to extensive locus heterogeneity, where only a small proportion of the studied cases are linked to a particular locus. Greater statistical power could be achieved by analyzing subsets of families from more homogeneous populations (for example isolates, such as Finnish, Icelandic or the Ashkenazi Jewish population) in which the number of risk loci might be smaller. This approach was proved to be successful in a few cases, such as in the identification of transmembrane protease, serine 6

Table I. *Description of non-BRCA1/2 families.*

No.	Family ID	Total	Male	Female	Dead	Alive	Breast cancer srict criteria			Breast cancer loose criteria			Genotyped		WES
							Affected	Unaffected	Unknown	Affected	Unaffected	Unknown	Total	Affected	
1	862	27	11	16	17	10	6	10	11	6	10	11	10	3	3
2	1056	34	14	20	18	16	9	8	17	9	9	16	16	7	3
3	1141	12	7	5	6	6	3	5	4	3	5	4	6	1	1
4	2060	23	10	13	7	16	5	6	12	5	11	7	16	4	2
5	3028	14	5	9	4	10	3	2	9	3	3	8	10	3	3
6	6067	10	4	6	8	2	3	3	4	3	3	4	2	2	2
7	6094	8	3	5	6	2	3	2	3	3	2	3	2	2	2
8	6100	8	2	6	4	4	6	1	1	6	1	1	4	4	2
9	6101	10	4	6	6	4	2	3	5	2	3	5	4	2	1
10	6104	13	5	8	5	8	2	2	9	2	2	9	8	2	2
11	6106	11	2	9	2	9	3	2	6	3	2	6	9	3	2
12	6123	5	1	4	2	3	3	0	2	3	0	2	3	3	2
13	6127	10	4	6	4	6	2	2	6	2	2	6	6	2	3

WES: Whole exome sequencing.

TMPRSS6 and RAD50 homolog *RAD50,* as susceptibility genes in the Finnish population (15, 16).

In the past 4 years, new susceptibility alleles have been found through large global collaborative approaches, either *via* the analysis of individual genes or, recently, through genome-wide tag SNP experiments. So far, 27 different breast cancer susceptibility variants were detected (17-30) by GWAS, with an additional locus (caspase 8, apoptosis-related cysteine peptidase *CASP8*) being identified through a candidate-gene approach (31). In total, these loci account for approximately 30% of familial breast cancer. Recently Michailidou *et al.* identified SNPs at 41 new breast cancer susceptibility loci (32). Despite their high frequency, the fraction of the familial risk explained by all known risk alleles is only 30% in European populations (23), suggesting that other loci possibly remain to be identified.

Recently, an exome-sequencing study of families with multiple breast cancer-affected individuals identified two families with mutations in X-ray repair complementing defective repair in Chinese hamster cells 2 gene *XRCC2* and further association studies confirmed that rare mutations in this gene increase the risk of breast cancer (33). Increasingly, several studies are using next-generation sequencing as an explorative tool to screen the whole genome or exome in order to find susceptible variants.

Exome sequencing developed from next-generation sequencing (NGS), also known as massively parallel sequencing technologies, is revolutionizing our ability to characterize cancer at the genomic level by cataloging all mutations, copy number variations and somatic rearrangements at base-pair resolution. The exome covers ~1% of the human genome (human exome ~30 Mb) and is so far the most functionally relevant in phenotype variation. NGS combined with efficient DNA capture

enables for use of whole-exome sequencing (WES) studies to target exons and is emerging as an efficient tool for testing for association of rare coding variants with complex diseases. Short tandem repeats or highly polymorphic microsatellite markers are distributed widely and evenly in the genome, are relatively easy to score, and information about the markers is readily accessible through several databases such as the Marshfield Institute, deCODE *etc*. Therefore, in the present study, we set out to utilize linkage analysis using microsatellite data to identify candidate regions in the non-BRCA1/2 families and furthermore, to use WES data to find potential high-penetrant genes in the identified linked regions.

Materials and Methods

Patients. The present study is based on a cohort of 13 large familial non-BRCA1/2 hereditary breast cancer families. All families have Swedish origin and were recruited from the Genetic Counseling Unit at the Department of Clinical Genetics, Karolinska University Hospital, Stockholm. The study was undertaken in accordance with the Swedish legislation of ethical permission and according to the decision made by the Stockholm Regional Ethics Committee (97/205 and 00/291 and 08/125-31.2). Informed consent from all the families was obtained prior to study initiation. DNA was obtained from 96 family members (Table I). A total of 50 family members were diagnosed with breast cancer raging between 2-9 affected individuals in each family. All families screened negatively for mutations in the *BRCA1* and *BRCA2* genes. The number of genotyped family members in each family ranged from 2-16, and for the WES study, 1-3 members from each family were chosen from the available DNA so that they represented the most distantly related affected woman.

Genotyping. Genomic DNA was extracted from peripheral blood by standard procedures. A quality control of the samples and microsatellite genotyping were performed at deCODE Genetics (Reykjavik, Iceland). The deCODE markers were validated and

Table II. *Summary of putative candidate regions identified by genome-wide linkage analysis.*

Locus	Region	Start Position deCode cM	Overall PL LOD ($\alpha=1$)[a]		HLOD[a]				Overall NPL LOD[c]			
			BS	BL	BS	α[b]	BL	α	BS	*p*-Value	BL	*p*-Value
8q12.3-q13.3	D8S512-D8S279	78.059	1.0	0.9	1.0	0.95	0.9	0.80	0.5	0.33	0.4	0.36
11q13.2	D11S4087	73.806	-	-	1.0	0.70	-	-	0.7	0.20	-	-
18q21.1	D18S450	70.781	1.0	0.9	1.4	0.75	1.3	0.75	1.2	0.06	1.2	0.06
18q21.32-q22.3	D18S64-D18S469	84.270055	-	-	-	-	1.2	0.50	0.8	0.16	0.8	0.16
22q11.1-11.21	D22S420-D22S427	3.048	2.2	1.7	2.2	1.00	1.7	1.00	0.7	0.18	0.7	0.18

BS: Breast cancer strict criteria, BL: breast cancer loose criteria, LOD: logarithm of odds, HLOD: heterogeneity LOD, PL: parametric linkage analysis, NPL: non-parametric linkage analysis. [a]The highest overall PL LOD or HLOD or single family PL LOD in the specified microsatellite marker region is reported. [b]"α" Alpha is '*a priori*' proportion of linked pedigrees, *i.e.* if there is a trait locus in this region, alpha is the proportion of these pedigrees that contains an affected genotype at this trait locus. Thus, alpha=1.00 implies no locus heterogeneity. [c]Overall NPL LOD score corresponding to that specific overall PL LOD score and/or HLOD score is reported.

originally selected from the Marshfield genetic map (34). A total of 540 fluorescently-labeled microsatellite markers located on the 22 autosomes and the X chromosome were genotyped. The average distance between the markers was 7.25 cM and the genetic map used in the analyses was the one provided by deCODE. The overall successful genotyping rate was 94.3% for the samples.

For the fine-mapping analysis, an additional 5, 10, 19, 2 and 4 highly informative markers on chromosomes 8, 11, 18, 22 and X respectively were chosen from the Marshfield genetic map. (D8S1785, D8S1775, D8S1807, D8S1475, D8S525, D11S1392, D11S4102, D11S905, D11S1993, D11S1983, D11S1765, D11S4178, D11S4136, D11S4184, D11S4081, D18S1157, D18S1145, D18S460, D18S1119, D18S1144, D18S55, D18S1131, D18S58, D18S1121, D18S454, D18S65, D18S970, D18S455, D18S467, D18S1126, D18S363, D18S473, D18S470, D18S1110, D22S686, GATA198B05, DXS1003, DXS8032, DXS8031 and DXS8082). Primers for these markers were obtained from uniSTS, NCBI database, labeled with 6-FAM fluorescein modification at the 5' end of forward primers. Markers were amplified according to an in-house polymerase chain reaction PCR protocol (available on request). The amplified products were analyzed on either an Applied Biosystems 3130xL Genetic Analyzer or 3500xL Genetic Analyzer (Thermo Fisher Scientific, Waltham, MA, USA) after being primed with Genescan 400HD ROX size standard (Applied Biosystems, Foster City, Canada) and Hi-Di Formamide (Applied Biosystems, Warrington, UK). Subsequently, Genemapper software (version 3.7 or 4.1) (Thermo Fisher Scientific, Waltham, MA, USA) was used to analyze the peaks.

Linkage and fine-mapping analysis. Linkage analysis assessed two criteria: the first more stringent, breast cancer-strict and the next less stringent, breast cancer-loose. Both criteria listed all patients with breast cancer as affected, healthy spouses as unaffected, and all other cancer cases as unknown. In the breast cancer-loose criteria analysis, healthy women over 50 years old were coded as unaffected and those less than 50 years old were coded as unknown whereas in breast cancer-strict criteria all healthy women were set as unknown.

All genotyping data were checked for Mendelian inconsistencies between parents and offspring using the Pedcheck (version 1.00) software (35). Pedcheck executes three different error detection algorithms at three levels 0, 1 and 2 to report all the possible

ambiguous genotypes violating Mendelian rules. Next, Mega2 (versions 4.4.3 and 4.5.4) was used to prepare the files for the genetic analysis of the family-based study (36). Mega2 calculates the marker allelic frequencies from all the genotyped individuals from pedigree, locus, map and marker input data while generating the input files for subsequent Simwalk2 (version 2.91) analysis (37).

The available linkage analysis programs use the Markov chain Monte Carlo simulation algorithms for computing the location scores which are directly comparable with multipoint logarithm of odds (LOD) scores and are presented in log10 units. For the parametric linkage analysis, we assumed a dominant mode of inheritance and a disease allelic frequency of 0.0001. The penetrance for homozygous-normal, heterozygous, and homozygous-affected was set to 5%, 80% and 80%, respectively, and the phenocopies account for 5% of the observed affected individuals. Simwalk2 linkage analysis program was run for both breast cancer-loose and -strict criteria to calculate multipoint parametric linkage LOD (PL LOD) scores and locus heterogeneity LOD (HLOD) scores, as well as nonparametric linkage LOD (NPL LOD) scores for the 22 autosomes. Similar multipoint analysis on chromosome X was performed using Merlin (version 1.1.2) software (38).

WES and bioinformatics analysis. The WES of 28 individuals from these 13 families was performed at the Science for Life Laboratory, Stockholm as paired end reads to 100 bp on an Illumina HiSeq2000 instrument (Illumina, San Francisco, California, USA). Bioinformatics analysis of the raw exome sequencing data included, alignment of sequence reads to the reference human genome (with chromosome coordinates GRCh37/hg19) using alignment tools: BWA (Wellcomme Trust Sanger Institute, Cambridge, UK), PICARD (Broadinstitute) and SAMTools (Wellcome Trust Sanger Institute, Cambridge, Massachussets, US), applying GATK (39) base quality score recalibration, indel realignment, duplicate removal and SNP and INDEL discovery. Genotyping was performed across all 30 samples simultaneously using standard hard filtering parameters or variant quality score recalibration (40) and the output generated as an annotated variant call format (VCF) file. The VCF is a generic format for storing next-generation sequence polymorphism

Table III. *Summary of putative candidate regions remain after fine-mapping analysis.*

Start Position	Overall PL (α=1)[a]		HLOD[a]				Overall NPL[c]			
deCode cM	BS	BL	BS	α[b]	BL	α	BS	*p*-Value	BL	*p*-Value
47.287	-	-	0.9	0.60	1.1	0.75	0.9	0.13	0.9	0.13
66.99	1.0	-	1.2	0.80	0.9	0.70	0.9	0.12	0.9	0.12

BS: Breast cancer strict criteria, BL: breast cancer loose criteria, LOD: logarithm of odds, HLOD: heterogeneity LOD, PL: parametric linkage analysis, NPL: non-parametric linkage analysis. [a]The highest overall PL LOD or HLOD or single family PL LOD in the specified microsatellite marker region is reported. [b]"α" Alpha is 'a priori' proportion of linked pedigrees, *i.e.* if there is a trait locus in this region, alpha is the proportion of these pedigrees that contains an affected genotype at this trait locus. Thus, alpha=1.00 implies no locus heterogeneity. [c]Overall NPL LOD score corresponding to that specific overall PL LOD score and/or HLOD score is reported.

Table IV. *Sequence analysis summary of the putative candidate regions.*

Locus	Region	Physical map position	Initial hits		Finalized hits	
		Mb	Indels	Missense	Indels	Missense
18q11.2-q12.3	D18S1107-D18S65	22127970-40458634	6	74	0	4
18q12.3-q21.1	D18S65-D18S455	40458287-45922020	4	51	0	2
18q21.2-q22.2	D18S1110-D18S1131	48610129-66816246	17	105	2	3

data such as SNPs, insertions, deletions and other variants including structural variants together with annotated information. ANNOVAR (Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA) is a software tool to functionally annotate the genetic variants detected from diverse genomes present in the VCF file. Sequence analysis by ANNOVAR was preceded by generating a compatible file from VCF, followed by a gene based annotation using University of California, Santa Cruz (*UCSC*) genome browser (41). The huge number of hits from the bioinformatics analysis was reduced to a sizable number through sequential reduction steps for final manual validation. The deleterious and missense sequence analysis hits initially found, flagged with LowCoverage filter, VeryLowQual, LowQual, LowQD, HARD_TO_VALIDATE and SNPCluster, which were suggested to be either false-positives or potential artifacts were eliminated. In the next step, all the hits not shared by all the cases of at least one family among the LOD score-contributing linked families were excluded. Subsequently, for each position of a SNP/variant, the minor allelic frequency (MAF) was either taken from the 1000 genome project or when not available in the 1000 genome, taken from an available Swedish population of patients with colon cancer. This MAF has been used as one of the criteria to remove the hits having MAF more than 10% in the normal population. The calculated allelic frequency (CAF) used one and the same case from our 13 families each time. Finally, the variants at each position with a CAF of at least twofold more than the MAF were shortlisted as candidate variants. Furthermore, the variants were analyzed by the Alamut, an Interactive Biosoftware (version 2.3 on 2014-08-21).

Results

Linkage and fine-mapping analysis. Genome-wide linkage analysis was performed using breast cancer-strict and -loose criteria, and no statistically significant results (LOD >3) were obtained. We chose to use an LOD of one, as of interest for further studies. PL LOD, NPL LOD and HLOD scores of interest are reported in Table II, which shows either individual markers or set of markers with PL LOD or HLOD greater than one in any of the analyses and the corresponding NPL scores. Concerning the sets of markers, the highest score in this region is shown. The six consequently selected regions on five chromosomes were 8q12.3-q13.3, 11q13.2, 18q21.1, 18q21.32-q22.3, 22q11.1-q11.21 and Xq13.1. These loci were further fine mapped with an additional 40 markers and the results are displayed in Table III. After fine mapping, three regions with LOD scores greater than one were suggested on chromosome 18. Indeed, the whole of chromosome 18 below the marker D18S464 could not be ruled-out. The families contributing prominently to this overall LOD score (Family ID: 862, 2060, 6100) each have an individual family LOD score of at least 0.8 in these regions. None of the loci on chromosome 8, 11, 22 and X resulted in positive LOD scores after fine mapping. The primary NPL analysis

Table V. *Candidate variants identified by Bioinformatics Analysis of Exome Sequencing data.*

S. No	Chr	Locus	Mutation	Start Posn	End Posn	SNP ID	Mutation Type	Variation Class	Ref	Alt	Gene ID/ Variants	MAF 1000g	CAF AC/28	Family ID Families Sharing Variants
1	18	18q11.2-q12.3	missense	23866185	23866185	rs74947492	nonsyn	SNV	G	C	TAF4B	C= 0.05/116	C=0.07/2	862
2	18	18q11.2-q12.3	missense	28650748	28650748	rs151024019	nonsyn	SNV	A	C	DSC2	N/A	C=0.04/1	862
3	18	18q11.2-q12.3	missense	29111109	29111109	rs193922639	nonsyn	SNV	G	A	DSG2	N/A	A=0.04/1	862
4	18	18q11.2-q12.3	missense	30847180	30847180	rs58448816	nonsyn	SNV	C	T	C18orf34	T=0.08/174	T=0.11/3	862
5	18	18q12.3-q21.1	missense	43262408	43262408	rs41301139	nonsyn	SNV	G	A	SLC14A2	A=0.006/14	A=0.071/2	862
6	18	18q12.3-q21.1	missense	44181227	44181227	rs101163657	nonsyn	SNV	C	T	LOXHD1	T=0.06/122	T=0.107/3	862
7	18	18q21.2-q22.2	Indels	56203253	56203255	rs112880633	nonframeshift	deletion	AAG	-	ALPK2	N/A	Var=0.214/6	2060, 6100
8	18	18q21.2-q22.2	Indels	56204388	56204402	rs67925233	nonframeshift	deletion	CAGTTGAT GTGTCCT	-	ALPK2	N/A	Var=0.429/12	862, 2060, 6100
9	18	18q21.2-q22.2	missense	56202120	56202120	rs7234999	nonsyn	SNV	G	A	ALPK2	A=0.071/156	A=0.21/6	2060
10	18	18q21.2-q22.2	missense	56202683	56202683	rs33910491	nonsyn	SNV	T	C	ALPK2	C=0.06/134	C=0.21/6	2060
11	18	18q21.2-q22.2	missense	56279025	56279025	rs6566987	nonsyn	SNV	T	G	ALPK2	T=0.09/185	T=0.18/5	2060

SNPID: Single nucleotide polymorphism ID, SNV: single nucleotide variant, MAF: minor allele frequency, 1000g: 1000 genomes. MAF is reported in dbSNP in a default global population. For example, for rs74947492 the MAF is C=0.053/115. This means that for that SNP, minor allele is 'C' and has a frequency of 5.3% in the 1000Genome phase 1 population and that 'C' is observed 115 times in the sample population of 1088 people (or 2176 chromosomes).

identified two other regions on chromosome 18 and one each on chromosomes 19 and X with LOD scores greater than one. As none of these loci were supported by the overall PL LOD or HLOD analysis, these regions were not fine-mapped.

Sequencing and bioinformatics analysis. WES data from 28 patients with breast cancer from the 13 non-BRCA1/2 families were used for analyzing the three linked regions. The exome sequencing data were analyzed using ANNOVAR, followed by a sequential variant reduction steps developed in our laboratory (see Materials and Methods). The bioinformatics analysis of the three putative candidate marker regions on 18q from our current linkage and fine-mapping analysis (Table III) resulted in several genes with either deleterious or missense mutations. Three families contributing to the positive LOD score on chromosome 18 were taken into consideration for finding genes harboring mutations in the linked regions (Family ID: 860, 2060, and 6100). The variants obtained before and after our variant reduction steps are summarized in Table IV. Among the total 257 variants found on chromosome 18, two deleterious and nine missense mutations were validated further by Alamut Interactive biosoftware (Table V).

Discussion

To the best of our knowledge, the present study represents the first single-Center genome-wide linkage analysis in Swedish non-BRCA1/2 families combined with sequence analysis published to date. We set out to test the hypothesis whether further susceptibility loci exist that confer moderate or high risk to breast cancer.

The current study was carried out in 13 extended non-BRCA1/2 breast cancer families. The advantage of our study is that large non-BRCA1/2 families were used. However, as breast cancer is a heterogeneous disease, using only 13 families was not likely to detect all involved breast cancer-predisposing genes, but we hypothesized that the larger size of the families would make up for the small number of families, in particular since Sweden can be considered a fairly homogenous population. The pedigrees from the families included in the current study indicated that the involved risk alleles were likely to be dominant. Such unusual high aggregation was not likely to be explained by shared environmental or low-penetrant variants alone, but rather by genes with a main effect, where the aforementioned factors could act as modifiers. Hence our primary analysis assumed a dominant mode of inheritance. In order to avoid overlooking a possible linkage signal due to misspecification of the model, we also used an allele-sharing approach (model-free or NPL analysis). However, this approach did not identify any linked regions.

The study showed no statistically significant results of LOD greater than 'three, but did find several putative candidate regions with an overall PL LOD or HLOD of greater than one on chromosome 18q. HLOD scores also supported the PL LOD scores but some loci were identified only by the HLOD score analysis. None of these loci overlapped with the regions previously identified in linkage studies or reported among the 25 SNP loci from the GWAS (17-30); since only a very limited part of all genetic contribution is currently known, this is not surprising, especially since our study used a small number of families with breast cancer, where chance alone can result in suggested risk loci.

As several sets of markers throughout chromosome 18 had LODs or HLODs greater than one without clear boundaries, it makes the whole chromosome 18 below marker D18S464 of interest. The bioinformatics analysis of exome sequencing data on the 18q candidate loci identified two deleterious and nine missense variants. Among the shortlisted variants, except desmocollin 2 *DSC2*, desmoglein 2 *DSG2*, solute carrier family 14 (urea transporter), member 2 *SLC14A2* and alpha-kinase 2 *ALPK2* (rs7234999, rs33910491), the remaining were shown by Alamut to be neutral/benign and not evolutionarily conserved. *DSC2* variant is reported as likely pathogenic in The Human Gene Mutation Database *HGMD* but only in arrhythmogenic right ventricular dysplasia/cardiomyopathy (CM0910201). Its clinical significance in breast cancer has not yet been studied. DSG2 was shown to be clinically pathogenic but conversely predicted to be benign by mutation- and conservation-prediction algorithms. *SLC14A2* was predicted to be probably damaging (PolyPhen2) and conserved (Grantham and GERP), but no clinical significance was found. Moreover, *ALPK2* (rs7234999 and rs33910491) were predicted to be possibly damaging and conserved, with no clinical significance being found. These breast cancer families will be further studied since we have whole-exome data for all of them. The present exome sequence analysis study was a follow-up of the linkage analysis, and therefore it focused only on the linked regions of chromosome18q and looked for the presence of potential risk variants in our candidate region.

Conclusion

Overall, the present study found three weakly-linked loci in the chromosome 18q region. Our findings are consistent with results of several other genetic studies, in that getting overall LOD score greater than 3 is not possible with such a small number of families. The three predicted to be probably/possibly damaging variants are of interest, but further confirmatory studies are needed to prove their clinical significance in causing breast cancer. Our data provide further support for the hypothesis that multiple loci on chromosomes harboring common low-penetrance, or rare

moderate -risk genetic variants are more likely responsible for familial breast cancer.

Conflicts of Interests

Financial Support

Acknowledgements

References

1 Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52,705 women with breast cancer and 108,411 women without breast cancer. Lancet *350*: 1047-1059, 1997.

2 Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A and Hemminki K: Environmental and heritable factors in the causation of cancer–analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med *343*: 78-85, 2000.

3 Nathanson KL, Wooster R, Weber BL and Nathanson KN: Breast cancer genetics: what we know and what we need. Nat Med *7*: 552-556, 2001.

4 Stratton MR and Rahman N: The emerging landscape of breast cancer susceptibility. Nat Genet *40*: 17-22, 2008.

5 Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, Liu Q, Cochran C, Bennett LM and Ding W: A strong candidate for the breast and ovarian cancer susceptibility gene *BRCA1*. Science *266*: 66-71, 1994.

6 Wooster R, Neuhausen SL, Mangion J, Quirk Y, Ford D, Collins N, Nguyen K, Seal S, Tran T and Averill D: Localization of a breast cancer susceptibility gene, *BRCA2*, to chromosome 13q12-13. Science *265*: 2088-2090, 1994.

7 Bergman A, Karlsson P, Berggren J, Martinsson T, Björck K, Nilsson S, Wahlström J, Wallgren A and Nordling M: Genome-wide linkage scan for breast cancer susceptibility loci in Swedish hereditary non-BRCA1/2 families: suggestive linkage to 10q23.32-q25.3. Genes Chromosomes Cancer *46*: 302-309, 2007.

8 Gonzalez-Neira A, Rosa-Rosa JM, Osorio A, Gonzalez E, Southey M, Sinilnikova O, Lynch H, Oldenburg RA, van Asperen CJ, Hoogerbrugge N, Pita G, Devilee P, Goldgar D and Benitez J: Genomewide high-density SNP linkage analysis of non-BRCA1/2 breast cancer families identifies various candidate regions and has greater power than microsatellite studies. BMC Genomics *8*: 299, 2007.

9 Hartikainen JM, Tuhkanen H, Kataja V, Dunning AM, Antoniou A, Smith P, Arffman A, Pirskanen M, Easton DF, Eskelinen M, Uusitupa M, Kosma V-M and Mannermaa A: 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.

10 Huusko P, Juo S-HH, Gillanders E, Sarantaus L, Kainu T, Vahteristo P, Allinen M, Jones M, Rapakko K, Eerola H, Markey C, Vehmanen P, Gildea D, Freas-Lutz D, Blomqvist C, Leisti J, Blanco G, Puistola U, Trent J, Bailey-Wilson J, Winqvist R, Nevanlinna H and Kallioniemi O-P: Genome-wide scanning for linkage in Finnish breast cancer families. Eur J Hum Genet *12*: 98-104, 2004.

11 Oldenburg RA, Kroeze-Jansema KHG, Houwing-Duistermaat JJ, Bayley J-P, Dambrot C, van Asperen CJ, van den Ouweland AMW, Bakker B, van Beers EH, Nederlof PM, Vasen H, Hoogerbrugge N, Cornelisse CJ, Meijers-Heijboer H and Devilee P: Genome-wide linkage scan in Dutch hereditary non-BRCA1/2 breast cancer families identifies 9q21-22 as a putative breast cancer susceptibility locus. Genes Chromosomes Cancer *47*: 947-956, 2008.

12 Rosa-Rosa JM, Pita G, Urioste M, Llort G, Brunet J, Lázaro C, Blanco I, Ramón y Cajal T, Díez O, de la Hoya M, Caldés T, Tejada M-I, González-Neira A and Benítez J: Genome-wide linkage scan reveals three putative breast-cancer-susceptibility loci. Am J Hum Genet *84*: 115-122, 2009.

13 Smith P, McGuffog L, Easton DF, Mann GJ, Pupo GM, Newman B, Chenevix-Trench G, Szabo C, Southey M, Renard H, Odefrey F, Lynch H, Stoppa-Lyonnet D, Couch F, Hopper JL, Giles GG, McCredie MRE, Buys S, Andrulis I, Senie R, Goldgar DE, Oldenburg R, Kroeze-Jansema K, Kraan J, Meijers-Heijboer H, Klijn JGM, van Asperen C, van Leeuwen I, Vasen HFA, Cornelisse CJ, Devilee P, Baskcomb L, Seal S, Barfoot R, Mangion J, Hall A, Edkins S, Rapley E, Wooster R, Chang-Claude J, Eccles D, Evans DG, Futreal PA, Nathanson KL, Weber BL, Rahman N and Stratton MR: A genome-wide linkage search for breast cancer susceptibility genes. Genes Chromosomes Cancer *45*: 646-655, 2006.

14 Thompson D, Szabo CI, Mangion J, Oldenburg RA, Odefrey F, Seal S, Barfoot R, Kroeze-Jansema K, Teare D, Rahman N, Renard H, Mann G, Hopper JL, Buys SS, Andrulis IL, Senie R, Daly MB, West D, Ostrander EA, Offit K, Peretz T, Osorio A, Benitez J, Nathanson KL, Sinilnikova OM, Olàh E, Bignon Y-J, Ruiz P, Badzioch MD, Vasen HFA, Futreal AP, Phelan CM, Narod SA, Lynch HT, Ponder BAJ, Eeles RA, Meijers-Heijboer H, Stoppa-Lyonnet D, Couch FJ, Eccles DM, Evans DG, Chang-Claude J, Lenoir G, Weber BL, Devilee P, Easton DF, Goldgar DE and Stratton MR: Evaluation of linkage of breast cancer to the putative BRCA3 locus on chromosome 13q21 in 128 multiple case families from the Breast Cancer Linkage Consortium. Proc Natl Acad Sci USA *99*: 827-831, 2002.

15 Hartikainen JM, Tuhkanen H, Kataja V, Eskelinen M, Uusitupa M, Kosma V-M and Mannermaa A: Refinement of the 22q12-q13 breast cancer--associated region: evidence of *TMPRSS6* as a candidate gene in an eastern Finnish population. Clin Cancer Res *12*: 1454-1462, 2006.

16 Heikkinen K, Rapakko K, Karppinen S-M, Erkko H, Knuutila S, Lundán T, Mannermaa A, Børresen-Dale A-L, Borg A, Barkardottir RB, Petrini J and Winqvist R: *RAD50* and *NBS1* are breast cancer susceptibility genes associated with genomic instability. Carcinogenesis *27*: 1593-1599, 2006.

17 Easton DF, Pooley KA, Dunning AM, Pharoah PDP, Thompson D, Ballinger DG, Struewing JP, Morrison J, Field H, Luben R, Wareham N, Ahmed S, Healey CS, Bowman R, Meyer KB, Haiman CA, Kolonel LK, Henderson BE, Le Marchand L, Brennan P, Sangrajrang S, Gaborieau V, Odefrey F, Shen C-Y, Wu P-E, Wang H-C, Eccles D, Evans DG, Peto J, Fletcher O, Johnson N, Seal S, Stratton MR, Rahman N, Chenevix-Trench G, Bojesen SE, Nordestgaard BG, Axelsson CK, Garcia-Closas M, Brinton L, Chanock S, Lissowska J, Peplonska B, Nevanlinna H, Fagerholm R, Eerola H, Kang D, Yoo K-Y, Noh D-Y, Ahn S-H, Hunter DJ, Hankinson SE, Cox DG, Hall P, Wedren S, Liu J, Low Y-L, Bogdanova N, Schürmann P, Dörk T, Tollenaar RAEM, Jacobi CE, Devilee P, Klijn JGM, Sigurdson AJ, Doody MM, Alexander BH, Zhang J, Cox A, Brock IW, MacPherson G, Reed MWR, Couch FJ, Goode EL, Olson JE, Meijers-Heijboer H, van den Ouweland A, Uitterlinden A, Rivadeneira F, Milne RL, Ribas G, Gonzalez-Neira A, Benitez J, Hopper JL, McCredie M, Southey M, Giles GG, Schroen C, Justenhoven C, Brauch H, Hamann U, Ko Y-D, Spurdle AB, Beesley J, Chen X, Mannermaa A, Kosma V-M, Kataja V, Hartikainen J, Day NE, Cox DR and Ponder BAJ: Genome-wide association study identifies novel breast cancer susceptibility loci. Nature *447*: 1087-1093, 2007.

18 Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, Wacholder S, Wang Z, Welch R, Hutchinson A, Wang J, Yu K, Chatterjee N, Orr N, Willett WC, Colditz GA, Ziegler RG, Berg CD, Buys SS, McCarty CA, Feigelson HS, Calle EE, Thun MJ, Hayes RB, Tucker M, Gerhard DS, Fraumeni JF Jr., Hoover RN, Thomas G and Chanock SJ: A genome-wide association study identifies alleles in *FGFR2*- associated with risk of sporadic postmenopausal breast cancer. Nat Genet *39*: 870-874, 2007.

19 Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, Gudjonsson SA, Masson G, Jakobsdottir M, Thorlacius S, Helgason A, Aben KK, Strobbe LJ, Albers-Akkers MT, Swinkels DW, Henderson BE, Kolonel LN, Le Marchand L, Millastre E, Andres R, Godino J, Garcia-Prats MD, Polo E, Tres A, Mouy M, Saemundsdottir J, Backman VM, Gudmundsson L, Kristjansson K, Bergthorsson JT, Kostic J, Frigge ML, Geller F, Gudbjartsson D, Sigurdsson H, Jonsdottir T, Hrafnkelsson J, Johannsson J, Sveinsson T, Myrdal G, Grimsson HN, Jonsson T, von Holst S, Werelius B, Margolin S, Lindblom A, Mayordomo JI, Haiman CA, Kiemeney LA, Johannsson OT, Gulcher JR, Thorsteinsdottir U, Kong A and Stefansson K: Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet *39*: 865-869, 2007.

20 Stacey SN, Manolescu A, Sulem P, Thorlacius S, Gudjonsson SA, Jonsson GF, Jakobsdottir M, Bergthorsson JT, Gudmundsson J, Aben KK, Strobbe LJ, Swinkels DW, van Engelenburg KCA, Henderson BE, Kolonel LN, Le Marchand L, Millastre E, Andres R, Saez B, Lambea J, Godino J, Polo E, Tres A, Picelli S, Rantala J, Margolin S, Jonsson T, Sigurdsson H, Jonsdottir T, Hrafnkelsson J, Johannsson J, Sveinsson T, Myrdal G, Grimsson HN, Sveinsdottir SG, Alexiusdottir K, Saemundsdottir J, Sigurdsson A, Kostic J, Gudmundsson L, Kristjansson K, Masson G, Fackenthal JD, Adebamowo C, Ogundiran T, Olopade OI, Haiman CA, Lindblom A, Mayordomo JI, Kiemeney LA, Gulcher JR, Rafnar T, Thorsteinsdottir U, Johannsson OT, Kong A and Stefansson K: Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet *40*: 703-706, 2008.

21 Gold B, Kirchhoff T, Stefanov S, Lautenberger J, Viale A, Garber J, Friedman E, Narod S, Olshen AB, Gregersen P, Kosarin K, Olsh A, Bergeron J, Ellis NA, Klein RJ, Clark AG, Norton L, Dean M, Boyd J and Offit K: Genome-wide association study provides evidence for a breast cancer risk locus at 6q22.33. Proc Natl Acad Sci USA *105*: 4340-4345, 2008.

22 Thomas G, Jacobs KB, Kraft P, Yeager M, Wacholder S, Cox DG, Hankinson SE, Hutchinson A, Wang Z, Yu K, Chatterjee N, Garcia-Closas M, Gonzalez-Bosquet J, Prokunina-Olsson L, Orr N, Willett WC, Colditz GA, Ziegler RG, Berg CD, Buys SS, McCarty CA, Feigelson HS, Calle EE, Thun MJ, Diver R, Prentice R, Jackson R, Kooperberg C, Chlebowski R, Lissowska J, Peplonska B, Brinton LA, Sigurdson A, Doody M, Bhatti P, Alexander BH, Buring J, Lee I-M, Vatten LJ, Hveem K, Kumle M, Hayes RB, Tucker M, Gerhard DS, Fraumeni JF Jr., Hoover RN, Chanock SJ and Hunter DJ: A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (*RAD51L1*). Nat Genet *41*: 579-584, 2009.

23 Ahmed S, Thomas G, Ghoussaini M, Healey CS, Humphreys MK, Platte R, Morrison J, Maranian M, Pooley KA, Luben R, Eccles D, Evans DG, Fletcher O, Johnson N, dos Santos Silva I, Peto J, Stratton MR, Rahman N, Jacobs K, Prentice R, Anderson GL, Rajkovic A, Curb JD, Ziegler RG, Berg CD, Buys SS, McCarty CA, Feigelson HS, Calle EE, Thun MJ, Diver WR, Bojesen S, Nordestgaard BG, Flyger H, Dörk T, Schürmann P, Hillemanns P, Karstens JH, Bogdanova NV, Antonenkova NN, Zalutsky IV, Bermisheva M, Fedorova S, Khusnutdinova E, SEARCH, Kang D, Yoo K-Y, Noh DY, Ahn S-H, Devilee P, van Asperen CJ, Tollenaar R a. EM, Seynaeve C, Garcia-Closas M, Lissowska J, Brinton L, Peplonska B, Nevanlinna H, Heikkinen T, Aittomäki K, Blomqvist C, Hopper JL, Southey MC, Smith L, Spurdle AB, Schmidt MK, Broeks A, van Hien RR, Cornelissen S, Milne RL, Ribas G, González-Neira A, Benitez J, Schmutzler RK, Burwinkel B, Bartram CR, Meindl A, Brauch H, Justenhoven C, Hamann U, GENICA Consortium, Chang-Claude J, Hein R, Wang-Gohrke S, Lindblom A, Margolin S, Mannermaa A, Kosma V-M, Kataja V, Olson JE, Wang X, Fredericksen Z, Giles GG, Severi G, Baglietto L, English DR, Hankinson SE, Cox DG, Kraft P, Vatten LJ, Hveem K, Kumle M, Sigurdson A, Doody M, Bhatti P, Alexander BH, Hooning MJ, van den Ouweland AMW, Oldenburg RA, Schutte M, Hall P, Czene K, Liu J, Li Y, Cox A, Elliott G, Brock I, Reed MWR, Shen C-Y, Yu J-C, Hsu G-C, Chen S-T, Anton-Culver H, Ziogas A, Andrulis IL, Knight JA, kConFab, Australian Ovarian Cancer Study Group, Beesley J, Goode EL, Couch F, Chenevix-Trench G, Hoover RN, Ponder BAJ, Hunter DJ, Pharoah PDP, Dunning AM, Chanock SJ and Easton DF: Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. Nat Genet *41*: 585-590, 2009.

24 Zheng W, Long J, Gao Y-T, Li C, Zheng Y, Xiang Y-B, Wen W, Levy S, Deming SL, Haines JL, Gu K, Fair AM, Cai Q, Lu W and Shu X-O: Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. Nat Genet *41*: 324-328, 2009.

25 Turnbull C, Ahmed S, Morrison J, Pernet D, Renwick A, Maranian M, Seal S, Ghoussaini M, Hines S, Healey CS, Hughes D, Warren-Perry M, Tapper W, Eccles D, Evans DG, Hooning M, Schutte M, van den Ouweland A, Houlston R, Ross G, Langford C, Pharoah PDP, Stratton MR, Dunning AM, Rahman N and Easton DF: Genome-wide association study identifies five new breast cancer susceptibility loci. Nat Genet *42*: 504-507, 2010.

26 Antoniou AC, Wang X, Fredericksen ZS, McGuffog L, Tarrell R, Sinilnikova OM, Healey S, Morrison J, Kartsonaki C, Lesnick T, Ghoussaini M, Barrowdale D, EMBRACE, Peock S, Cook M, Oliver C, Frost D, Eccles D, Evans DG, Eeles R, Izatt L, Chu C, Douglas F, Paterson J, Stoppa-Lyonnet D, Houdayer C, Mazoyer S, Giraud S, Lasset C, Remenieras A, Caron O, Hardouin A, Berthet P, GEMO Study Collaborators, Hogervorst FBL, Rookus MA, Jager A, van den Ouweland A, Hoogerbrugge N, van der Luijt RB, Meijers-Heijboer H, Gómez García EB, HEBON, Devilee P, Vreeswijk MPG, Lubinski J, Jakubowska A, Gronwald J, Huzarski T, Byrski T, Górski B, Cybulski C, Spurdle AB, Holland H, kConFab, Goldgar DE, John EM, Hopper JL, Southey M, Buys SS, Daly MB, Terry M-B, Schmutzler RK, Wappenschmidt B, Engel C, Meindl A, Preisler-Adams S, Arnold N, Niederacher D, Sutter C, Domchek SM, Nathanson KL, Rebbeck T, Blum JL, Piedmonte M, Rodriguez GC, Wakeley K, Boggess JF, Basil J, Blank SV, Friedman E, Kaufman B, Laitman Y, Milgrom R, Andrulis IL, Glendon G, Ozcelik H, Kirchhoff T, Vijai J, Gaudet MM, Altshuler D, Guiducci C, SWE-BRCA, Loman N, Harbst K, Rantala J, Ehrencrona H, Gerdes A-M, Thomassen M, Sunde L, Peterlongo P, Manoukian S, Bonanni B, Viel A, Radice P, Caldes T, de la Hoya M, Singer CF, Fink-Retter A, Greene MH, Mai PL, Loud JT, Guidugli L, Lindor NM, Hansen TVO, Nielsen FC, Blanco I, Lazaro C, Garber J, Ramus SJ, Gayther SA, Phelan C, Narod S, Szabo CI, MOD SQUAD, Benitez J, Osorio A, Nevanlinna H, Heikkinen T, Caligo MA, Beattie MS, Hamann U, Godwin AK, Montagna M, Casella C, Neuhausen SL, Karlan BY, Tung N, Toland AE, Weitzel J, Olopade O, Simard J, Soucy P, Rubinstein WS, Arason A, Rennert G, Martin NG, Montgomery GW, Chang-Claude J, Flesch-Janys D, Brauch H, GENICA, Severi G, Baglietto L, Cox A, Cross SS, Miron P, Gerty SM, Tapper W, Yannoukakos D, Fountzilas G, Fasching PA, Beckmann MW, Dos Santos Silva I, Peto J, Lambrechts D, Paridaens R, Rüdiger T, Försti A, Winqvist R, Pylkäs K, Diasio RB, Lee AM, Eckel-Passow J, Vachon C, Blows F, Driver K, Dunning A, Pharoah PPD, Offit K, Pankratz VS, Hakonarson H, Chenevix-Trench G, Easton DF and Couch FJ: A locus on 19p13 modifies risk of breast cancer in *BRCA1* mutation carriers and is associated with hormone receptor-negative breast cancer in the general population. Nat Genet *42*: 885-892, 2010.

27 Fletcher O, Johnson N, Orr N, Hosking FJ, Gibson LJ, Walker K, Zelenika D, Gut I, Heath S, Palles C, Coupland B, Broderick P, Schoemaker M, Jones M, Williamson J, Chilcott-Burns S, Tomczyk K, Simpson G, Jacobs KB, Chanock SJ, Hunter DJ, Tomlinson IP, Swerdlow A, Ashworth A, Ross G, dos Santos Silva I, Lathrop M, Houlston RS and Peto J: Novel breast cancer susceptibility locus at 9q31.2: results of a genome-wide association study. J Natl Cancer Inst *103*: 425-435, 2011.

28 Haiman CA, Chen GK, Vachon CM, Canzian F, Dunning A, Millikan RC, Wang X, Ademuyiwa F, Ahmed S, Ambrosone CB, Baglietto L, Balleine R, Bandera EV, Beckmann MW, Berg CD, Bernstein L, Blomqvist C, Blot WJ, Brauch H, Buring JE, Carey LA, Carpenter JE, Chang-Claude J, Chanock SJ, Chasman DI, Clarke CL, Cox A, Cross SS, Deming SL, Diasio RB, Dimopoulos AM, Driver WR, Dunnebier T, Durcan L, Eccles D, Edlund CK, Ekici AB, Fasching PA, Feigelson HS, Flesch-Janys D, Fostira F, Forsti A, Fountzilas G, Gerty SM, Giles GG, Godwin AK, Goodfellow P, Graham N, Greco D, Hamann U, Hankinson SE, Hartmann A, Hein R, Heinz J, Holbrook A, Hoover RN, Hu JJ, Hunter DJ, Ingles SA, Irwanto A, Ivanovich J, John EM, Johnson N, Jukkola-Vuorinen A, Kaaks R, Ko Y-D, Kolonel LN, Konstantopoulou I, Kosma V-M, Kulkarni S, Lambrechts D, Lee AM, Le Marchand L, Lesnick T, Liu J, Lindstrom S, Mannermaa A, Margolin S, Martin NG, Miron P, Montgomery GW, Nevanlinna H, Nickels S, Nyante S, Olswold C, Palmer J, Pathak H, Pectasides D, Perou CM, Peto J, Pharoah PDP, Pooler LC, Press MF, Pylkas K, Rebbeck TR, Rodriguez-Gil JL, Rosenberg L, Ross E, Rudiger T, Silva I dos S, Sawyer E, Schmidt MK, Schulz-Wendtland R, Schumacher F, Severi G, Sheng X, Signorello LB, Sinn H-P, Stevens KN, Southey MC, Tapper WJ, Tomlinson I, Hogervorst FBL, Wauters E, Weaver J, Wildiers H, Winqvist R, Berg DVD, Wan P, Xia LY, Yannoukakos D, Zheng W, Ziegler RG, Siddiq A, Slager SL, Stram DO, Easton D, Kraft P, Henderson BE and Couch FJ: A common variant at the *TERT-CLPTM1L* locus is associated with estrogen receptor-negative breast cancer. Nat Genet *43*: 1210-1214, 2011.

29 Ghoussaini M, Fletcher O, Michailidou K, Turnbull C, Schmidt MK, Dicks E, Dennis J, Wang Q, Humphreys MK, Luccarini C, Baynes C, Conroy D, Maranian M, Ahmed S, Driver K, Johnson N, Orr N, dos Santos Silva I, Waisfisz Q, Meijers-Heijboer H, Uitterlinden AG, Rivadeneira F, Hall P, Czene K, Irwanto A, Liu J, Nevanlinna H, Aittomaki K, Blomqvist C, Meindl A, Schmutzler RK, Muller-Myhsok B, Lichtner P, Chang-Claude J, Hein R, Nickels S, Flesch-Janys D, Tsimiklis H, Makalic E, Schmidt D, Bui M, Hopper JL, Apicella C, Park DJ, Southey M, Hunter DJ, Chanock SJ, Broeks A, Verhoef S, Hogervorst FBL, Fasching PA, Lux MP, Beckmann MW, Ekici AB, Sawyer E, Tomlinson I, Kerin M, Marme F, Schneeweiss A, Sohn C, Burwinkel B, Guenel P, Truong T, Cordina-Duverger E, Menegaux F, Bojesen SE, Nordestgaard BG, Nielsen SF, Flyger H, Milne RL, Alonso MR, Gonzalez-Neira A, Benitez J, Anton-Culver H, Ziogas A, Bernstein L, Dur CC, Brenner H, Muller H, Arndt V, Stegmaier C, Justenhoven C, Brauch H, Bruning T, Wang-Gohrke S, Eilber U, Dork T, Schurmann P, Bremer M, Hillemanns P, Bogdanova NV, Antonenkova NN, Rogov YI, Karstens JH, Bermisheva M, Prokofieva D, Khusnutdinova E, Lindblom A, Margolin S, Mannermaa A, Kataja V, Kosma V-M, Hartikainen JM, Lambrechts D, Yesilyurt BT, Floris G, Leunen K, Manoukian S, Bonanni B, Fortuzzi S, Peterlongo P, Couch FJ, Wang X, Stevens K, Lee A, Giles GG, Baglietto L, Severi G, McLean C, Alnaes GG, Kristensen V, Borrensen-Dale A-L, John EM, Miron A, Winqvist R, Pylkas K, Jukkola-Vuorinen A, Kauppila S, Andrulis IL, Glendon G, Mulligan AM, Devilee P, van Asperen CJ, Tollenaar RAEM, Seynaeve C, Figueroa JD, Garcia-Closas M, Brinton L, Lissowska J, Hooning MJ, Hollestelle A, Oldenburg RA, van den Ouweland AMW, Cox A, Reed MWR, Shah M, Jakubowska A, Lubinski J, Jaworska K, Durda K, Jones M, Schoemaker M, Ashworth A, Swerdlow A, Beesley J, Chen X, Muir KR, Lophatananon A, Rattanamongkongul S, Chaiwerawattana A, Kang D, Yoo K-Y, Noh D-Y, Shen C-Y, Yu J-C, Wu P-E, Hsiung C-N, Perkins A, Swann R, Velentzis L, Eccles DM, Tapper WJ, Gerty SM, Graham NJ, Ponder BAJ, Chenevix-Trench G, Pharoah PDP, Lathrop M, Dunning AM, Rahman N, Peto J and Easton DF: Genome-wide association analysis identifies three new breast cancer susceptibility loci. Nat Genet *44*: 312-318, 2012.

30 Siddiq A, Couch FJ, Chen GK, Lindström S, Eccles D, Millikan RC, Michailidou K, Stram DO, Beckmann L, Rhie SK, Ambrosone CB, Aittomäki K, Amiano P, Apicella C, Australian Breast Cancer Tissue Bank Investigators, Baglietto L, Bandera EV, Beckmann MW, Berg CD, Bernstein L, Blomqvist C, Brauch H, Brinton L, Bui QM, Buring JE, Buys SS, Campa D, Carpenter JE, Chasman DI, Chang-Claude J, Chen C, Clavel-Chapelon F, Cox A, Cross SS, Czene K, Deming SL, Diasio RB, Diver WR, Dunning AM, Durcan L, Ekici AB, Fasching PA, Familial Breast Cancer Study, Feigelson HS, Fejerman L, Figueroa JD, Fletcher O, Flesch-Janys D, Gaudet MM, GENICA Consortium, Gerty SM, Rodriguez-Gil JL, Giles GG, van Gils CH, Godwin AK, Graham N, Greco D, Hall P, Hankinson SE, Hartmann A, Hein R, Heinz J, Hoover RN, Hopper JL, Hu JJ, Huntsman S, Ingles SA, Irwanto A, Isaacs C, Jacobs KB, John EM, Justenhoven C, Kaaks R, Kolonel LN, Coetzee GA, Lathrop M, Le Marchand L, Lee AM, Lee I-M, Lesnick T, Lichtner P, Liu J, Lund E, Makalic E, Martin NG, McLean CA, Meijers-Heijboer H, Meindl A, Miron P, Monroe KR, Montgomery GW, Müller-Myhsok B, Nickels S, Nyante SJ, Olswold C, Overvad K, Palli D, Park DJ, Palmer JR, Pathak H, Peto J, Pharoah P, Rahman N, Rivadeneira F, Schmidt DF, Schmutzler RK, Slager S, Southey MC, Stevens KN, Sinn H-P, Press MF, Ross E, Riboli E, Ridker PM, Schumacher FR, Severi G, Dos Santos Silva I, Stone J, Sund M, Tapper WJ, Thun MJ, Travis RC, Turnbull C, Uitterlinden AG, Waisfisz Q, Wang X, Wang Z, Weaver J, Schulz-Wendtland R, Wilkens LR, Van Den Berg D, Zheng W, Ziegler RG, Ziv E, Nevanlinna H, Easton DF, Hunter DJ, Henderson BE, Chanock SJ, Garcia-Closas M, Kraft P, Haiman CA and Vachon CM: A meta-analysis of genome-wide association studies of breast cancer identifies two novel susceptibility loci at 6q14 and 20q11. Hum Mol Genet *21*: 5373-5384, 2012.

31 Cox A, Dunning AM, Garcia-Closas M, Balasubramanian S, Reed MWR, Pooley KA, Scollen S, Baynes C, Ponder BAJ, Chanock S, Lissowska J, Brinton L, Peplonska B, Southey MC, Hopper JL, McCredie MRE, Giles GG, Fletcher O, Johnson N, dos Santos Silva I, Gibson L, Bojesen SE, Nordestgaard BG, Axelsson CK, Torres D, Hamann U, Justenhoven C, Brauch H, Chang-Claude J, Kropp S, Risch A, Wang-Gohrke S, Schurmann P, Bogdanova N, Dork T, Fagerholm R, Aaltonen K, Blomqvist C, Nevanlinna H, Seal S, Renwick A, Stratton MR, Rahman N, Sangrajrang S, Hughes D, Odefrey F, Brennan P, Spurdle AB, Chenevix-Trench G, Beesley J, Mannermaa A, Hartikainen J, Kataja V, Kosma V-M, Couch FJ, Olson JE, Goode EL, Broeks A, Schmidt MK, Hogervorst FBL, Veer LJV, Kang D, Yoo K-Y, Noh D-Y, Ahn S-H, Wedren S, Hall P, Low Y-L, Liu J, Milne RL, Ribas G, Gonzalez-Neira A, Benitez J, Sigurdson AJ, Stredrick DL, Alexander BH, Struewing JP, Pharoah PDP and Easton DF: A common coding variant in *CASP8* is associated with breast cancer risk. Nat Genet *39*: 352-358, 2007.

32 Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, Schmidt MK, Chang-Claude J, Bojesen SE, Bolla MK, Wang Q, Dicks E, Lee A, Turnbull C, Rahman N, Breast and Ovarian Cancer Susceptibility Collaboration, Fletcher O, Peto J, Gibson L, Dos Santos Silva I, Nevanlinna H, Muranen TA, Aittomäki K, Blomqvist C, Czene K, Irwanto A, Liu J, Waisfisz Q, Meijers-Heijboer H, Adank M, Hereditary Breast and Ovarian Cancer Research Group Netherlands (HEBON), van der Luijt RB, Hein R, Dahmen N, Beckman L, Meindl A, Schmutzler RK, Müller-Myhsok B, Lichtner P, Hopper JL, Southey MC, Makalic E, Schmidt DF, Uitterlinden AG, Hofman A, Hunter DJ, Chanock

SJ, Vincent D, Bacot F, Tessier DC, Canisius S, Wessels LFA, Haiman CA, Shah M, Luben R, Brown J, Luccarini C, Schoof N, Humphreys K, Li J, Nordestgaard BG, Nielsen SF, Flyger H, Couch FJ, Wang X, Vachon C, Stevens KN, Lambrechts D, Moisse M, Paridaens R, Christiaens M-R, Rudolph A, Nickels S, Flesch-Janys D, Johnson N, Aitken Z, Aaltonen K, Heikkinen T, Broeks A, Veer LJV, van der Schoot CE, Guénel P, Truong T, Laurent-Puig P, Menegaux F, Marme F, Schneeweiss A, Sohn C, Burwinkel B, Zamora MP, Perez JIA, Pita G, Alonso MR, Cox A, Brock IW, Cross SS, Reed MWR, Sawyer EJ, Tomlinson I, Kerin MJ, Miller N, Henderson BE, Schumacher F, Le Marchand L, Andrulis IL, Knight JA, Glendon G, Mulligan AM, kConFab Investigators, Australian Ovarian Cancer Study Group, Lindblom A, Margolin S, Hooning MJ, Hollestelle A, van den Ouweland AMW, Jager A, Bui QM, Stone J, Dite GS, Apicella C, Tsimiklis H, Giles GG, Severi G, Baglietto L, Fasching PA, Haeberle L, Ekici AB, Beckmann MW, Brenner H, Müller H, Arndt V, Stegmaier C, Swerdlow A, Ashworth A, Orr N, Jones M, Figueroa J, Lissowska J, Brinton L, Goldberg MS, Labrèche F, Dumont M, Winqvist R, Pylkäs K, Jukkola-Vuorinen A, Grip M, Brauch H, Hamann U, Brüning T, GENICA (Gene Environment Interaction and Breast Cancer in Germany) Network, Radice P, Peterlongo P, Manoukian S, Bonanni B, Devilee P, Tollenaar RAEM, Seynaeve C, van Asperen CJ, Jakubowska A, Lubinski J, Jaworska K, Durda K, Mannermaa A, Kataja V, Kosma V-M, Hartikainen JM, Bogdanova NV, Antonenkova NN, Dörk T, Kristensen VN, Anton-Culver H, Slager S, Toland AE, Edge S, Fostira F, Kang D, Yoo K-Y, Noh D-Y, Matsuo K, Ito H, Iwata H, Sueta A, Wu AH, Tseng C-C, Van Den Berg D, Stram DO, Shu X-O, Lu W, Gao Y-T, Cai H, Teo SH, Yip CH, Phuah SY, Cornes BK, Hartman M, Miao H, Lim WY, Sng J-H, Muir K, Lophatananon A, Stewart-Brown S, Siriwanarangsan P, Shen C-Y, Hsiung C-N, Wu P-E, Ding S-L, Sangrajrang S, Gaborieau V, Brennan P, McKay J, Blot WJ, Signorello LB, Cai Q, Zheng W, Deming-Halverson S, Shrubsole M, Long J, Simard J, Garcia-Closas M, Pharoah PDP, Chenevix-Trench G, Dunning AM, Benitez J and Easton DF: Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet *45*: 353-361, 361e1-2, 2013.

33 Park DJ, Lesueur F, Nguyen-Dumont T, Pertesi M, Odefrey F, Hammet F, Neuhausen SL, John EM, Andrulis IL, Terry MB, Daly M, Buys S, Le Calvez-Kelm F, Lonie A, Pope BJ, Tsimiklis H, Voegele C, Hilbers FM, Hoogerbrugge N, Barroso A, Osorio A, Breast Cancer Family Registry, Kathleen Cuningham Foundation Consortium for Research into Familial Breast Cancer, Giles GG, Devilee P, Benitez J, Hopper JL, Tavtigian SV, Goldgar DE and Southey MC: Rare mutations in *XRCC2* increase the risk of breast cancer. Am J Hum Genet *90*: 734-739, 2012.

34 Broman KW, Murray JC, Sheffield VC, White RL and Weber JL: Comprehensive human genetic maps: individual and sex-specific variation in recombination. Am J Hum Genet 63: 861-869, 1998.

35 O'Connell JR and Weeks DE: PedCheck: a program for identification of genotype incompatibilities in linkage analysis. Am J Hum Genet *63*: 259-266, 1998.

36 Mukhopadhyay N, Almasy L, Schroeder M, Mulvihill WP and Weeks DE: Mega2: data-handling for facilitating genetic linkage and association analyses. Bioinformatics *21*: 2556-2557, 2005.

37 Sobel E and Lange K: Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. Am J Hum Genet *58*: 1323-1337, 1996.

38 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.

39 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M and DePristo MA: The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res *20*: 1297-1303, 2010.

40 DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D and Daly MJ: A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet *43*: 491-498, 2011.

41 Wang K, Li M and Hakonarson H: ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res *38*: e164, 2010.