ORIGINAL ARTICLE

Genetic sequence variations of BRCA1-interacting genes *AURKA*, *BAP1*, *BARD1* and *DHX9* in French Canadian Families with high risk of breast cancer

Frédéric Guénard, Yvan Labrie, Geneviève Ouellette, Charles Joly Beuparlant and Francine Durocher, INHERIT BRCA¹

Breast cancer is a heterogeneous disease displaying some degree of familial clustering. Highly penetrant breast cancer susceptibility genes represent approximately 20–25% of the familial aggregation of breast cancer. A significant proportion of this familial aggregation of breast cancer is thus yet to be explained by other breast cancer susceptibility genes. Given the high susceptibility conferred by the two major breast cancer predisposition genes, *BRCA1* and *BRCA2* and the implication of these genes in many key cellular processes, assessment of genes encoding BRCA1-interacting proteins as plausible breast cancer candidate genes is thus attractive. In this study, four genes encoding BRCA1-interacting proteins were analyzed in a cohort of 96 breast cancer individuals from high-risk non-*BRCA1/BRCA2* French Canadian families. Although no deleterious truncating germline mutations or aberrant spliced mRNA species were identified, a total of 10, 4, 11 and 6 variants were found in the *AURKA*, *BAP1*, *BARD1* and *DHX9* genes, respectively. The allele frequency of each variant was further ascertained in a cohort of 98 healthy French Canadian unrelated women and a difference in allele frequency was observed for one *BARD1* variant based on single-marker analysis. Haplotype estimation, haplotype blocks and tagging SNPs identification were then performed for each gene, providing a valuable tool for further searches of common disease-associated variants in these genes and therefore further analyses on these genes in larger cohorts is warranted in the search of low-to-moderate penetrance breast cancer susceptibility alleles.

Journal of Human Genetics (2009) 54, 152–161; doi:10.1038/jhg.2009.6; published online 6 February 2009

Keywords: BRCA1-interacting genes; breast cancer; haplotypes; tagging SNPs; variants

INTRODUCTION

Breast cancer is the most common malignant form of cancer among occidental women, with hereditary cases representing approximately 5–10%.¹ Mutations in high penetrance *BRCA1* and *BRCA2* genes account for only ~25% of families with hereditary breast cancer² whereas other lower penetrance genes such as *ATM*, *CHEK2*, *TP53*, *PTEN*, *STK11*, *CDH1*, *PALB2/FANCF* and *BRIP1/FANCD1*³ explain a much smaller proportion of families with hereditary breast cancer. Deleterious mutations in *BRCA1/BRCA2* were identified in 24% of our 256 high-risk French Canadian breast/ovarian cancer families,⁴ supporting that a significant proportion of high-risk breast cancer families still remain unexplained by either germline mutations in high penetrance *BRCA1/2* alleles or other lower penetrance genes.^{5–10}

Several structural motifs allow the BRCA1 protein to interact with cellular proteins to regulate diverse biological functions such as control of transcription¹¹ and DNA damage repair.¹² Therefore, BRCA1-interacting proteins implicated in these cellular pathways

represent attractive candidates with regard to breast cancer susceptibility. Based on a candidate gene strategy, four genes encoding BRCA1-interacting proteins, namely *AURKA*, *BAP1*, *BARD1* and *DHX9*, were thoroughly investigated in an attempt to identify germline deleterious mutations and/or aberrant splicing events, which could potentially be associated with an increase in breast cancer risk. In this study, familial breast cancer cases were purposely selected. Indeed, such a study design has been demonstrated to substantially decrease the number of cases and controls to achieve the same power, as compared with studies where cases are unselected for family history.¹³

The *AURKA* protein is a centrosome-associated Ser/Thr kinase which localizes to the centrosome and could be involved in centrosome duplication,¹⁴ maturation and separation, as well as in the control of bipolar spindle assembly.¹⁵ *AURKA* binds and phosphorylates BRCA1, and it was suggested that this BRCA1 phosphorylation plays a role in G2/M transition.¹⁶ The *AURKA* gene has been recently

Cancer Genomics Laboratory, Oncology and Molecular Endocrinology Research Centre, Centre Hospitalier Universitaire de Québec and Laval University, Québec, Canada
Correspondence: Dr F Durocher, Cancer Genomics Laboratory, Oncology and Molecular Endocrinology Research Centre, Centre Hospitalier Universitaire de Québec and Laval University, 2705 Laurier Boulevard, Bloc T2-53, Québec, Canada, G1V 4G2.
E-mail: Francine.durocher@crchul.ulaval.ca

¹Other members of INHERIT BRCA¹ involved in this study are listed in the Appendix.

Received 20 August 2008; revised 23 December 2008; accepted 6 January 2009; published online 6 February 2009

identified as a low-penetrance tumor susceptibility gene,¹⁷ and was subsequently analyzed for its implication with breast cancer albeit leading conflicting results.^{18–27}

BAP1 is an ubiquitin carboxy-terminal hydrolase which binds to the BRCA1 RING finger domain and enhances BRCA1 growth suppression properties.²⁸ BAP1 was suggested to regulate the BRCA1/BRCA2-RAD51 DNA repair complex.²⁹ Given the known implication of BRCA1 in DNA repair¹² and its interaction with RNA pol II,¹¹ which is linked to transcription-coupled repair (TCR) processes, a role for BAP1 in TCR has been suggested.³⁰ It has been reported that mutation of an amino acid of BAP1 in its BRCA1-binding domain (L691P) abolishes the BAP1-BRCA1 interaction.²⁸ On the other hand, a study on 47 non-BRCA1/BRCA2 familial breast cancer cases from the French population did not reveal any association with breast cancer predisposition.³¹

The BARD1 protein interacts with BRCA1 and both proteins form a complex which possesses a dual E3 ubiquitin ligase activity.³² The heterodimer BRCA1/BARD1 promotes its own ubiquitination³³ and ubiquitinates RNA pol II subunit A.³⁴ The interaction of BRCA1 with RNA pol II¹¹ and DHX9³⁵ suggested a role for BARD1 in TCR. The *BARD1* gene has been reported to be targeted by somatic³⁶ and germline^{37–39} mutations in breast and in ovarian^{36,40} cancer cohorts. *BARD1* mutations have been identified in hereditary breast/ovarian cancers from BRCA1/2-negative patients^{37,41} and it was suggested that the Cys557Ser variant may represent a low-penetrance breast cancer susceptibility allele.^{39,42,43} However, a recent study did not identify any increased risk of breast cancer associated with specific *BARD1* variations,⁴⁴ whereas another study reported an association of *BARD1* haplotypes with either an increased or decreased risk of breast cancer.⁴⁵

As for the DHX9 helicase, a member of the DEAH helicase family,⁴⁶ it binds directly to both CBP and pol II and is required for complex formation between both proteins.⁴⁷ The DHX9 protein (aa 230–325) binds the full-length BRCA1 protein through a subregion of the BRCT domain of BRCA1.³⁵ DHX9 truncating mutations were reported to affect the interaction of both proteins (BRCA1 and RNA pol II) and to result in decreased transcriptional activity of BRCA1,³⁵ thus possibly affecting its function in transcription and TCR. Despite an interaction with BRCA1 and a role in transcriptional activity of BRCA1, DHX9 mutations have not been assessed thus far in non-BRCA1/BRCA2 breast cancer families.

Based on the multiple functions of these BRCA1-interacting genes, they obviously represent attractive candidates that might explain a proportion of the remaining inherited breast cancer cases. The goal of this study being the identification of deleterious mutation, and given that such deleterious disease-causing mutations producing a premature truncation of the protein occur mostly within the coding region, complementary DNA (cDNA) mutation screening was therefore performed. In addition, since alternative splicing is now recognized as the primary source of human proteomic diversity and gene regulation and that an increasing number of disease-causing mutations affecting the correct splicing of genes are identified,⁴⁸ cDNA material definitively represents a resource of choice as alternative splice forms could not be detected when extensive re-sequencing is performed on genomic DNA.

Thus the goal of this study was to evaluate the possible involvement of *AURKA*, *BAP1*, *BARD1* and *DHX9* germline deleterious mutations or on breast cancer susceptibility by analyzing all mRNA-encoding sequences as well as alternative splicing in cDNA samples from 96 breast cancer cases, in which no BRCA1/2 mutations were detected.

MATERIALS AND METHODS

Ascertainment of families and DNA extraction

High-risk French Canadian breast and/or ovarian families were recruited through a large ongoing interdisciplinary research program designated INHERIT BRCAs. Clinicians (listed in Appendix) were directly involved in this translational research program and were responsible for the BRCA1/2 test result disclosure to participants. Approval was obtained from ethics committees corresponding to the different participating institutions. More details regarding ascertainment criteria, experimental and clinical procedures, as well as the INHERIT BRCAs research program have been described elsewhere.^{4,5} Subsequently, another component was designed for the 'Localization and identification of novel breast cancer susceptibility loci/genes'. Ethics approval for this latter study was also obtained and each participant knowing their inconclusive BRCA1/2 test results status had to sign a specific informed consent for their participation in this component. A subset of 96 high-risk French Canadian breast/ovarian cancer families were recruited in this study according to the ascertainment criteria described previously.^{5,7,9} Recruitment of 98 healthy unrelated French Canadian individuals was conducted on a non-nominative basis, as part of a long term study aiming at the characterization of the genetic variability in human population and approved by the Institutional Ethics Review Board of the Hôpital Sainte-Justine.

DNA extraction, RNA isolation from immortalized cell lines and cDNA synthesis

Lymphocytes from breast cancer individuals were isolated and immortalized as previously described.⁵ Total RNA extraction, storage and reverse transcription of RNA, as well as genomic DNA extraction from healthy unrelated French Canadian individuals were also performed as previously described.^{5,6,8–10}

PCR amplification and mutation analysis

cDNA amplicons covering the entire mRNA-encoding sequences of the GenBank mRNA records (*AURKA*: NM_198436.1; *BAP1*: NM_004656.2; *BARD1*: NM_000465.1; *DHX9*: NM_001357.3) were obtained by PCR amplification from cDNA of breast cancer cases using the primer pairs listed in Supplementary Table 1. Direct sequencing and sequence data analysis were performed as previously described.⁵ Genomic DNA sequences from healthy unrelated French Canadian individuals were obtained using the same direct sequencing method, but with amplification of relevant variant-containing exons from amplicons obtained using intronic primers (Supplementary Table 1). The segregation of rare missense and deletion variants (minor allele frequency; MAF <5%), and variants demonstrating a suggestive difference in genotype frequency between breast cancer cases and unaffected individuals were also analyzed in families for which DNA material was available from multiple individuals.

Each variant was tested for deviation from HWE by means of a χ^2 test. All *P*-values were two-sided with 1 degree of freedom. To estimate the genetic association of each variation or haplotype with breast cancer, differences in genotype distributions between the case and control groups were tested using a χ^2 test. *P*-values less than 0.05 are considered as significant. However, to address the issue of multiple comparison, Bonferroni correction was applied for each estimate of genetic association.

Conservation in other species and computational analyses

Protein alignment was performed using data extracted from the UCSC database, and both the SIFT and PolyPhen web-based softwares were used to predict the effect of amino acid substitutions. All tests were run under default threshold values.

LD analysis and haplotype estimation

To estimate the pattern of LD, the LDA program was used to calculate pairwise LD (D' and r^2) values. Haplotype analysis was performed with the PHASE 2.1.1 software. Haplotype frequencies were estimated using all variants genotyped in both sample series (cases and controls) and a permutation test was conducted to determine the significance of differences in the distribution of inferred haplotypes between both groups. All association tests were run under default conditions, with 1000 permutations. Haplotype blocks were identified using

genotyping data from both series using the Haploview software. tSNPs from each LD block were thereafter identified using the same software, and haplotype frequencies were also estimated using only the tSNPs identified.

RESULTS

The entire coding sequence of *AURKA*, *BAP1*, *BARD1* and *DHX9* genes were analyzed using cDNA from 96 high-risk breast cancer individuals, each coming from a distinct high-risk family. No deleterious truncating germline mutations were found. In addition, based on gel electrophoresis and sequence chromatogram analysis, no major alternative splice forms have been detected for any of the genes analyzed, unlikely involving this kind of alteration with breast cancer risk. However, a number of sequence variations were identified within each of these four genes, as described below.

AURKA gene analysis

Five of the 10 variants identified in the *AURKA* gene are coding substitutions, whereas the remaining five are located within the 3'UTR region (Table 1). Four of the *AURKA* variants are novel, not being reported in dbSNP database (Build 127). Albeit no significant difference was observed between both sample sets based on single-marker analysis, the novel c.1130T>A variation is present exclusively among three breast cancer cases at the heterozygous state. As for the c.1117A>G (Met373Val) variant present in four breast cancer individuals at the heterozygous state, segregation analysis suggests a possible association with breast cancer. Indeed as illustrated in the pedigree from one of the four carrier's families (Figure 1), all heterozygous carriers are affected with breast cancer whereas among the two non-carriers, one is affected (diagnosed at 45 years) and the other is unaffected (55 years). A DNA sample from an additional family member was also available for two other families. An unaffected first-degree relative (59 years) could be genotyped and was found to be a non-carrier of the Met373Val variant in one family, whereas an affected third-degree relative diagnosed with breast cancer at 74 years was found to be a heterozygote carrier for this variant in the other family (data not shown). No additional family member was available for the remaining family. With the exception of the c.1212+423G>C variant which displays a slight excess of homozygotes (Hardy-Weinberg equilibrium (HWE) $P=0.03$) among the breast cancer cases, no variants display a significant deviation from HWE.

As shown in Table 2, all variants located in the coding sequence result in an amino acid change. Of these, three residues (Val57, Ala213 and Val377) are completely conserved in distant species (apart from Leu377 in *Md*), suggesting that they are under strong functional constraint or may have a specific role on protein conformation. The three variants (Ala213Val, Met373Val and Val377Glu) located within the Ser/Thr protein kinase domain are predicted to be not tolerated by SIFT, with Val377Glu also evaluated as probably damaging by the PolyPhen program (Table 3).

A graphical representation of linkage disequilibrium (LD) with both r^2 and Lewontin's D' and using data from both series is shown in Figure 2. Most single nucleotide polymorphism (SNP) pairs display strong LD ($0.8 < D' \leq 1$), whereas lower LD is observed for a combination involving SNPs 1, 2, 4 and 6. The r^2 coefficient calculated displays a large spectrum of values ranging from 0 to 0.6, given the high dependence of r^2 on allele frequency. Analysis of haplotypes estimated with the PHASE program led to six haplotypes exhibiting a frequency $\geq 1\%$ (Table 4). Based on solid block algorithm, two LD blocks were identified at the *AURKA* locus, with LD breakage being located in the region of exon 9 (Figure 3). However, when performing a similar analysis using CEPH/CEU data, three LD blocks are

obtained, including two relatively small blocks (0.5 and 1.2 kb) in the 5' region and a larger block covering intron 1 to intron 7 (17 kb) (data not shown). As indicated in Figure 3, three tagging SNPs (tSNPs) (rs2273535, rs1047972 and rs34189236) have been identified, representing 95% of our estimated haplotypes.

BAP1 gene analysis

Analysis of *BAP1* sequence in our breast cancer individuals led to the identification of only three variants (Table 1). Two are novel synonymous variants, c.294C>T being detected only once among our 96 breast cancer individuals, whereas c.1026C>T was present exclusively in two unaffected individuals at the heterozygous state. No significant difference in allele frequency was obtained between both series for any of these nucleotide substitutions based on single-marker analysis.

All *BAP1* variations are comprised within one block of strong LD (Figure 3), which is confirmed by the perfect LD ($D'=1$) observed between the two most distant intragenic SNPs (SNP1 and SNP4) (Figure 2). As displayed in Table 4, PHASE analysis estimated two major haplotypes, which are represented by only one tSNP (rs123598; SNP4) (Figure 3), and both haplotypes display a similar frequency in both sample sets. Given that only three SNPs, all located in intron 3, are available from HapMap data, it was not relevant to perform haplotype and tSNP identification using the CEU population.

BARD1 gene analysis

Among the *BARD1* sequence variations identified in this study, six are missense variants, whereas another consists of an in frame 21 bp deletion (seven amino acids) in exon 4 (c.1075del21). Three (c.820A>G, c.1203T>C and c.2212A>G) are considered as novel, being not reported in dbSNP database (Table 1). All variants follow HWE, with the exception of the c.1670G>C variant, where a significant deviation is observed due to the occurrence of one rare homozygote. A significant difference is also observed for the c.1518C>T substitution, which is slightly more frequent within cases ($P=0.013$, data not shown). However, this association does not reach statistical significance following Bonferroni correction, and segregation of this variant in a large informative family does not reveal any obvious association with breast cancer (data not shown). Of the seven amino acid changes identified, only the Ile738Val residue displays a certain level of conservation in higher species, with the alternative Val residue being observed in rodents and *Tetraodon nigroviridis* (Table 2). Apart from the deletion of 7 amino acids, *in silico* evaluation of the potential effect of amino acid changes on the protein function (Table 3) revealed that four amino acid changes are predicted to affect the protein function by at least one of the two programs used, including the Arg378Ser change which is predicted to potentially affect the protein function by both SIFT and PolyPhen softwares.

The graphical representation of the pairwise LD between the 11 variants identified in both series combined is shown in Figure 2. The low pairwise D' values observed involve mainly SNP8, whereas moderate pairwise values are observed for SNP1. A large spectrum of r^2 values ranging from 0 to 1.0 is observed for *BARD1*. The Haploview software (Figure 3) determined two LD blocks, with breakage of LD located between SNP3 and SNP4 in the region of exon 4. Within these two LD blocks, 2 and 6 tSNPs were identified, block 1 being defined by SNPs 1 and 2, whereas SNPs 4, 6, 8, 9, 10 and 11 represent the second block. Similar analyses using CEPH/CEU data set led to the identification of seven LD blocks (data not shown). The discordance obtained is present mainly in the 5'-region of the *BARD1*

Table 1 Sequence variations and genotype frequencies of *AURKA*, *BAP1*, *BARD1* and *DHX9* genes in cases and controls

Gene	SNP	SNP ID ^a	dbSNP ID	Amino acid change	Location	Series	Number of individuals	Common homozygote (expected) ^b	Heterozygote (expected) ^b	Rare homozygote (expected) ^b	Minor allele frequency (MAF)	Dev from HWE P-value
AURKA	1	c.91T>A	rs2273535	Phe31Ile	Exon 4	Cases	96	64 (61.0)	25 (31.1)	7 (4.0)	0.203	0.06
						Controls	96	63 (61.8)	28 (30.5)	5 (3.8)	0.198	0.43
	2	c.169G>A	rs1047972	Val57Ile	Exon 4	Cases	96	62 (61.8)	30 (30.5)	4 (3.8)	0.198	0.88
						Controls	96	69 (67.5)	23 (26.0)	4 (2.5)	0.161	0.26
	3	c.638C>T	NA	Ala213Val	Exon 7	Cases	96	95 (95.0)	1 (1.0)	0 (0.0)	0.005	0.96
						Controls	97	97 (0.0)	0 (0.0)	0 (0.0)	0.000	—
	4	c.1117A>G	rs33923703	Met373Val	Exon 10	Cases	96	92 (92.0)	4 (3.9)	0 (0.0)	0.021	0.83
						Controls	95	89 (89.1)	6 (5.8)	0 (0.1)	0.032	0.75
	5	c.1130T>A	NA	Val377Glu	Exon 10	Cases	96	93 (93.0)	3 (3.0)	0 (0.0)	0.016	0.88
						Controls	96	96 (0.0)	0 (0.0)	0 (0.0)	0.000	—
	6	c.1212+3A>C	rs34361943	—	3'-UTR	Cases	96	92 (92.0)	4 (3.9)	0 (0.0)	0.021	0.83
						Controls	95	89 (89.1)	6 (5.8)	0 (0.1)	0.032	0.75
	7	c.1212+69G>A	rs34189236	—	3'-UTR	Cases	96	94 (94.0)	2 (2.0)	0 (0.0)	0.010	0.92
						Controls	95	94 (94.0)	1 (1.0)	0 (0.0)	0.005	0.96
	8	c.1212+370C>T	NA	—	3'-UTR	Cases	96	94 (94.0)	2 (2.0)	0 (0.0)	0.010	0.92
						Controls	95	95 (0.0)	0 (0.0)	0 (0.0)	0.000	—
	9	c.1212+422C>G	NA	—	3'-UTR	Cases	96	95 (95.0)	1 (1.0)	0 (0.0)	0.005	0.96
						Controls	95	95 (0.0)	0 (0.0)	0 (0.0)	0.000	—
	10	c.1212+423G>C	rs8173	—	3'-UTR	Cases	96	60 (56.3)	27 (34.5)	9 (5.3)	0.234	0.03
						Controls	95	54 (54.6)	36 (34.9)	5 (5.6)	0.242	0.75
BAP1	1	c.294C>T	NA	Ser98Ser	Exon 5	Cases	96	95 (95.0)	1 (1.0)	0 (0.0)	0.005	0.96
						Controls	94	94 (0.0)	0 (0.0)	0 (0.0)	0.000	—
	2	c.1002A>G	rs28997577	Leu334Leu	Exon 11	Cases	96	95 (95.0)	1 (1.0)	0 (0.0)	0.005	0.96
						Controls	92	91 (91.0)	1 (1.0)	0 (0.0)	0.005	0.96
	3	c.1026C>T	NA	Ser342Ser	Exon 11	Cases	96	96 (0.0)	0 (0.0)	0 (0.0)	0.000	—
						Controls	93	91 (91.0)	2 (2.0)	0 (0.0)	0.011	0.92
BARD1	4	c.2190+444C>T	rs123598	-	3'-UTR	Cases	96	81 (80.7)	14 (14.7)	1 (0.7)	0.083	0.66
						Controls	87	74 (73.6)	12 (12.9)	1 (0.6)	0.080	0.53
	1	c.70C>T	rs1048108	Pro24Ser	Exon 1	Cases	96	52 (48.9)	33 (39.2)	11 (7.9)	0.286	0.12
						Controls	95	46 (45.2)	39 (40.7)	10 (9.2)	0.311	0.69
	2	c.609A>C	rs28997574	Gly203Gly	Exon 4	Cases	96	95 (95.0)	1 (1.0)	0 (0.0)	0.005	0.96
						Controls	86	81 (81.1)	5 (4.8)	0 (0.1)	0.029	0.78
	3	c.820A>G	NA	Ser274Gly	Exon 4	Cases	96	95 (95.0)	1 (1.0)	0 (0.0)	0.005	0.96
						Controls	90	90 (0.0)	0 (0.0)	0 (0.0)	0.000	—
	4	c.1053G>C	rs2070096	Thr351Thr	Exon 4	Cases	96	56 (59.4)	39 (32.2)	1 (4.4)	0.214	0.04
						Controls	89	51 (52.0)	33 (32.6)	5 (5.2)	0.242	0.91
	5	c.1075del21	rs28997575	delLPECSSP aa 359-365	Exon 4	Cases	96	92 (92.0)	4 (3.9)	0 (0.0)	0.021	0.83
						Controls	87	85 (85.0)	2 (2.0)	0 (0.0)	0.000	0.91
	6	c.1134G>C	rs2229571	Arg378Ser	Exon 4	Cases	96	12 (13.9)	49 (45.2)	35 (36.9)	0.620	0.42
						Controls	88	12 (13.9)	46 (42.2)	30 (32.0)	0.602	0.39
	7	c.1203T>C	NA	Ser401Ser	Exon 4	Cases	96	95 (95.0)	1 (1.0)	0 (0.0)	0.005	0.96
						Controls	87	87 (0.0)	0 (0.0)	0 (0.0)	0.000	—
	8	c.1518C>T	rs2070093	His506His	Exon 6	Cases	96	52 (52.5)	38 (37.0)	6 (6.5)	0.260	0.79
						Controls	98	68 (69.5)	29 (26.1)	1 (2.5)	0.158	0.27
	9	c.1519G>A	rs2070094	Val507Met	Exon 6	Cases	96	46 (45.4)	40 (41.3)	10 (9.4)	0.313	0.77
						Controls	98	39 (39.9)	47 (45.3)	12 (12.9)	0.362	0.71

Table 1 Continued

Gene	SNP	SNP ID ^a	dbSNP ID	Amino acid change	Location	Series	Number of individuals	Common homozygote (expected) ^b	Heterozygote (expected) ^b	Rare homozygote (expected) ^b	Minor allele frequency (MAF)	Dev from HWE P-value
DHX9	10	c.1670G>C	rs28997576	Cys557Ser	Exon 7	Cases	96	93 (93.0)	3 (3.0)	0 (0.0)	0.016	0.88
						Controls	98	95 (94.0)	2 (3.9)	1 (0.0)	0.020	<0.01
	11	c.2212A>G	NA	Ile738Val	Exon 11	Cases	96	95 (95.0)	1 (1.0)	0 (0.0)	0.005	0.96
						Controls	97	92 (92.1)	5 (4.9)	0 (0.1)	0.026	0.79
	1	c.265C>G	NA	Pro89Ala	Exon 4	Cases	96	95 (95.0)	1 (1.0)	0 (0.0)	0.005	0.96
						Controls	94	94 (0.0)	0 (0.0)	0 (0.0)	0.000	—
	2	c.915T>G	NA	Ser305Ser	Exon 10	Cases	96	93 (93.0)	3 (3.0)	0 (0.0)	0.016	0.88
						Controls	96	92 (92.0)	4 (3.9)	0 (0.0)	0.021	0.83
	3	c.1745C>T	rs2275177	Thr582Thr	Exon 16	Cases	96	81 (78.8)	12 (16.3)	3 (0.8)	0.094	0.01
						Controls	91	80 (79.4)	10 (11.2)	1 (0.4)	0.066	0.30
	4	c.1873C>G	NA	Ser625Cys	Exon 17	Cases	96	96 (0.0)	0 (0.0)	0 (0.0)	0.000	—
						Controls	91	90 (90.0)	1 (1.0)	0 (0.0)	0.005	0.96
	5	c.1976C>T	rs35533583	Gly659Gly	Exon 17	Cases	96	91 (91.1)	5 (4.9)	0 (0.1)	0.026	0.79
						Controls	91	84 (84.1)	7 (6.7)	0 (0.1)	0.038	0.70
	6	c.2708C>T	NA	p.Ile903Ile	Exon 23	Cases	96	95 (95.0)	1 (1.0)	0 (0.0)	0.005	0.96
						Controls	94	93 (93.0)	1 (1.0)	0 (0.0)	0.005	0.96

^aAccording to the nomenclature of the Human Genome Variation Society.^bAs expected under Hardy-Weinberg equilibrium.

gene, in which the four haplotype blocks identified from HapMap data are represented by a single block in the French Canadian population. Regarding haplotypes estimation, the analysis led to the identification of 11 haplotypes, all exhibiting a frequency $\geq 1\%$ (Table 4). No significant difference in single or global haplotype distribution was identified between both groups.

DHX9 gene analysis

The analysis of *DHX9* in our breast cancer individuals led to the identification of five variants, with an additional nucleotide substitution present exclusively among unaffected individuals (Table 1). Out of these variations, two involve an amino acid change, among which c.1873C>G is located in the proximity of the helicase C-terminal domain (aa 631–776). The c.1745C>T (rs2275177) sequence variation demonstrates a significant deviation from HWE in breast cancer cases, where an excess of rare homozygous genotypes is observed. Single-marker analysis did not reveal a significant difference in MAF for any of the nucleotide changes identified. Regarding the amino acid conservation of the Pro89Ala variant in other orthologs, the Pro allele is conserved in the three species in which this residue is available (*Canis* and rodents, Table 2). The two *DHX9* amino acid changes identified are predicted to be possibly damaging by the PolyPhen program. In addition, the Ser625Cys located within the RNA pol II-binding domain (aa 230–650) according to SIFT is predicted to be not tolerated (Table 3).

Most pairs of adjacent SNPs display strong LD ($0.8 < D' \leq 1$) following LDA analysis. As displayed in Figure 2, r^2 values under 0.3 are obtained for all pairwise comparisons. The perfect D' value associated with the two most distant variations identified is confirmed by Haploview analysis (Figure 3). Three tSNPs were highlighted for *DHX9*, namely c.915T>G, c.1745C>T and c.1976C>T (Figure 3). Four haplotypes exhibiting a frequency $\geq 1\%$ have been estimated with the PHASE program (Table 4).

DISCUSSION

Given that deleterious mutations in *BRCA1/BRCA2* were identified in only 24% of the 256 high-risk French Canadian breast/ovarian cancer families,⁴ the analysis of the coding region of four genes encoding *BRCA1*-interacting proteins was undertaken in 96 high-risk individuals drawn from non-*BRCA1/BRCA2* French Canadian families. To our knowledge, this is the first study to investigate the potential association of *DHX9* germline mutations with breast cancer susceptibility in high-risk non-*BRCA1/BRCA2* breast cancer families. Although *AURKA* and *BARD1* genes have been investigated in tumor samples, population-based studies or high-risk families, only one investigation assessed the involvement of *BAP1* in breast cancer susceptibility.³¹

The Phe31Ile variant was reported to confer an increased risk of breast cancer, either alone or in combination with Val57Ile.^{18–21,23,27} However, our study and others did not replicate this association.^{22,24–26} With the exception of c.1212+423G>C (rs8173), evaluated in both breast⁴⁴ and ovarian⁴⁹ cancer, none of the 3'UTR *AURKA* variants identified in our study were assessed in other studies. As for the Met373Val variant, we report here a possible implication of this variant with breast cancer susceptibility, segregation analysis indicating that all heterozygous carriers are affected with breast cancer. However, this rare variant is observed at a similar frequency in affected and unaffected individuals from our study, thus requiring further analyses.

Many *BARD1* variants identified in this study have been evaluated in the context of breast cancer. The c.1670G>C variant was associated with an increased risk of breast cancer either in high-risk families or in

Figure 1 Pedigree of AURKA:Met373Val carriers family. The age of onset of cancer, age of death or current age are indicated below family members. The affected case that was sequenced first is designated by an arrow. The genotype call is indicated below the respective individual analyzed. d, deceased.

Table 2 Non-synonymous sequence variants detected in *AURKA*, *BAP1*, *BARD1* and *DHX9* genes and corresponding residues present in orthologs

Gene	SNP ^a	SNP ID ^b	Amino acid change	Canis familiaris	Mus musculus	Rattus norvegicus	Monodelphis domestica	Gallus gallus	Xenopus tropicalis	Tetraodon nigroviridis
AURKA	1	c.91T>A	Phe31Ile	Val	Ile	Ile	Ile	—	—	—
	2	c.169G>A	Val57Ile	Val	Val	Val	Val	Val	—	—
	3	c.638C>T	Ala213Val	Ala	Ala	Ala	Ala	Ala	Ala	Ala
	4	c.1117A>G	Met373Val	Thr	Thr	Thr	Thr	Pro	Ser	Pro
	5	c.1130T>A	Val377Glu	Val	Val	Val	Leu	Val	Val	Val
BARD1	1	c.70C>T	Pro24Ser	Pro	Pro	Pro	Ser	Ala	—	—
	3	c.820A>G	Ser274Gly	Ser	Ser	Ser	Val	Asn	—	—
	5	c.1075del21	delLPECSSP	LPGCSP	LVDCSSL	LVDCSSL	LTESAPS	Q-DCPIP	QR—	—
	6	c.1134G>C	Arg378Ser	Arg	Arg	Arg	Val	Lys	Arg	—
	9	c.1519G>A	Val507Met	Leu	Ile	Ile	Val	Val	Ile	Leu
	10	c.1670G>C	Cys557Ser	Tyr	Cys	Cys	Cys	Pro	Cys	—
DHX9	11	c.2212A>G	Ile738Val	Ile	Val	Val	Ile	Ile	Val	Ile
	1	c.265C>G	Pro89Ala	Pro	Pro	Pro	—	—	—	—
	4	c.1873C>G	Ser625Cys	Ala	Ser	Ser	Ala	—	Ser	Ala

^aAccording to Table 1.

^bAccording to the nomenclature of the Human Genome Variation Society.

—, No corresponding residue.

population-based studies,^{38,39} whereas our results and others^{43,50} do not support these findings. None of the other variants identified here (that is, c.609A>C, c.1053G>C, c.1075del21 and c.1518C>T) was previously associated with an increased risk of breast cancer,^{38,41} despite our results suggesting an association of the c.1518C>T (His506His) with an increased risk of breast cancer ($P=0.013$), which is however not significant following Bonferroni correction. This finding would nonetheless have to be confirmed in much bigger cohorts.

The *BAP1* analysis led to the identification of the c.1026C>T (Ser342Ser) variant, which was reported once in a cohort of 47 French familial breast cancer cases negative for *BRCA1/BRCA2* mutations.³¹

Interestingly the common c.2190+444C>T sequence variation localized in the 3'UTR region was identified in our population with a MAF>5% in both series but was not analyzed in the French cohort.³¹

As for *DHX9* sequence variations, we report here the first assessment of its implication in breast cancer. Of the six sequence variants identified, four are rare, novel variants, two of which being missense variants. The Ser625Cys variation is located within the RNA pol II-binding domain (aa 230–650) close to the C-terminal helicase domain (aa 631–776), whereas the Pro89Ala is situated in the N-terminal basic region capable of binding efficiently the C/H3 region of CBP transcription coactivator (aa 1–250).⁴⁷

Table 3 Location of non-synonymous sequence variants detected in *AURKA*, *BAP1*, *BARD1* and *DHX9* genes and prediction on protein function

Gene	SNP ^a	Amino acid change	Domain	SIFT	PolyPhen (score difference)
AURKA	1	Phe31Ile	None	Tolerated	Benign
	2	Val57Ile	None	Tolerated	Benign
	3	Ala213Val	Ser/Thr protein kinase	Not tolerated	Benign
	4	Met373Val	Ser/Thr protein kinase	Not tolerated	Benign
	5	Val377Glu	Ser/Thr protein kinase	Not tolerated	Probably damaging (2.39)
BARD1	1	Pro24Ser	None	Not tolerated	Benign
	3	Ser274Gly	None	Tolerated	Benign
	5	delLPECSSP	None	NA	NA
	6	Arg378Ser	None	Not tolerated	Possibly damaging (1.79)
	9	Val507Met	Ankyrin repeats	Not tolerated	Benign
	10	Cys557Ser	Close from BRCT (aa 568–767)	Tolerated	Probably damaging (2.52)
DHX9	11	Ile738Val	BRCT	Tolerated	Benign
DHX9	1	Pro89Ala	CBP-binding domain (aa 1–250)	Tolerated	Possibly damaging (1.56)
	4	Ser625Cys	RNA pol II-binding domain (aa 230–650)	Not tolerated	Possibly damaging (1.53)

^aAccording to Table 1. **Figure 2** Graphical overview of pairwise linkage disequilibrium (LD) measures of (D') (lower left part) and r^2 (higher right part) for each SNPs pair identified in breast cancer cases and healthy individuals. All SNPs are denoted numerically with reference to Table 1. Distance between first and last SNPs for each gene is indicated on the left end on each LD representation.

In addition to establishing the allele frequency of each variant in a series of unaffected controls, the effect of the variants identified were also assessed using conservation analysis and web-based programs. Both sequencing chromatograms and gel electrophoresis were thoroughly analyzed to detect any aberrant spliced forms, and we were able to conclude that no major spliced mRNAs are highly expressed in the cDNA material from our high-risk breast cancer individuals. Despite the fact that alternative transcripts expressed at low levels (~10%

compared to the wild type form) can be detected by sequencing chromatogram analysis,^{5,51} we cannot exclude the presence of minor, weakly expressed forms which could not have been detected here. The identification of alternative spliced forms of *BARD1*^{40,52,53} and *AURKA*⁵⁴ mRNA have been reported, which is in accordance with UCSC database data and other alternative transcript databases such as ASAP II and Fast DB. As for *BAP1*, a splice variant was reported in a small cell lung cancer cell line.²⁸ However, discordant results are presented in different databases, UCSC identifying many alternative spliced mRNA species for both *BAP1* and *DHX9* genes, whereas the database Fast DB identified only few alternative spliced mRNA species for *BAP1*, and none for *DHX9*.

In terms of linkage disequilibrium, relatively strong LD ($0.8 < D' \leq 1$) is observed for the four genes when using the D' measure (Figure 2), and the D' values obtained here are concordant with data from the sole study reporting LD values (D') for the *AURKA* gene.²² Using the r^2 correlation coefficient, much lower LD values were obtained. Indeed differences between D' and r^2 values are likely to be explained by the fact that r^2 is more stringent and sensitive to frequency than D' .⁵⁵

To our knowledge, the current study is the first to report haplotype estimation for the *BAP1* and *DHX9* genes. The sole other study assessing *BARD1* haplotypes⁴⁵ reported a significant difference in haplotype distribution between cases and controls for two different haplotypes. Although both haplotypes have indeed been estimated in our analysis (*BARD1*-h1 and *BARD1*-h6), this association was not observed. The current study is also the most extensive investigation conducted on *AURKA* haplotypes, which includes 11 variations encompassing exons 3–9. Our results do not confirm any association of *AURKA* haplotypes with an increased risk of breast cancer, as previously demonstrated in other studies.^{19,22,23}

The identification of haplotype blocks was thereafter conducted for the *AURKA*, *BAP1*, *BARD1* and *DHX9* genes. With the exception of the *AURKA* gene, for which the 5'-end can be roughly separated into two clusters,²² this analysis was not previously performed or reported for the *BAP1*, *BARD1* and *DHX9* genes. The haplotype block identification for *BAP1* is concordant with results obtained from HapMap data, whereas similar analyses using French Canadian data for the *AURKA*, *BARD1* and *DHX9* genes revealed slight differences when compared with HapMap data (data not shown). However, we

Table 4 Estimated haplotype frequencies in cases and controls using PHASE

Gene	Haplotype number ^a	Haplotypes (all SNPs used) ^b	Haplotypes (tSNPs only)	Estimated haplotype frequencies (tSNPs only)			
				Cases	Controls	Combined	Cumulative
AURKA	h1	TGCATAGCCG	TGA ^c	0.600	0.635	0.618	0.618
	h2	AGCATAGCCC	AGA ^c	0.196	0.186	0.191	0.809
	h3	TACATAGCCG	TAA ^c	0.182	0.145	0.164	0.973
	h4	AGCATAGCCG	AGA ^c	—	—	—	—
	h5	TGCATAGCCC	TGA ^c	—	—	—	—
	h6	TACATAGCCC	TAA ^c	—	—	—	—
BAP1	h1	CACC	C	0.917	0.919	0.918	0.918
	h2	CACT	T	0.083	0.081	0.082	1.000
BARD1	h1	TAAGACTCAGA	TAGCCAGA	0.214	0.227	0.220	0.220
	h2	CAACAGTCGGA	CACGCCGA	0.187	0.233	0.210	0.430
	h3	CAAGACTTGGGA	CAGCTGGA	0.176	0.099	0.137	0.567
	h4	CAAGACTCGGA	CAGCCGGA	0.097	0.088	0.093	0.660
	h5	CAAGAGTCGGA	CAGGCCGA	0.083	0.092	0.088	0.748
	h6	CAAGACTCAGA	CAGCCAGA	0.053	0.080	0.066	0.814
	h7	CAAGAGTTGGA	CAGGTGGA	0.061	0.045	0.053	0.867
	h8	TAAGACTCGGA	TAGCCGGA	0.039	0.044	0.042	0.909
	h9	CAAGACTCACCA	CAGCCACA	0.021	0.020	0.021	0.930
	h10	CCAGACTCAGG	CCGCCAGG	0.005	0.025	0.015	0.945
	h11	TAAGAGTCGGA	TAGGCCGA	0.014	0.013	0.014	0.959
DHX9	h1	CTCCCC	TCC	0.905	0.887	0.856	0.856
	h2	CTTCCC	TTC	0.036	0.072	0.054	0.910
	h3	CTTCTC	TTT	0.031	0.020	0.025	0.935
	h4	CGCCCC	GCC	0.021	0.015	0.018	0.953

^aOnly haplotypes exhibiting a frequency $\geq 1\%$ are displayed.^bWild type allele for the *BARD1* SNP 5 (c.1075del21) is represented by the letter T in haplotypes whereas variant allele for this SNP is not present in haplotypes exhibiting a frequency $\geq 1\%$.^cThese tSNPs haplotypes both tag simultaneously a common and a rare *AURKA* haplotype estimated using all SNPs.

have to keep in mind that the greater extent of LD generally found in the French Canadian population can be responsible for a lower number of haplotype blocks.⁵⁶ In addition such a difference between HapMap and this study may be due to the higher coverage of the gene regions in the HapMap data set which include both exonic and intronic SNPs, the latter not assessed in our study.

The analysis of the coding region of the *AURKA*, *BAP1*, *BARD1* and *DHX9* genes did not identify any deleterious truncating mutation or aberrant splicing mRNAs. We report here a possible association of the *BARD1* c.1518C>T variant with an increased risk of breast cancer. Further studies on much bigger cohorts are however needed to fully evaluate the association of this variant with breast cancer risk and to assess the functional contribution of this variant on mRNA stability and protein expression. Haplotype analysis, LD block and tSNPs identification were also conducted for all four genes, and permitted to state that haplotypes displaying a frequency $\geq 1\%$ represent more than 94% of all haplotypes estimated for each gene in our population. This first haplotype-based analysis to establish tSNPs for both the *BAP1* and *DHX9* genes will provide a valuable tool for further large association studies.

ACKNOWLEDGEMENTS

We are indebted to the participants and their families for their generosity and providing DNA samples. We thank Dr Martine Dumont, Gilles Leblanc, Carolle

Figure 3 Haplotype blocks and tSNPs identified for the *AURKA* (a), *BAP1* (b), *BARD1* (c) and *DHX9* (d) genes. Tagging SNPs identified on a block-by-block basis using haplotypes showing a frequency higher than 1% are denoted with an asterisk (*). Haplotype frequencies are displayed on the right of each haplotype combination whereas the level of recombination is displayed above the connections between two blocks. Thin connections represent haplotypes with frequencies between 1 and 10%, whereas haplotypes with frequencies higher than 10% are represented by thick lines.

Samson and Martine Tranchant for sample management, mutation screening and skillful technical assistance as well as Claire Brousseau, Marie-Andrée Lajoie, Pascale Léger, Hélène Malouin and Josée Rhéaume, for genetic counseling and clinical data management. We also thank Anne-Marie Moisan and Lucie Larouche for MLPA analyses. We thank Damian Labuda and Claudia Moreau at the Centre de Recherche de l'Hôpital Ste-Justine for providing control DNA samples. We thank Professor Bartha Maria Knoppers and her colleagues from the Centre de recherche en droit public de l'Université de Montréal for their precious help with ELSI issues related to our research program. We also appreciate advice received from ethics committees. This work was supported by the Canadian Institutes of Health Research (CIHR) through the INHERIT BRCA research program, the Fonds de la Recherche en Santé du Québec (FRSQ)/Réseau de Médecine Génétique Appliquée (RMGA), the CURE Foundation and the Canadian Breast Cancer Research Alliance (CBCRA). FG holds a studentship from Fondation René Bussièrès, CJB holds a Frederick Banting and Charles Best Canada Graduate Scholarships—Master's Award studentship from CIHR, JS is Chairholder of the Canada Research Chair in Oncogenetics, and FD is a recipient of a chercheur-boursier from the Fonds de la Recherche en Santé du Québec (FRSQ) and a Research Career Award in the Health Sciences from CIHR/Rx&D Health Research Foundation.

- 1 Bradbury, A. R. & Olopade, O. I. Genetic susceptibility to breast cancer. *Rev. Endocr. Metab. Disord.* **8**, 255–267 (2007).
- 2 Easton, D. F. How many more breast cancer predisposition genes are there? *Breast Cancer Res.* **1**, 14–17 (1999).
- 3 Stratton, M. R. & Rahman, N. The emerging landscape of breast cancer susceptibility. *Nat. Genet.* **40**, 17–22 (2008).
- 4 Simard, J., Dumont, M., Moisan, A. M., Gaborieau, V., Malouin, H., Durocher, F. *et al.* Evaluation of BRCA1 and BRCA2 mutation prevalence, risk prediction models and a multistep testing approach in French-Canadian families with high risk of breast and ovarian cancer. *J. Med. Genet.* **44**, 107–121 (2007).
- 5 Durocher, F., Labrie, Y., Soucy, P., Sinilnikova, O., Labuda, D., Bessette, P. *et al.* Mutation analysis and characterization of ATR sequence variants in breast cancer cases from high-risk French Canadian breast/ovarian cancer families. *BMC Cancer* **6**, 230 (2006).
- 6 Durocher, F., Labrie, Y., Ouellette, G., INHERIT BRCA & Simard, J. Genetic sequence variations and ADPRT haplotype analysis in French Canadian families with high risk of breast cancer. *J. Hum. Genet.* **52**, 963–977 (2007).
- 7 Guénard, F., Labrie, Y., Ouellette, G., Beauparlant, C. J., Bessette, P., Chiquette, J. *et al.* Germline mutations in the breast cancer susceptibility gene PTEN are rare in high-risk non-BRCA1/2 French Canadian breast cancer families. *Fam. Cancer* **6**, 483–490 (2007).
- 8 Guénard, F., Labrie, Y., Ouellette, G., Joly Beauparlant, C., Simard, J., INHERIT BRCA & Durocher, F. Mutational analysis of the breast cancer susceptibility gene BRIP1/BACH1/FANCI in high-risk non-BRCA1/BRCA2 breast cancer families. *J. Hum. Genet.* **53**, 579–591 (2008).
- 9 Desjardins, S., Belleau, P., Labrie, Y., Ouellette, G., Bessette, P., Chiquette, J. *et al.* Genetic variants and haplotype analyses of the ZBRK1/ZNF350 gene in high-risk non BRCA1/2 French Canadian breast and ovarian cancer families. *Int. J. Cancer* **122**, 108–116 (2008a).
- 10 Desjardins, S., Ouellette, G., Labrie, Y., Simard, J., INHERIT BRCA & Durocher, F. Analysis of GADD45A sequence variations in French Canadian families with high risk of breast cancer. *J. Hum. Genet.* **53**, 490–498 (2008b).
- 11 Scully, R., Anderson, S. F., Chao, D. M., Wei, W., Ye, L., Young, R. A. *et al.* BRCA1 is a component of the RNA polymerase II holoenzyme. *Proc. Natl. Acad. Sci. USA* **94**, 5605–5610 (1997).
- 12 Durocher, F., Guénard, F., Desjardins, S., Ouellette, G. & Labrie, Y. Inherited susceptibility to breast cancer: accomplishments and challenges. In: Sinnott D (ed). *Molecular genetics of cancer* Research Signpost, Kerala, pp 19–93 (2005).
- 13 Antoniou, A. C. & Easton, D. Polygenic inheritance of breast cancer: Implications for design of association studies. *Genet. Epidemiol.* **25**, 190–202 (2003).
- 14 Zhou, H., Kuang, J., Zhong, L., Kuo, W. L., Gray, J. W., Sahin, A. *et al.* Tumour amplified kinase STK15/BTAK induces centrosome amplification, aneuploidy and transformation. *Nat. Genet.* **20**, 189–193 (1998).
- 15 Dutertre, S., Descamps, S. & Prigent, C. On the role of aurora-A in centrosome function. *Oncogene* **21**, 6175–6183 (2002).
- 16 Ouchi, M., Fujiuchi, N., Sasai, K., Katayama, H., Minamishima, Y. A., Ongusaha, P. P. *et al.* BRCA1 phosphorylation by Aurora-A in the regulation of G2 to M transition. *J. Biol. Chem.* **279**, 19643–19648 (2004).
- 17 Ewart-Toland, A., Briassoulis, P., de Koning, J. P., Mao, J. H., Yuan, J., Chan, F. *et al.* Identification of STK6/STK15 as a candidate low-penetrance tumor-susceptibility gene in mouse and human. *Nat. Genet.* **34**, 403–412 (2003).
- 18 Dai, Q., Cai, Q. Y., Shu, X. O., Ewart-Toland, A., Wen, W. Q., Balmain, A. *et al.* Synergistic effects of STK15 gene polymorphisms and endogenous estrogen exposure in the risk of breast cancer. *Cancer Epidemiol. Biomarkers Prev.* **13**, 2065–2070 (2004).
- 19 Egan, K. M., Newcomb, P. A., Ambrosone, C. B., Trentham-Dietz, A., Titus-Ernstoff, L., Hampton, J. M. *et al.* STK15 polymorphism and breast cancer risk in a population-based study. *Carcinogenesis* **25**(11):2149–2153. *Erratum in: Carcinogenesis* **26**, 701 (2004).
- 20 Sun, T., Miao, X., Wang, J., Tan, W., Zhou, Y., Yu, C. *et al.* Functional Phe31Ile polymorphism in Aurora A and risk of breast carcinoma. *Carcinogenesis* **25**, 2225–2230 (2004).
- 21 Ewart-Toland, A., Dai, Q., Gao, Y. T., Nagase, H., Dunlop, M. G., Farrington, S. M. *et al.* Aurora-A/STK15 T+91A is a general low penetrance cancer susceptibility gene: a meta-analysis of multiple cancer types. *Carcinogenesis* **26**, 1368–1373 (2005).
- 22 Lo, Y. L., Yu, J. C., Chen, S. T., Yang, H. C., Fann, C. S., Mau, Y. C. *et al.* Breast cancer risk associated with genotypic polymorphism of the mitosis-regulating gene Aurora-A/STK15/BTAK. *Int. J. Cancer* **115**, 276–283 (2005).
- 23 Cox, D. G., Hankinson, S. E. & Hunter, D. J. Polymorphisms of the AURKA (STK15/Aurora Kinase) Gene and Breast Cancer Risk (United States). *Cancer Causes Control.* **17**, 81–83 (2006).
- 24 Fletcher, O., Johnson, N., Palles, C., dos Santos Silva, I., McCormack, V., Whittaker, J. *et al.* Inconsistent association between the STK15 F31I genetic polymorphism and breast cancer risk. *J. Natl. Cancer Inst.* **98**, 1014–1018 (2006).
- 25 Couch, F. J., Sinilnikova, O., Vierkant, R. A., Pankratz, V. S., Fredericksen, Z. S., Stoppa-Lyonnet, D. *et al.* AURKA F31I polymorphism and breast cancer risk in BRCA1 and BRCA2 mutation carriers: a consortium of investigators of modifiers of BRCA1/2 study. *Cancer Epidemiol. Biomarkers Prev.* **16**, 1416–1421 (2007).
- 26 Tchatchou, S., Wirtenberger, M., Hemminki, K., Sutter, C., Meindl, A., Wappenschmidt, B. *et al.* Aurora kinases A and B and familial breast cancer risk. *Cancer Lett.* **247**, 266–272 (2007).
- 27 Vidarsdottir, L., Bodvarsdottir, S. K., Hilmarsdottir, H., Tryggvadottir, L. & Eyfjord, J. E. Breast cancer risk associated with AURKA 91T → A polymorphism in relation to BRCA mutations. *Cancer Lett.* **250**, 206–212 (2007).
- 28 Jensen, D. E., Proctor, M., Marquis, S. T., Gardner, H. P., Ha, S. I., Chodosh, L. A. *et al.* BAP1: a novel ubiquitin hydrolase which binds to the BRCA1 RING finger and enhances BRCA1-mediated cell growth suppression. *Oncogene* **16**, 1097–1112 (1998).
- 29 Jensen, D. E. & Rauscher, F. J. III. BAP1, a candidate tumor suppressor protein that interacts with BRCA1. *Ann. NY Acad. Sci.* **886**, 191–194 (1999).
- 30 Hu, J. J., Mohrenweiser, H. W., Bell, D. A., Leadon, S. A. & Miller, M. S. Symposium overview: genetic polymorphisms in DNA repair and cancer risk. *Toxicol. Appl. Pharmacol.* **185**, 64–73 (2002).
- 31 Coupié, I., Cousin, P. Y., Hughes, D., Legoix-Ne, P., Trehin, A., Sinilnikova, O. M. *et al.* BAP1 and breast cancer risk. *Fam. Cancer* **4**, 273–277 (2005).
- 32 Hashizume, R., Fukuda, M., Maeda, I., Nishikawa, H., Oyake, D., Yabuki, Y. *et al.* The RING heterodimer BRCA1-BARD1 is a ubiquitin ligase inactivated by a breast cancer-derived mutation. *J. Biol. Chem.* **276**, 14537–14540 (2001).
- 33 Chen, A., Kleiman, F. E., Manley, J. L., Ouchi, T. & Pan, Z. Q. Autoubiquitination of the BRCA1-BARD1 RING ubiquitin ligase. *J. Biol. Chem.* **277**, 22085–22092 (2002).
- 34 Starita, L. M., Horwitz, A. A., Keogh, M. C., Ishioka, C., Parvin, J. D. & Chiba, N. BRCA1/BARD1 ubiquitinate phosphorylated RNA polymerase II. *J. Biol. Chem.* **280**, 24498–24505 (2005).
- 35 Anderson, S. F., Schlegel, B. P., Nakajima, T., Wolpin, E. S. & Parvin, J. D. BRCA1 protein is linked to the RNA polymerase II holoenzyme complex via RNA helicase A. *Nat. Genet.* **19**, 254–256 (1998).
- 36 Thai, T. H., Du, F., Tsan, J. T., Jin, Y., Phung, A., Spillman, M. A. *et al.* Mutations in the BRCA1-associated RING domain (BARD1) gene in primary breast, ovarian and uterine cancers. *Hum. Mol. Genet.* **7**, 195–202 (1998).
- 37 Ghimenti, C., Sensi, E., Prescittini, S., Brunetti, I. M., Conte, P., Bevilacqua, G. *et al.* Germline mutations of the BRCA1-associated ring domain (BARD1) gene in breast and breast/ovarian families negative for BRCA1 and BRCA2 alterations. *Genes Chromosomes Cancer* **33**, 235–242 (2002).
- 38 Karppinen, S. M., Heikkinen, K., Rapakko, K. & Winqvist, R. Mutation screening of the BARD1 gene: evidence for involvement of the Cys557Ser allele in hereditary susceptibility to breast cancer. *J. Med. Genet.* **41**, e114 (2004).
- 39 Stacey, S. N., Sulem, P., Johannsson, O. T., Helgason, A., Gudmundsson, J., Kistic, J. P. *et al.* The BARD1 Cys557Ser variant and breast cancer risk in Iceland. *PLoS Med.* **3**, e217 (2006).
- 40 Wu, J. Y., Vlastos, A. T., Pelte, M. F., Caligo, M. A., Bianco, A., Krause, K. H. *et al.* Aberrant expression of BARD1 in breast and ovarian cancers with poor prognosis. *Int. J. Cancer* **118**, 1215–1226 (2006).
- 41 Ishitobi, M., Miyoshi, Y., Hasegawa, S., Egawa, C., Tamaki, Y., Monden, M. *et al.* Mutational analysis of BARD1 in familial breast cancer patients in Japan. *Cancer Lett.* **200**, 1–7 (2003).
- 42 Karppinen, S. M., Barkardottir, R. B., Backenhorn, K., Sydenham, T., Syrjakoski, K., Schleutker, J. *et al.* Nordic collaborative study of the BARD1 Cys557Ser allele in 3956 patients with cancer: enrichment in familial BRCA1/BRCA2 mutation-negative breast cancer but not in other malignancies. *J. Med. Genet.* **43**, 856–862 (2006).
- 43 Jakubowska, A., Cybulski, C., Szymanska, A., Huzarski, T., Byrski, T., Gronwald, J. *et al.* BARD1 and breast cancer in Poland. *Breast Cancer Res. Treat.* **107**, 119–122 (2008).
- 44 Pharoah, P. D., Tyrer, J., Dunning, A. M., Easton, D. F., Ponder, B. A. & SEARCH Investigators. Association between common variation in 120 candidate genes and breast cancer risk. *PLoS Genet.* **3**, e42 (2007).
- 45 Huo, X., Hu, Z., Zhai, X., Wang, Y., Wang, S., Wang, X. *et al.* Common non-synonymous polymorphisms in the BRCA1 associated RING Domain (BARD1) gene are associated with breast cancer susceptibility: a case-control analysis. *Breast Cancer Res. Treat.* **102**, 329–337 (2007).
- 46 Lee, C. G. & Hurwitz, J. A new RNA helicase isolated from HeLa cells that catalytically translocates in the 3' to 5' direction. *J. Biol. Chem.* **267**, 4398–4407 (1992).
- 47 Nakajima, T., Uchida, C., Anderson, S. F., Lee, C. G., Hurwitz, J., Parvin, J. D. *et al.* RNA helicase A mediates association of CBP with RNA polymerase II. *Cell* **90**, 1107–1112 (1997).
- 48 Srebrow, A. & Kornbliht, A. R. The connection between splicing and cancer. *J. Cell. Sci.* **119** (Part 13), 2635–2641 (2006).
- 49 Dicioccio, R. A., Song, H., Waterfall, C., Kimura, M. T., Nagase, H., McGuire, V. *et al.* STK15 polymorphisms and association with risk of invasive ovarian cancer. *Cancer Epidemiol. Biomarkers Prev.* **13**, 1589–1594 (2004).
- 50 Goringe, K. L., Choong, D. Y., Visvader, J. E., Lindeman, G. J. & Campbell, I. G. BARD1 variants are not associated with breast cancer risk in Australian familial breast cancer. *Breast Cancer Res. Treat.* **111**, 505–509 (2008).
- 51 Fortin, J., Moisan, A. M., Dumont, M., Leblanc, G., Labrie, Y., Durocher, F. *et al.* A new alternative splice variant of BRCA1 containing an additional in-frame exon. *Biochim. Biophys. Acta* **1731**, 57–65 (2005).
- 52 Lombardi, G., Falaschi, E., Di Cristofano, C., Naccarato, A. G., Sensi, E., Aretini, P. *et al.* Identification of novel alternatively spliced BRCA1-associated RING domain (BARD1) messenger RNAs in human peripheral blood lymphocytes and in sporadic breast cancer tissues. *Genes Chromosomes Cancer* **46**, 791–795 (2007).
- 53 Tsuzuki, M., Wu, W., Nishikawa, H., Hayami, R., Oyake, D., Yabuki, Y. *et al.* A truncated splice variant of human BARD1 that lacks the RING finger and ankyrin repeats. *Cancer Lett.* **233**, 108–116 (2006).
- 54 Shin, S. O., Lee, K. H., Kim, J. H., Baek, S. H., Park, J. W., Gabrielson, E. W. *et al.* Alternative splicing in 5'-untranslational region of STK-15 gene, encoding centrosome associated kinase, in breast cancer cell lines. *Exp. Mol. Med.* **32**, 193–196 (2000).

- 55 Devlin, B. & Risch, N. A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics* **29**, 311–322 (1995).
- 56 Vézina, H., Durocher, F., Dumont, M., Houde, L., Szabo, C., Tranchant, M. *et al*. Molecular and genealogical characterization of the R1443X BRCA1 mutation in high-risk French-Canadian breast/ovarian cancer families. *Hum. Genet.* **117**, 119–132 (2005).

APPENDIX

Clinicians and other members (*) of INHERIT BRCAs involved in this study

Paul Bessette: Service de gynécologie, Centre Hospitalier Universitaire de Sherbrooke, Fleurimont, QC, J1H 5N4, Canada; Peter Bridge: Molecular Diagnostic Laboratory, Alberta Children's Hospital,

Calgary, AB, T2T 5C7, Canada; Jocelyne Chiquette: Clinique des maladies du sein Deschênes-Fabia, Hôpital du Saint-Sacrement, Québec, QC, G1S 4L8, Canada; Rachel Laframboise: Service de médecine génétique, CHUQ, Pavillon CHUL, Québec, QC, G1V 4G2, Canada; Jean Lépine: Centre Hospitalier régional de Rimouski, Rimouski, QC, G5L 5T1, Canada; Bernard Lespérance, Roxane Pichette: Service d'hémo-oncologie, Hôpital du Sacré-Cœur, Montréal, QC, H4J 1C5, Canada; Marie Plante: Service de gynécologie, CHUQ, L'Hôtel-Dieu de Québec, Québec, QC, G1R 2J6, Canada; *Jacques Simard, Canada Research Chair in Oncogenetics: Cancer Genomics Laboratory, Oncology and Molecular Endocrinology Research Centre, CHUQ, Pavillon CHUL, Québec, QC G1V 4G2, Canada.

Supplementary Information accompanies the paper on Journal of Human Genetics website (<http://www.nature.com/jhg>)