Mutational analysis of the breast cancer susceptibility gene *BRIP1/BACH1/FANCF* in high-risk non-*BRCA1/BRCA2* breast cancer families

Frédéric Guénard · Yvan Labrie · Geneviève Ouellette ·
Charles Joly Beauparlant · Jacques Simard ·
INHERIT BRCAs · Francine Durocher

Received: 11 October 2007 / Accepted: 16 March 2008 / Published online: 15 April 2008
© The Japan Society of Human Genetics and Springer 2008

Abstract The *BRIP1* gene encodes a helicase interacting with *BRCA1*, which contributes to *BRCA1*-associated DNA repair function. Germ-line *BRIP1* mutations affecting the helicase domain activity have been identified in early onset breast cancer patients. In addition, *BRIP1* was recently identified as deficient in Fanconi anemia (FA) complementation group J. Given the growing evidence now linking *BRCA1*, *BRCA2*, and the FA pathway, as well as the involvement of FA proteins (*BRCA2/FANCD1* and *PALB2/FANCF*) in breast cancer susceptibility, we sought to evaluate the contribution of *FANCF* gene alterations regarding breast cancer susceptibility among our cohort of 96 breast cancer individuals from high-risk non-*BRCA1/2* French Canadian families. No deleterious mutation, exon deletion, or retention of intronic portions could be identified. However, extensive analysis of the promoter and whole exonic and flanking intronic regions of *FANCF* led to the identification of 42 variants, including 22 novel variants not previously reported, four of which were located in the promoter region. Transcription factors analysis revealed a potential involvement of *FANCF* promoter

variants in regulation of *FANCF* expression, and reporter gene assays were performed. The allelic frequency was assessed in a cohort of 73 unaffected French Canadian individuals, and haplotype analysis and tagging single nucleotide polymorphism (SNP) identification were also performed. Although our study unlikely involves *FANCF* as a high-risk predisposition gene in non-*BRCA1/2* high-risk French Canadian families, the possible association of *FANCF* missense variants with phenotypes associated with FA, such as childhood cancer, cannot be excluded.

Keywords *BRIP1/FANCF* gene · Breast cancer · Sequence analysis · Haplotype · Tagging SNP

Introduction

In developed countries, breast cancer incidence is estimated as 7.8% at 80 years of age (Collaborative Group on Hormonal Factors in Breast Cancer 2001). Mutations in the two most important breast cancer susceptibility genes, *BRCA1* and *BRCA2*, account for only 25% of families with hereditary breast cancer (Easton 1999; Ponder 2001). A lower proportion of hereditary breast cancer is explained by mutations in *TP53* (Laloo et al. 2003; Gasco et al. 2003), *PTEN* (Ueda et al. 1998; Ball et al. 2001; Guenard et al. 2007), *STK11* (Giardiello et al. 2000), *ATM* (Stankovic et al. 1998; Chenevix-Trench et al. 2002; Szabo et al. 2004; Renwick et al. 2006), and *CHEK2* (*CHEK2* Breast Cancer Case-Control Consortium 2004) genes, and more recently, specific *FANCF* genes have been associated with breast cancer susceptibility, namely, *FANCD1* subsequently identified as *BRCA2*, *FANCF/PALB2*, and *FANCF/BRIP1* (Durocher et al. 2005; Seal et al. 2006; Rahman et al. 2007; Erkkö et al. 2007; Tischkowitz et al. 2007).

Jacques Simard holds Canada Research Chair in Oncogenetics. Other members of INHERIT BRCAs involved in clinical aspects of this study are listed in the Appendix.

Electronic supplementary material The online version of this article (doi:10.1007/s10038-008-0285-z) contains supplementary material, which is available to authorized users.

F. Guénard · Y. Labrie · G. Ouellette · C. Joly Beauparlant ·
J. Simard · F. Durocher (✉)
Cancer Genomics Laboratory, Oncology and Molecular
Endocrinology Research Centre, Centre Hospitalier Universitaire
de Québec and Laval University, 2705 Laurier Boulevard,
Québec city, QC, Canada G1V 4G2
e-mail: francine.durocher@crchul.ulaval.ca

FANCI, also called *BRIP1/BACH1*, has been identified as deficient in Fanconi anemia (FA) complementation group J (Levitus et al. 2005). Many of the proteins encoded by FA genes, including the more recently identified *FANCI* (Smogorzewska et al. 2007), interact directly with each other to form a multisubunit nuclear complex and are involved in DNA repair processes (Medhurst et al. 2001). An increased risk to develop cancer has been reported for homozygous FA patients (Alter et al. 2003).

The *BRIP1/BACH1/FANCI* gene is located at 17q22–q24 and encodes a helicase in which the C-terminal part is known to interact with *BRCA1* via its *BRCA1* C-terminal (BRCT) repeats (Cantor et al. 2001). The corresponding protein contributes to *BRCA1*-associated DNA repair function, and mutation of an essential catalytic residue (Pro47) found in most members of the DEAH family interferes with normal double-strand break (DSB) repair in a manner that was dependent on *FANCI* binding to *BRCA1* (Cantor et al. 2001). In addition, phosphorylation of the Ser990 residue triggers *FANCI* interaction with BRCT domains of *BRCA1*, which is required for establishment of the G2 cell-cycle checkpoint response to DNA damage (Yu et al. 2003). Proper localization of *FANCI* to the nucleus is impaired by the c.517C > T variant (p.Arg173Cys) (Lei et al. 2003), which has been suggested to have an effect on breast cancer susceptibility. Furthermore, enzymatic activity of *FANCI* was found defective in two patients with germline *FANCI* coding sequence mutations who experienced early onset breast cancer (Cantor et al. 2004). Germline mutations affecting domain activity or messenger ribonucleic acid (mRNA) expression were also identified in early onset breast cancer patients, suggesting an implication for *FANCI* in breast cancer susceptibility (Cantor et al. 2001, 2004). Given the close relationship between *BRCA1* and *FANCI* and the observation that other helicases belonging to the RecQ family are implicated in cancer-predisposing syndromes, namely, the Bloom (*BLM* gene), the Werner (*WRN* gene), and the Rothmund-Thomson syndromes (*RecQL4* gene) (Ellis et al. 1995; Yu et al. 1996; Kitao et al. 1998, 1999), this reinforces the rational basis for a potential implication of *FANCI* in breast cancer predisposition.

In recent years, a number of studies have been conducted to evaluate the possible implication of the *FANCI* gene with regard to breast cancer susceptibility (Rutter et al. 2003; Sigurdson et al. 2004; Lewis et al. 2005; Garcia-Closas et al. 2006; Vahteristo et al. 2006; Seal et al. 2006; Song et al. 2007; Frank et al. 2007). However, the great majority was performed either on specific variants (such as P919S and –64G > A) (Sigurdson et al. 2004; Frank et al. 2007), on breast cancer cases not selected on the basis of a family history (Garcia-Closas et al. 2006; Song et al. 2007), or on a limited number of

cases from *BRCA1/2*-negative breast cancer families (Rutter et al. 2003; Vahteristo et al. 2006). Recently published results have clearly shown that inactivating truncating mutations of *FANCI* confer susceptibility to breast cancer in high-risk non-*BRCA1/BRCA2* families (Seal et al. 2006).

In this regard, given that deleterious mutations in *BRCA1/BRCA2* were identified in only 24% of 256 high-risk French Canadian breast/ovarian cancer families (Simard et al. 2007), and that germline mutations in other genes are rare and do not seem to contribute substantially to breast cancer susceptibility in high-risk French Canadian breast cancer families (Durocher et al. 2006, 2007; Guenard et al. 2007; Desjardins et al. 2008), evaluation of the *FANCI* contribution to breast cancer in these families is imperative. Therefore, the complete coding, promoter, and flanking intronic regions of *FANCI* were analyzed in 96 breast cancer cases drawn from nonrelated *BRCA1/BRCA2*-negative French Canadian families with a high risk of breast cancer.

Materials and methods

Ascertainment of families

Recruitment of high-risk French Canadian breast and/or ovarian cancer families from Canada was part of a large ongoing interdisciplinary research program Interdisciplinary Health Research International Team on Breast Cancer (INHERIT BRCA). More details regarding ascertainment criteria, experimental procedures, and the INHERIT BRCA research program are described elsewhere (Vezina et al. 2005; Antoniou et al. 2006; Simard et al. 2007). Briefly, all patients were referred by physicians from Québec province, and they were also responsible for *BRCA1/2* test result disclosure to participants. Ethics committees approved the study, and patients signed informed consent. Patients were screened for mutations and large genomic rearrangements of the *BRCA1* and *BRCA2* genes (Moisan et al. 2006). Subsequently, another component was designed for the “localization and identification of new breast cancer susceptibility loci/genes.” Ethics approval for this latter study was also obtained from the different institutions participating in this research project, and each participant knowing their inconclusive *BRCA1/2* test results status had to sign a specific informed consent for their participation in this component. All participants had to be at least 18 years of age and mentally capable. A subset of 96 high-risk French Canadian breast/ovarian cancer families were recruited for this study according to the ascertainment criteria described previously (Durocher et al. 2006, 2007; Guenard et al. 2007;

Desjardins et al. 2008). The mean age at diagnosis of these 96 individuals affected with breast cancer was 48.5 years (32–74). Genomic DNA extraction of the 96 French Canadian breast cancer cases as well as 73 healthy unrelated French Canadian individuals was performed as previously described (Durocher et al. 2006, 2007; Guenard et al. 2007; Desjardins et al. 2008). These control individual samples were recruited on a nonnominative basis as part of a long-term study aimed at characterizing the genetic variability in human populations approved by Institutional Ethics Review Board.

DNA/RNA isolation from immortalized cell lines and cDNA synthesis

Lymphocytes were isolated and immortalized from blood samples using the Epstein-Barr virus (EBV) in 15% Roswell Park Memorial Institute (RPMI) media, as previously described (Boukamp et al. 1990, 1998). Total RNA was extracted from EBV-transformed β -lymphoblastoid cell lines using TRI REAGENT® (Molecular Research Center Inc., Cincinnati, OH, USA) according to the manufacturer's instructions. Following RNA extraction, reverse transcription of 5 μ g of RNA was performed as previously described (Durocher et al. 2006). Genomic DNA from 96 high-risk French Canadian breast cancer individuals and 73 controls was extracted from peripheral blood using standard methods, either phenol–chloroform, Gentra kits (Minneapolis, MN, USA), or the QIAamp DNA Blood MAXI Kit (QIAGEN Inc., Mississauga, ON, Canada) following supplier instructions. Genomic DNA from immortalized cell lines was extracted using the cetyl trimethylammonium bromide (CTAB) method (Sambrook et al. 2001).

Polymerase chain reaction amplification, sequence analysis, and variant characterization

Amplification and direct sequencing of complementary DNA (cDNA) and genomic DNA of the *FANCF* gene were conducted using primers selected with the Primer Express 2.0 software (Applied Biosystems, Foster City, CA, USA). These primers, listed in Supplemental Table 1, were chosen to amplify exonic, promoter, and flanking intronic regions of the *FANCF* gene. A fluorescence-based direct sequencing method using BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) was performed to analyze both cDNA and genomic sequences. Sequence reactions were run on an Applied Biosystems 3730xl DNA analyzer, and sequence data were analyzed using the Staden Pregap4 and Gap4 programs (Bonfield et al. 1998). Each single nucleotide polymorphism (SNP) was tested for

departure from Hardy–Weinberg equilibrium (HWE) by means of a χ^2 test. All *P* values were two-sided, with 1 degree of freedom. Allelic distribution in both series was tested using a χ^2 test. *P* values <0.05 were considered significant.

Computational analysis

To identify potential effects of variants located in the promoter region, a transcription factor search was performed using the program MatInspector from the GenomatixSuite (Quandt et al. 1995; Cartharius et al. 2005). The Exonic Splicing Enhancer (ESE) finder Web-based program (Cartegni et al. 2003) was used to determine the potential effect of variants found in the coding region of the *FANCF* gene on pre-mRNA splicing. All tests were run under default threshold values. The NNSPLICE 0.9 program (Reese et al. 1997) was used to evaluate the potential effect of intronic variants identified in this study on the strength of potential splice sites leading to the splicing of pre-mRNA, whereas the SIFT and PolyPhen Web-based softwares were used to predict the effect of amino-acid substitution on protein structure.

Reporter gene assay

FANCF promoter constructs were amplified from genomic DNA using either a 1,068-bp (long) or a 633-bp (short) fragment with primers introducing 5' *KpnI* and 3' *XhoI* sites (Supplemental Table 2), and these amplicons were thereafter subcloned into pGL3 basic vector (Promega, Madison, WI). Variant constructs were produced by site-directed mutagenesis (QuikChange Site-Directed Mutagenesis Kit; Stratagene, La Jolla, CA) or obtained by polymerase chain reaction (PCR) amplification and subcloning. Transient transfection assays of MCF-7 cells and dual-luciferase reporter assays were conducted as previously described (Duguay et al. 2004). Cells were harvested 48 h after transfection, which was performed in triplicate. Reporter-gene activity was corrected both by subtracting background signal and by normalizing firefly activity against renilla activity obtained from the same sample. Results are expressed as mean $\pm$ standard deviation (SD) for each triplicate, and luciferase activity was compared between the variant and its respective wild-type (WT) construct, with a two-tailed unpaired Student *t* test. *P* values < 0.05 were considered significant.

Linkage disequilibrium and haplotype analyses

The linkage disequilibrium analysis (LDA) program (Ding et al. 2003) was used to calculate pairwise linkage

disequilibrium (LD; Lewontin's $|D'|$ and r^2 values) for each SNP pair identified in our breast cancer series (Lewontin 1964; Devlin and Risch 1995). The PHASE 2.1.1 software (Stephens et al. 2001; Stephens and Donnelly 2003) was used to estimate haplotype frequencies. SNPs having a minor allele frequency (MAF) $\geq 5\%$ identified in both sample sets were used to perform this haplotype analysis. The algorithm was run five times, with a minimum of 100 permutations, under default conditions. Genotyping data from both series and from the Centre d'Etude du Polymorphisme Humain (CEPH) cohort (HapMap database) were used to identify LD blocks within the *FANCI* gene using the Haploview software. Identification of tagging SNPs (tSNPs) within each LD block was conducted using the same software.

Electronic databases

ESE finder: http://rulai.cshl.edu/cgi-bin/tools/ESE3/ese_finder.cgi?process=home

Haploview: <http://www.broad.mit.edu/mpg/haploview>

HapMap: <http://www.hapmap.org>

Matinspector: <http://www.genomatix.de/matinspector.html>

NCBI: <http://www.ncbi.nlm.nih.gov>

PHASE: <http://www.stat.washington.edu/stephens/software.html>

PolyPhen: <http://genetics.bwh.harvard.edu/pph/>

SIFT: <http://blocks.fhcrc.org/sift/SIFT.html>

Splice Site Prediction Program using Neural Networks (NNSPLICE; SSPNN): http://www.fruitfly.org/seq_tools/splice.html

TFSEARCH: <http://www.cbrc.jp/research/db/TFSEARCH.html>

Transcription Element Search System (TESS): <http://www.cbil.upenn.edu/tess>

UCSC Genome Bioinformatics: <http://genome.ucsc.edu/>

Results

FANCI sequence analysis

Analysis of the promoter, whole-exonic, and intronic flanking sequences of the *FANCI* gene in our 96 breast cancer individuals led to the identification of 42 variants, of which 17 displayed an MAF $\geq 5\%$ (Table 1). These sequence variations were composed of 37 nucleotide substitutions, three deletions, and two insertions. Among these 42 variants, 13 were exonic, five and 24 were located in promoter and intronic regions, respectively, and 22 were novel, with the remaining 20 being reported in the SNP database (dbSNP build 127). Analysis of the promoter region, consisting of approximately 1,400 bp upstream of

exon 1, led to identification of five additional variants, including four not reported in databases.

Sequence variations displaying an MAF $\geq 5\%$ in breast cancer cases were further assessed in a series of 73 unaffected individuals also of French Canadian origin. Rare changes (MAF $< 5\%$) located in the proximal region of these common variants (MAF $\geq 5\%$) were also analyzed, yielding a total of 38 nucleotide variants characterized in both sample sets. Of these, 14 were not observed in control individuals. Corresponding frequencies are denoted in Table 1. No deviation from HWE was observed for any of the nucleotide changes identified, nor was there a significant difference in terms of allelic frequency observed between both series based on single-marker analysis. The segregation of the variants identified was examined in families for which DNA material was available from multiple individuals (34 of the 42 variants), but no clear segregation between the identified variants and breast cancer was observed (data not shown).

Conservation of human *FANCI* residues

Among the five amino-acid substitutions found in our breast cancer cases, the common c.2755T > C variant (p.Pro919Ser) is located within the BRCA1-binding domain (Cantor et al. 2001) and the four remaining variants (c.415T > G, c.517C > T, c.577G > A, c.823A > G) are situated within the Rad3-related DNA helicase domain of *FANCI* (residues 1–888) (Cantor et al. 2001). As illustrated in Table 2, comparison of these amino-acid changes was performed across relevant species using data extracted from the University of California Santa Cruz (UCSC) database. Alignment of *FANCI* orthologue sequences revealed that p.Ser139, p.Arg173, and p.Ile275 are strongly conserved in higher species, with p.Ile275 also found in more distant species such as *Xenopus tropicalis* and *Tetraodon nigroviridis*. The p.Val193 residue is poorly conserved, whereas the p.Pro919 residue is highly variable and not conserved in other species. Taken together, the comparison of these orthologues indicates that three variants, namely, c.415T > G (p.Ser139Ala), c.517C > T (p.Arg173Cys), and c.823A > G (p.Ile275Val), are under strong functional constraint. Of these five amino-acid substitutions, only p.Arg173Cys is predicted to be damaging or not tolerated for protein folding. The remaining four, which include two unreported amino-acid changes, are considered to be tolerated or benign.

In silico analysis

Variants located in the promoter region were evaluated for their potential effect on transcription-factor binding sites (TFBS). As illustrated in Table 3, whereas two variants

Table 1 Observed coding and intronic sequence variants and genotype frequencies in familial breast cancer cases and controls

SNP	SNP ID ^a	dbSNP ID	Amino-acid change	Location	Cases (<i>n</i> = 96)			Controls				
					Common homozygote	Heterozygote	Rare homozygote	MAF	Common homozygote	Heterozygote	Rare homozygote	MAF (<i>n</i>)
1	c.-141 – 785delCTTTT	N/A	–	Promoter	94	2	0	0.010	62	0	0	0.000 (62)
2	c.-141 – 331T > C	N/A	–	Promoter	93	3	0	0.016	59	2	0	0.016 (61)
3	c.-141 – 280G > T	N/A	–	Promoter	95	1	0	0.005	61	0	0	0.000 (61)
4	c.-141 – 64G > A	rs2048718	–	Promoter	38	41	17	0.391	23	32	15	0.443 (70)
5	c.-141 – 55A > G	N/A	–	Promoter	95	1	0	0.005	70	0	0	0.000 (70)
6	c.-138T > C	N/A	–	5' UTR	95	1	0	0.005	70	0	0	0.000 (70)
7	c.-31 + 12G > A	rs4988340	–	Intron 1	52	40	4	0.250	36	23	10	0.312 (69)
8	c.-31 + 22G > C	rs4988341	–	Intron 1	95	1	0	0.005	68	1	0	0.007 (69)
9	c.205 + 162T > C	N/A	–	Intron 3	94	2	0	0.010	N/A	N/A	N/A	N/A
10	c.249A > G	N/A	p.Gln83Gln	Exon 4	95	1	0	0.005	72	0	0	0.000 (72)
11	c.379 + 76A > G	N/A	–	Intron 4	95	1	0	0.005	72	0	0	0.000 (72)
12	c.380 – 226insA	N/A	–	Intron 4	31	48	17	0.427	33	28	11	0.347 (72)
13	c.380 – 223A > T	N/A	–	Intron 4	95	1	0	0.005	72	0	0	0.000 (72)
14	c.380 – 28G > A	rs4988343	–	Intron 4	91	5	0	0.026	70	2	0	0.014 (72)
15	c.415T > G	N/A	p.Ser139Ala	Exon 5	95	1	0	0.005	71	0	0	0.000 (71)
16	c.508 – 281A > G	rs9908659	–	Intron 5	31	48	17	0.427	29	28	11	0.368 (68)
17	c.508 – 31C > G	rs4988344	–	Intron 5	75	18	3	0.125	53	15	0	0.110 (68)
18	c.517C > T	rs4988345	p.Arg173Cys	Exon 6	95	1	0	0.005	68	0	0	0.000 (68)
19	c.577G > A	rs4988346	p.Val193Ile	Exon 6	94	2	0	0.010	67	1	0	0.007 (68)
20	c.823A > G	N/A	p.Ile275Val	Exon 7	95	1	0	0.005	66	0	0	0.000 (66)
21	c.1140 + 11G > A	N/A	–	Intron 8	95	1	0	0.005	69	0	0	0.000 (69)
22	c.1140 + 91insT	rs11390869	–	Intron 8	3	33	60	0.797	2	26	41	0.783 (69)
23	c.1140 + 198C > G	N/A	–	Intron 8	35	49	12	0.380	27	31	9	0.366 (67)
24	c.1340 + 53G > A	N/A	–	Intron 9	95	1	0	0.005	72	0	0	0.000 (72)
25	c.1340 + 109G > A	rs2191248	–	Intron 9	36	47	13	0.380	32	31	9	0.340 (72)
26	c.1474 – 196delCCT	N/A	–	Intron 10	88	8	0	0.042	N/A	N/A	N/A	N/A
27	c.1628 + 99G > A	N/A	–	Intron 11	93	3	0	0.016	N/A	N/A	N/A	N/A
28	c.1628 + 163G > A	N/A	–	Intron 11	95	1	0	0.005	N/A	N/A	N/A	N/A
29	c.1795 – 47C > G	rs4988351	–	Intron 12	58	30	8	0.240	44	22	3	0.203 (69)
30	c.2097 + 7G > A	rs4988352	–	Intron 14	94	2	0	0.010	70	0	0	0.000 (70)
31	c.2379 + 204T > G	rs12937232	–	Intron 16	62	31	3	0.193	33	15	2	0.190 (50)
32	c.2379 + 218G > A	N/A	–	Intron 16	92	4	0	0.021	50	1	0	0.010 (51)
33	c.2379 + 306T > C	rs9904294	–	Intron 16	62	31	3	0.193	32	16	3	0.216 (51)

Table 1 continued

SNP	SNP ID ^a	dbSNP ID	Amino-acid change	Location	Cases (n = 96)			Controls			MAF (n)
					Common homozygote	Heterozygote	Rare homozygote	Common homozygote	Heterozygote	Rare homozygote	
34	c.2637G > A	rs4986765	p.Glu879Glu	Exon 19	36	49	11	35	29	7	0.303 (71)
35	c.2755T > C	rs4986764	p.Pro919Ser	Exon 19	18	48	30	12	32	26	0.600 (70)
36	c.2905 + 83T > A	rs4988357	–	Intron 19	45	40	11	27	29	11	0.375 (67)
37	c.2906 – 86G > A	N/A	–	Intron 19	95	1	0	61	0	0	0.000 (61)
38	c.3411C > T	rs4986763	p.Tyr1137Tyr	Exon 20	27	49	20	25	34	12	0.408 (71)
39	c.3750 + 128A > G	N/A	–	3' UTR	93	3	0	58	2	0	0.017 (60)
40	c.3750 + 158C > T	rs1978111	–	3' UTR	27	49	20	27	32	12	0.394 (71)
41	c.3750 + 396delCT	N/A	–	3' UTR	95	1	0	61	2	0	0.016 (63)
42	c.3750 + 483T > C	rs7213430	–	3' UTR	28	49	19	23	29	10	0.395 (62)

N/A Information not available for this single nucleotide polymorphism (SNP) or not reported in the SNP database (dbSNP), MAF minor allele frequency

^a According to the nomenclature of the Human Genome Variation Society

(c.–141 – 331T > C and c.–141 – 280G > T) seem to have a limited effect, the c.–141 – 785delCTTTT, c.–141 – 64G > A and c.–141 – 55A > G changes demonstrate a stronger potential impact on several promoter binding sites (BS), involving in particular androgen receptor (AR), myogenic factor 4 (MF4), KLF4, ras-responsive element binding 1 (RREB1), and aryl hydrocarbon receptor (AHR) BS.

Regarding the possible effect of exonic variants on ESE, analysis of the scores obtained revealed that certain exonic variants markedly alter the binding capacity of putative ESE elements, either by reducing their values or by creating new BS for SR proteins. As illustrated in Table 4, BS scores affecting binding of SC35, SRp55, and SRp40 proteins are mainly disturbed by four variants, namely, c.517C > T (SRp55), c.2637G > A (SC35), c.2755T > C (SC35, SRp40 and SRp55), and the 3'-UTR c.3750 + 396delCT variation, which seems to reduce the binding of the SC35 protein. All these variations could result in promotion or repression of splicing, or more particularly, on alternative splicing.

The potential effect of the 24 intronic variants on the strength of splice sites leading to the splicing of pre-mRNA was also evaluated by in silico analyses. The results obtained indicated that few sequence variations could affect pre-mRNA splicing, the score variations unlikely causing any marked changes in alternative splicing (data not shown). To further assess the possible effect of exonic and intronic variants on splicing at the mRNA level, cDNA material from our 96 high-risk breast cancer individuals was also analyzed. This analysis allowed us to confirm the presence of all exonic variants previously identified on genomic DNA with the exception of c.–138T > C, which could not be detected due to 5'-end localization of this variant on mRNA. In agreement with in silico studies, analysis of cDNA material revealed no aberrant splicing, even in carriers of variants that were suggestive to alter splicing. In addition, analysis of cDNA material from immortalized lymphocytes of a number of these breast cancer individuals treated with puromycin, an inhibitor of nonsense-mediated decay, confirmed these results (data not shown).

Luciferase assay

As illustrated in Fig. 1, transient transfection of *FANCF* promoter constructs revealed a slightly significant reduction for the c.–141 – 280G > T variant in reporter activity relative to the WT *FANCF* promoter construct, showing a 15% reduced ability to drive the firefly luciferase reporter gene. As for the four other promoter variants examined, no significant effect was observed on transcriptional activation.

Table 2 Sequence variants detected in human *FANCF* and residues found in orthologues

SNP	SNP ID ^a	Amino-acid change	<i>Canis familiaris</i>	<i>Mus musculus</i>	<i>Rattus norvegicus</i>	<i>Gallus gallus</i>	<i>Monodelphis domestica</i>	<i>Xenopus tropicalis</i>	<i>Tetraodon nigroviridis</i>
15	c.415T > G	p.Ser139Ala	Ser	Ser	Ser	Ser	Ser	–	–
18	c.517C > T	p.Arg173Cys	Arg	Arg	Arg	Arg	Arg	–	–
19	c.577G > A	p.Val193Ile	Thr	Val	Val	Lys	Met	–	–
20	c.823A > G	p.Ile275Val	Ile	Ile	Ile	Ile	Ile	Ile	Ile
35	c.2755T > C	p.Pro919Ser	Leu	Phe	Leu	Ser	Leu	–	–

– No corresponding residue in this species

^a According to the nomenclature of the Human Genome Variation Society**Table 3** Identification of transcription-factor binding sites (TFBS) potentially affected by variants located in the *FANCF* promoter region

Variant	TFBS potentially affected	Strand	Sequence affected (wt sequence shown) ^a	Matrix similarity score change (wt/var) ^b
c.–141 – 785delCTTTT	Androgen Receptor	Plus	gatgaacctcc TTTT cttt	0.836/ND
	Krüppel-like Factor 4	Minus	ctcAAAG aaaagg aggtt	ND/0.990
	Myogenic Factor 4	Minus	ctttaATTTactcaag aaaagg	0.786/ND
c.–141 – 331T > C	Forkhead box I1	Plus	tttaaacAAACactagg	0.990/0.997
c.–141 – 280G > T	Krüppel-like Factor 4	Minus	aaaaaaacaAGGG	0.955/0.972
c.–141 – 64G > A	RREB1	Plus	cCCCAgtcctgcaca	ND/0.823
	Sp2	Minus	ttgtgtgcaGGACtg	0.800/ND
c.–141 – 55A > G	Aryl hydrocarbon Receptor	Minus	cccgcccttg TGTG caggactgg	ND/0.974
	Sp2	Minus	ttgtgtgcaGGACtg	0.800/ND

ND not detected

^a Positions in bold are variant positions; capital letters are positions representing the core sequence of the binding site^b wt wild-type allele; var variant allele

Linkage disequilibrium and haplotype analyses

Pairwise LD between all 42 SNPs identified in the case series was calculated using the LDA program (data not shown). Complete LD was found between SNPs 1 (c.–141 – 785delCTTTT) and 42 (c.3750 + 483T > C), which are located ~181 kb apart ($|D'| = 1.0$), indicating that LD at the *FANCF* locus does not decline significantly with distance. However, regarding specific association involving SNPs 16 (c.508 – 281A > G) and 17 (c.508 – 31C > G), lower LD values were observed, mainly with variants located within the 3'UTR region of the gene. As expected, r^2 coefficient calculated for the *FANCF* gene region displayed lower values, as this measure was dependent on allelic frequency, which was well represented by the large spectrum of r^2 values ranging from 0 to 1.0. Indeed, the majority of lowest r^2 values were observed for 19 of the 25 SNPs displaying an MAF <5%.

PHASE program estimation of *FANCF* haplotypes reconstructed from the 17 SNPs having an MAF $\geq 5\%$ in both sample sets identified 17 haplotypes exhibiting a frequency $\geq 1\%$ (Fig. 2a). These 17 haplotypes represent >80% of all haplotypes estimated in both series. The permutation test of these 17 SNP haplotypes indicated no

significant difference in the estimated haplotype frequency distributions between both groups ($P = 0.327$, data not shown).

Given that no significant difference was observed between cases and controls following PHASE analyses, that no deviation from HWE for all SNPs in either series was observed, and that no significant difference in haplotype frequencies was identified, these observations support the use of individuals from both series as one population to identify LD blocks and tSNPs of the *FANCF* gene in the French Canadian population.

LD-block assignment was conducted using the Haploview software and led to the identification of three LD blocks in the *FANCF* gene (Fig. 2b). The 5'-block identified (promoter region to intron 5) encompasses approximately 16 kb, whereas LD blocks two (intron 8–16) and three (exon 19–20) covered 59 kb and 3 kb, respectively. Following LD-block partition, tSNPs were identified within each block. When considering haplotypes having a frequency $\geq 5\%$, eight tSNPs were identified in the French Canadian population, which represents >95% of all haplotypes. SNPs 4 (rs2048718), 7 (rs4988340), and 12 were identified as tSNPs within the first block, whereas SNPs 22 and 25 (rs11390869, rs2191248) represent the middle

Table 4 Identification of exonic splicing enhancer (ESE) potentially affected by exonic variants in the *FANCF* gene

Variant	Location	Binding-site potentially affected	Motif in wild-type sequence ^a	Score	Motif in variant sequence ^a	Score
c.415T > G	Exon 5	SC35	ATCTGCTA	2.473	AGCTGCTA	3.121
c.517C > T	Exon 6	SRp55	AACGTC	3.640	AATGTC	2.089
c.577G > A	Exon 7	SRp40	TGACTAT	2.308	TGACTGT	2.875
c.2637G > A	Exon19	SC35	GAGTCCTT	1.292	GAATCCTT	2.589
		SC35	AGTCCTTG	2.401	AATCCTTG	2.041
c.2755T > C	Exon19	SF2/ASF	CTCACCT	2.536	CCCACCT	1.932
		SC35	GTACCTCA	1.970	GTACCCCA	3.700
		SRp40	CCTCACC	3.772	CCCCACC	2.424
		SRp55	TACCTC	3.223	TACCCC	0.618
c.3411C > T	Exon 20	SC35	CGATCCTG	3.181	TGATCCT	3.160
c.3750 + 396delCT	Exon 20	SC35	CTTCTCTG	2.447	CTTCTGCC	1.400

^a Variant positions are in bold

Fig. 1 Promoter activity of long (1,068-bp) and short (633-bp) constructs carrying variations present in the promoter. An asterisk (*) indicates $P < 0.05$ (Student t test) relative to its respective wild-type (WT) construct

block, with the third block being defined by SNPs 34, 35, and 36 (rs4986765, rs4986764, and rs4988357).

Using HapMap data from the CEPH/CEU cohort, three LD blocks were also identified (Fig. 2b). Out of 17 SNPs used in our haplotype block analysis in French Canadian individuals, 11 were part of the CEPH/CEU sample set, with nine present at an MAF $\geq 5\%$ in the CEPH/CEU sample set. Despite using a different panel of SNPs, composition of the three blocks was relatively the same between French Canadian and CEPH/CEU data sets. Indeed, similar regions of LD breakage were observed following both analyses.

Discussion

To increase the power of our study aiming at finding genetic variants involved in breast cancer susceptibility, we

selected individuals from our cohort of French Canadian non-*BRCA1/2* high-risk breast cancer families (one individual per family). Given that this population is considered as a founder population, this allowed us to increase the likelihood of potentially identifying genetic variants associated with breast cancer (Antoniou and Easton 2003).

The previous identification of *BRCA2/FANCD1* and *PALB2/FANCN* (Rahman et al. 2007; Erkkö et al. 2007) as breast cancer susceptibility genes further strengthens the implication of FA genes in breast cancer susceptibility. Inactivating truncating mutations of *FANCF* have been demonstrated to confer susceptibility to breast cancer in high-risk non-*BRCA1/BRCA2* families from the UK (Seal et al. 2006). Although no deleterious germline mutation leading to a premature termination of the protein was identified in our study, 42 nucleotide variations were found, including 22 novel variants not previously reported in nucleotide databases (dbSNP and UCSC databases).

A number of case-control studies investigated the potential association of *FANCF* germline mutations with breast cancer susceptibility. Among the few polymorphisms/mutations not identified in our study, and for which a significant association has been demonstrated, P47A and M299I were originally described by Cantor et al. (2001). However, these breast cancer associations were not subsequently confirmed in large breast cancer cohorts (Rutter et al. 2003; Lewis et al. 2005; Seal et al. 2006). Two silent variants observed in our French Canadian cohort (c.2637G > A: p.Glu879Glu and c.3411C > T: p.Tyr1137Tyr) have been investigated in several breast cancer case-control studies regarding their implication with breast cancer susceptibility, but none of these studies reported a significant association with breast cancer for both nucleotide alterations (Cantor et al. 2001; Luo et al. 2002; Karppinen et al. 2003; Rutter et al. 2003; Lewis et al. 2005; Vahteristo et al. 2006).

Fig. 2 **a** Haplotype frequencies and combined frequencies of haplotypes exhibiting a minor allele frequency (MAF) $\geq 1\%$ in cases and controls, as reconstructed by the PHASE program. Single nucleotide polymorphisms (SNPs) denoted with an asterisk (*) are insertions and are represented by the *I* letter in haplotypes. **b** Haplotype blocks and tagging SNPs (tSNPs) identified using SNPs showing an MAF $> 5\%$ (17 SNPs). Tagging SNPs identified on a block-by-block basis using

haplotypes showing a frequency $> 5\%$ 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 $> 10\%$ are represented by *thick lines*

Whereas the c.517C $>$ T (p.Arg173Cys) variant has been reported to impair nuclear translocation (Yu et al. 2003), this nucleotide change, together with another rare missense variant (c.577G $>$ A: p.Val193Ile), do not seem to be associated with breast cancer in other studies (Luo et al. 2002; Rutter et al. 2003; Lewis et al. 2005; Vahteristo et al. 2006; Seal et al. 2006). As for the common c.2755T $>$ C variant (p.Pro919Ser) located in the BRCA1-binding domain spanning residues 888–1,063 (Cantor et al. 2001) and which has been previously reported to be associated with an increased risk of breast cancer to age 50 (Sigurdson et al. 2004), we and others did not highlight any significant association with breast cancer risk (Cantor et al. 2001; Luo et al. 2002; Karppinen et al. 2003; Rutter et al. 2003; Vahteristo et al. 2006; Garcia-Closas et al. 2006; Seal et al. 2006; Frank et al. 2007).

Of the five variations identified in the promoter region, the sole variant assessed in previous studies (c.-141 – 64G $>$ A) demonstrated no association with either breast cancer (Rutter et al. 2003; Sigurdson et al. 2004; Song et al. 2007; Pharoah et al. 2007; Frank et al. 2007) or with bladder cancer risk (Figueroa et al. 2007). In concordance

with these findings, we found this sequence variation to be more frequent in the control series, further supporting its nonassociation with breast cancer risk. With the exception of c.205 + 162T $>$ C and c.1281 + 91insT, all reported intronic variants were evaluated in the context of breast cancer (Karppinen et al. 2003; Rutter et al. 2003; Lewis et al. 2005; Vahteristo et al. 2006; Garcia-Closas et al. 2006; Song et al. 2007; Pharoah et al. 2007), and none were shown to be associated with an increased risk.

In agreement with our results, two recent studies, performed on large cohorts of breast cancer cases, included four variants identified in the study (c.1340 + 109G $>$ A, c.-141 – 64G $>$ A, c.508 – 31C $>$ G, c.508 – 281A $>$ G), and they came to the conclusion that none were associated significantly with breast cancer risk (Pharoah et al. 2007; Song et al. 2007), although rs4988344 (c.508 – 31C $>$ G) seemed to be associated with an increased risk of ovarian cancer (Song et al. 2007). Although eight variants could not be evaluated further due to material availability from other family members, none of the variations examined clearly displayed a segregation with the disease. However, partial or incomplete segregation patterns were recently

identified for susceptibility alleles that confer modest increase in breast cancer risk (Rahman et al. 2007), thus reinforcing the need for extensive analysis of the variants identified in our study.

Three of the five promoter variants reported here are predicted to impact TFBS affinity, which is consistent with other results obtained in 16 cell-cycle checkpoint genes (Belanger et al. 2005). The identification of RREB1 and KLF4 BS in the *FANCF* promoter, involved in the regulation of the cell-cycle checkpoint regulator p16, and in G1/S and G2/M transitions of the cell cycle, respectively, points toward a possible cell-cycle-dependent regulation of *FANCF*, which is concordant with the known cell-cycle-dependent phosphorylation and interaction of FANCF with BRCA1 (Yu et al. 2003). The promoter variants identified could also be involved in hormone activation or hormone response. Indeed, the c.-141 – 785delCTTTT variant seems to abolish the sole AR BS present in the promoter region analyzed, whereas the c.-141 – 64G > A variant creates a potential AHR BS, the latter known to be implicated in repression of E2F interaction with promoter of cell-cycle genes (Marlowe et al. 2004) and of estrogen-receptor (ER) binding to estrogen-responsive elements (Kharat and Saatcioglu 1996), resulting in inhibition of proliferation. Moreover, a BS for the Sp2 transcription factor, disrupted by both c.-141 – 64G > A and c.-141 – 55A > G variants, is known to interact with the G1/S phase transition-associated E2F1 transcription factor and to regulate luteinizing hormone (LH) receptor at the transcription level.

With the exception of the c.-141 – 280G > T variant, for which a slightly significant difference in the ability to drive the firefly luciferase reporter gene was identified, no other variant identified in this study seems to cause a change in the basal transcriptional activity of the *FANCF* gene. Further investigations are required to determine the potential impact of this promoter variant on binding of transcription factors to the *FANCF* promoter region.

Regarding alternative splicing of the *FANCF* gene, our findings are consistent with results from Cantor et al. (2001) in which no aberrant spliced mRNA species was identified in individuals from breast cancer families. Altered splicing has, however, been reported in individuals from FA complementation group J (Levrin et al. 2005; Levitus et al. 2005) and has also been observed in an individual affected with breast cancer from a *BRCA1/2*-negative family (Seal et al. 2006).

Haplotype reconstruction from the 17 SNPs having an MAF $\geq 5\%$ in both sample sets led to identification of 11 haplotypes (frequency $\geq 2\%$) (Fig. 2a). The only other haplotype analysis identified nine haplotypes, with a frequency $\geq 2\%$ (Rutter et al. 2003). Although similar, the greater diversity obtained here is a reflection of the number of SNPs

used for haplotype analysis: 17 in this study as opposed to nine (all included in our set) in Rutter et al. (2003).

Analysis conducted with the Haploview software identified three distinct blocks (Fig. 2b). A previous study identified two haplotype blocks located at the 5' and 3' ends with a large region of LD breakage comprising exons 1–12 (Rutter et al. 2003). Another investigation performed by Song et al. (2007) also identified two haplotype blocks but with LD breakage between introns 6 and 9. We identified three haplotype blocks, which is concordant with results obtained from HapMap data, in which three haplotype blocks were also identified. Another study including 9-kb upstream and 9-kb downstream of the *FANCF* gene identified 12 tSNPs (Song et al. 2007), whereas eight tSNPs were identified among our French Canadian individuals. This number of tSNP is consistent with the number of tSNPs normally required at other gene loci and in other populations (Johnson et al. 2001; Weale et al. 2003), including the French Canadian population (Durocher et al. 2006, 2007). Although only two tSNPs (rs2191248 and rs4986765) have been tagged in both sample sets, this could be explained by the fact that several equivalent SNPs may represent a given LD block.

Analysis of promoter, exonic, and flanking intronic regions of *FANCF* in 96 high-risk individuals from non-*BRCA1/BRCA2* French Canadian families did not lead to the identification of deleterious truncating mutation, exon deletion, or retention of an intronic portion. However, this analysis led to the identification of 42 variants, 22 of which are novel, including two common variants displaying a frequency of 43% and 38%. Moreover, this study reports four novel variations in the promoter region of *FANCF* potentially affecting transcription factor BS, one of these significantly altering the promoter activity in a reporter-gene assay. Further studies are therefore warranted to fully establish the functional contribution of these sequence changes on *FANCF* gene expression.

Acknowledgments The authors are indebted to the participants and their families for their generosity and providing DNA samples. We thank Dr. Damian Labuda and Claudia Moreau at the Centre de recherche de l'Hôpital Ste-Justine for providing control DNA samples. We would like to thank Dr. Martine Dumont, Dr. Penny Soucy, Gilles Leblanc, Carole 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 counselling and clinical data management at the Cancer Genomics Laboratory. A special thank to Mélanie Houde for skillful laboratory assistance. We thank Professor Bartha Maria Knoppers and her colleagues from the Centre de recherche en droit public de l'Université de Montréal for their valuable 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) and Institute of Cancer and Institute of Gender and Health for the INHERIT BRCA research program, the Fonds de la Recherche

en Santé du Québec (FRSQ)/Réseau de Médecine Génétique Appliquée (RMGA), the Canadian Breast Cancer Research Alliance, the CURE Foundation, and the Canadian Fanconi Anemia Research Fund. FG is recipient of a studentship from Fondation René Bussires, 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 Sant du Qubec (FRSQ) and a Research Career Award in the Health Sciences from CIHR/Rx&D Health Research Foundation.

Appendix

Other members of INHERIT BRCAs involved in clinical aspects of this study:

Paul Bessette: Service de gyncologie, Centre Hospitalier Universitaire de Sherbrooke, Fleurimont, QC, J1H 5N4, Canada.

Jocelyne Chiquette: Clinique des maladies du sein Deschnes-Fabia, Hpital du Saint-Sacrement, Qubec, QC, G1S 4L8, Canada.

Rachel Laframboise: Service de mdecine gntique, CHUQ, Pavillon CHUL, Qubec, QC, G1V 4G2, Canada.

Jean Lpine: Centre Hospitalier rgional de Rimouski, Rimouski, QC, G5L 5T1, Canada.

Bernard Lesprance, Roxane Pichette: Service d'hmato-oncologie, Hpital du Sacr-Cur, Montral, QC, H4J 1C5, Canada.

Marie Plante: Service de gyncologie, CHUQ, L'Htel-Dieu de Qubec, Qubec, QC, G1R 2J6, Canada.

References

- Alter BP (2003) Cancer in Fanconi anemia, 1927–2001. *Cancer* 97(2):425–440
- Antoniou AC, Easton DF (2003) Polygenic inheritance of breast cancer: implications for design of association studies. *Genet Epidemiol* 25:190–202
- Antoniou AC, Durocher F, Smith P, Simard J, Easton DF; INHERIT BRCAs Program Members (2006) BRCA1 and BRCA2 mutation predictions using the BOADICEA and BRCAPRO models and penetrance estimation in high-risk French-Canadian families. *Breast Cancer Res* 8(1):R3
- Ball S, Arolker M, Purushotham AD (2001) Breast cancer, Cowden disease and PTEN-MATCHS syndrome. *Eur J Surg Oncol* 27(6):604–606
- Belanger H, Beaulieu P, Moreau C, Labuda D, Hudson TJ, Sinnett D (2005) Functional promoter SNPs in cell cycle checkpoint genes. *Hum Mol Genet* 14(18):2641–2648
- Bonfield JK, Rada C, Staden R (1998) Automated detection of point mutations using fluorescent sequence trace subtraction. *Nucleic Acids Res* 26(14):3404–3409
- Boukamp P, Stanbridge EJ, Foo DY, Cerutti PA, Fusenig NE (1990) c-Ha-ras oncogene expression in immortalized human keratinocytes (HaCaT) alters growth potential in vivo but lacks correlation with malignancy. *Cancer Res* 50(9):2840–2847
- Boukamp P, Petrussevska RT, Breitkreutz D, Hornung J, Markham A, Fusenig NE (1998) Normal keratinization in a spontaneously immortalized aneuploid human keratinocyte cell line. *J Cell Biol* 106(3):761–771
- Cantor SB, Bell DW, Ganesan S, Kass EM, Drapkin R, Grossman S, Wahrer DC, Sgroi DC, Lane WS, Haber DA, Livingston DM (2001) BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function. *Cell* 105(1):149–160
- Cantor S, Drapkin R, Zhang F, Lin Y, Han J, Pamidi S, Livingston DM (2004) The BRCA1-associated protein BACH1 is a DNA helicase targeted by clinically relevant inactivating mutations. *Proc Natl Acad Sci USA* 101(8):2357–2362
- Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR (2003) ESEfinder: a web resource to identify exonic splicing enhancers. *Nucleic Acids Res* 31(13):3568–3571
- Cartharius K, Frech K, Grote K, Klocke B, Haltmeier M, Klingenhoff A, Frisch M, Bayerlein M, Werner T (2005) MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics* 21(13):2933–2942
- CHEK2 Breast Cancer Case-Control Consortium (2004) CHEK2*1100delC and susceptibility to breast cancer: a collaborative analysis involving 10,860 breast cancer cases and 9,065 controls from 10 studies. *Am J Hum Genet* 74(6):1175–1182
- Chenevix-Trench G, Spurdle AB, Gatei M, Kelly H, Marsh A, Chen X, Donn K, Cummings M, Nyholt D, Jenkins MA, Scott C, Pupo GM, Dork T, Bendix R, Kirk J, Tucker K, McCredie MR, Hopper JL, Sambrook J, Mann GJ, Khanna KK (2002) Dominant negative ATM mutations in breast cancer families. *J Natl Cancer Inst* 94(3):205–215
- Collaborative Group on Hormonal Factors in Breast Cancer (2001) Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. *Lancet* 358(9291):1389–1399
- Desjardins S, Belleau P, Labrie Y, Ouellette G, Bessette P, Chiquette J, Laframboise R, Lpine J, Lesprance B, Pichette R, Plante M, INHERIT BRCAs, Durocher F (2008) 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(1):108–116
- Devlin B, Risch N (1995) A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics* 29(2):311–322
- Ding K, Zhou K, He F, Shen Y (2003) LDA—a java-based linkage disequilibrium analyzer. *Bioinformatics* 19(16):2147–2148
- Duguay Y, Br C, Skorpn F, Guillemette C (2004) A novel functional polymorphism in the uridine diphosphate-glucuronosyltransferase 2B7 promoter with significant impact on promoter activity. *Clin Pharmacol Ther* 75(3):223–233
- Durocher F, Gunard F, Desjardins S, Ouellette G, Labrie Y (2005) Inherited susceptibility to breast cancer: accomplishments and challenges. In: Sinnett D (ed) *Molecular genetics of cancer*. Research Signpost, Kerala, pp 19–93
- Durocher F, Labrie Y, Soucy P, Sinilnikova O, Labuda D, Bessette P, Chiquette J, Laframboise R, Lpine J, Lesprance B, Ouellette G, Pichette R, Plante M, Tavtigian SV, Simard J (2006) 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
- Durocher F, Labrie Y, Ouellette G, INHERIT BRCAs, Simard J (2007) Genetic sequence variations and ADPRT haplotype analysis in French Canadian families with high risk of breast cancer. *J Hum Genet* 52(12):963–977
- Easton DF (1999) How many more breast cancer predisposition genes are there? *Breast Cancer Res* 1(1):14–17
- Ellis NA, Groden J, Ye TZ, Straughen J, Lennon DJ, Ciocci S, Proytcheva M, German J (1995) The Bloom's syndrome gene product is homologous to RecQ helicases. *Cell* 83(4):655–666

- Erkko H, Xia B, Nikkila J, Schleutker J, Syrjakoski K, Mannermaa A, Kallioniemi A, Pylkas K, Karppinen SM, Rapakko K, Miron A, Sheng Q, Li G, Mattila H, Bell DW, Haber DA, Grip M, Reiman M, Jukkola-Vuorinen A, Mustonen A, Kere J, Aaltonen LA, Kosma VM, Kataja V, Soini Y, Drapkin RI, Livingston DM, Winqvist R (2007) A recurrent mutation in PALB2 in Finnish cancer families. *Nature* 446(7133):316–319
- Figuerola JD, Malats N, Rothman N, Real FX, Silverman D, Kogevinas M, Chanock S, Yeager M, Welch R, Dosemeci M, Tardon A, Serra C, Carrato A, Garcia-Closas R, Castano-Vinyals G, Garcia-Closas M (2007) Evaluation of genetic variation in the double-strand break repair pathway and bladder cancer risk. *Carcinogenesis* 28(8):1788–1793
- Frank B, Hemminki K, Meindl A, Wappenschmidt B, Sutter C, Kiechle M, Bugert P, Schmutzler RK, Bartram CR, Burwinkel B (2007) BRIP1 (BACH1) variants and familial breast cancer risk: a case-control study. *BMC Cancer* 7:83
- Garcia-Closas M, Egan KM, Newcomb PA, Brinton LA, Titus-Ernstoff L, Chanock S, Welch R, Lissowska J, Peplonska B, Szeszenia-Dabrowska N, Zatonski W, Bardin-Mikolajczak A, Struewing JP (2006) Polymorphisms in DNA double-strand break repair genes and risk of breast cancer: two population-based studies in USA and Poland, and meta-analyses. *Hum Genet* 119(4):376–388
- Gasco M, Yulug IG, Crook T (2003) TP53 mutations in familial breast cancer: functional aspects. *Hum Mutat* 21(3):301–306
- Giardiello FM, Brensinger JD, Tersmette AC, Goodman SN, Petersen GM, Booker SV, Cruz-Correa M, Offerhaus JA (2000) Very high risk of cancer in familial Peutz-Jeghers syndrome. *Gastroenterology* 119(6):1447–1453
- Guenard F, Labrie Y, Ouellette G, Beauparlant CJ, Bessette P, Chiquette J, Laframboise R, Lépine J, Lespérance B, Pichette R, Plante M; INHERIT BRCA, Durocher F (2007) 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(4):483–490
- Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RC, Payne F, Hughes W, Nuttall S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SC, Clayton DG, Todd JA (2001) Haplotype tagging for the identification of common disease genes. *Nat Genet* 29(2):233–237
- Karppinen SM, Vuosku J, Heikkinen K, Allinen M, Winqvist R (2003) No evidence of involvement of germline BACH1 mutations in Finnish breast and ovarian cancer families. *Eur J Cancer* 39(3):366–371
- Kharat I, Saatcioglu F (1996) Antiestrogenic effects of 2,3,7,8-tetrachlorodibenzo-p-dioxin are mediated by direct transcriptional interference with the liganded estrogen receptor. Cross-talk between aryl hydrocarbon- and estrogen-mediated signaling. *J Biol Chem* 271(18):10533–10537
- Kitao S, Ohsugi I, Ichikawa K, Goto M, Furuichi Y, Shimamoto A (1998) Cloning of two new human helicase genes of the RecQ family: biological significance of multiple species in higher eukaryotes. *Genomics* 54(3):443–452
- Kitao S, Shimamoto A, Goto M, Miller RW, Smithson WA, Lindor NM, Furuichi Y (1999) Mutations in RECQL4 cause a subset of cases of Rothmund-Thomson syndrome. *Nat Genet* 22(1):82–84
- Laloo F, Varley J, Ellis D, Moran A, O'Dair L, Pharoah P, Evans DG, Early Onset Breast Cancer Study Group (2003) Prediction of pathogenic mutations in patients with early-onset breast cancer by family history. *Lancet* 361(9363):1101–1102
- Lei H, Vorechovsky I (2003) BACH1 517C- > T transition impairs protein translocation to nucleus: a role in breast cancer susceptibility? *Int J Cancer* 104(3):389–391
- Levitov M, Waisfisz Q, Godthelp BC, de Vries Y, Hussain S, Wiegant WW, Elghalbzouri-Maghrani E, Steltenpool J, Rooimans MA, Pals G, Arwert F, Mathew CG, Zdzienicka MZ, Hiom K, De Winter JP, Joenje H (2005) The DNA helicase BRIP1 is defective in Fanconi anemia complementation group J. *Nat Genet* 37(9):934–935
- Levrin O, Attwooll C, Henry RT, Milton KL, Neveling K, Rio P, Batish SD, Kalb R, Velleuer E, Barral S, Ott J, Petrini J, Schindler D, Hanenberg H, Auerbach AD (2005) The BRCA1-interacting helicase BRIP1 is deficient in Fanconi anemia. *Nat Genet* 37(9):931–933
- Lewis AG, Flanagan J, Marsh A, Pupo GM, Mann G, Spurdle AB, Lindeman GJ, Visvader JE, Brown MA, Chenevix-Trench G, Kathleen Cuninghame Foundation Consortium for Research into Familial Breast Cancer (2005) Mutation analysis of FANCD2, BRIP1/BACH1, LMO4 and SFN in familial breast cancer. *Breast Cancer Res* 7(6):R1005–1016
- Lewontin RC (1964) The interaction of selection and linkage. II. Optimum models. *Genetics* 50:757–782
- Luo L, Lei H, Du Q, von Wachenfeldt A, Kockum I, Luthman H, Vorechovsky I, Lindblom A (2002) No mutations in the BACH1 gene in BRCA1 and BRCA2 negative breast-cancer families linked to 17q22. *Int J Cancer* 98(4):638–639
- Marlowe JL, Knudsen ES, Schwemberger S, Puga A (2004) The aryl hydrocarbon receptor displaces p300 from E2F-dependent promoters and represses S phase-specific gene expression. *J Biol Chem* 279(28):29013–29022
- Medhurst AL, Huber PA, Waisfisz Q, de Winter JP, Mathew CG (2001) Direct interactions of the five known Fanconi anaemia proteins suggest a common functional pathway. *Hum Mol Genet* 10(4):423–429
- Moisan AM, Fortin J, Moisan MD, Samson C, Bessette P, Chiquette J, Laframboise R, Lépine J, Lesperance B, Pichette R, Plante M, Provencher L, Voyer P, Goldgar D, Bridge P, Simard J (2006) No evidence of BRCA1/2 genomic rearrangements in high-risk French-Canadian breast/ovarian cancer families. *Genet Test* 10(2):104–115
- Pharoah PD, Tyrer J, Dunning AM, Easton DF, Ponder BA; SEARCH Investigators (2007) Association between common variation in 120 candidate genes and breast cancer risk. *PLoS Genet* 3(3):e42
- Ponder BA (2001) Cancer genetics. *Nature* 411(6835):336–341
- Quandt K, Frech K, Karas H, Wingender E, Werner T (1995) MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res* 23(23):4878–4884
- Rahman N, Seal S, Thompson D, Kelly P, Renwick A, Elliott A, Reid S, Spanova K, Barfoot R, Chagtai T, Jayatilake H, McGuffog L, Hanks S, Evans DG, Eccles D, Easton DF, Stratton MR (2007) PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nat Genet* 39(2):165–167
- Reese MG, Eeckman FH, Kulp D, Haussler D (1997) Improved splice site detection in Genie. *J Comput Biol* 4(3):311–323
- Renwick A, Thompson D, Seal S, Kelly P, Chagtai T, Ahmed M, North B, Jayatilake H, Barfoot R, Spanova K, McGuffog L, Evans DG, Eccles D, Breast Cancer Susceptibility Collaboration (UK), Easton DF, Stratton MR, Rahman N (2006) ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. *Nat Genet* 38(8):873–875
- Rutter JL, Smith AM, Davila MR, Sigurdson AJ, Giusti RM, Pineda MA, Doody MM, Tucker MA, Greene MH, Zhang J, Struewing JP (2003) Mutational analysis of the BRCA1-interacting genes ZNF350/ZBRK1 and BRIP1/BACH1 among BRCA1 and BRCA2-negative probands from breast-ovarian cancer families and among early onset breast cancer cases and reference individuals. *Hum Mutat* 22(2):121–128

- Sambrook J, Russell DW (2001) Irwin N, Janssen KA, Argentine J, Curtis S, Zierler M, Dickerson M, Sialiano I, McInerney N, Brown D, Schaefer S, deBruin D, Atkeson E, Weiss D, Curtis MD (eds) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor
- Seal S, Thompson D, Renwick A, Elliott A, Kelly P, Barfoot R, Chagtai T, Jayatilake H, Ahmed M, Spanova K, North B, McGuffog L, Evans DG, Eccles D, Easton DF, Stratton MR, Rahman N (2006) Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. *Nat Genet* 38(11):1239–1241
- Sigurdson AJ, Hauptmann M, Chatterjee N, Alexander BH, Doody MM, Rutter JL, Struwing JP (2004) Kin-cohort estimates for familial breast cancer risk in relation to variants in DNA base excision repair, BRCA1 interacting and growth factor genes. *BMC Cancer* 4:9
- Simard J, Dumont M, Moisan AM, Gaborieau V, Malouin H, Durocher F, Chiquette J, Plante M, Avard D, Bessette P, Brousseau C, Dorval M, Godard B, Houde L; INHERIT BRCA, Joly Y, Lajoie MA, Leblanc G, Lepine J, Lesperance B, Vezina H, Parboosingh J, Pichette R, Provencher L, Rheaume J, Sinnett D, Samson C, Simard JC, Tranchant M, Voyer P, Easton D, Tavtigian SV, Knoppers BM, Laframboise R, Bridge P, Goldgar D (2007) 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(2):107–121
- Smogorzewska A, Matsuoka S, Vinciguerra P, McDonald ER 3rd, Hurov KE, Luo J, Ballif BA, Gygi SP, Hofmann K, D'Andrea AD, Elledge SJ (2007) Identification of the FANCI protein, a monoubiquitinated FANCD2 paralog required for DNA repair. *Cell* 129(2):289–301
- Song H, Ramus SJ, Kjaer SK, Hogdall E, Dicioccio RA, Whittemore AS, McGuire V, Hogdall C, Jacobs IJ, Easton DF, Ponder BA, Dunning AM, Gayther SA, D P Pharoah P (2007) Tagging single nucleotide polymorphisms in the BRIP1 gene and susceptibility to breast and ovarian cancer. *PLoS ONE* 2:e268
- Stankovic T, Kidd AM, Sutcliffe A, McGuire GM, Robinson P, Weber P, Bedenham T, Bradwell AR, Easton DF, Lennox GG, Haites N, Byrd PJ, Taylor AM (1998) ATM mutations and phenotypes in ataxia-telangiectasia families in the British Isles: expression of mutant ATM and the risk of leukemia, lymphoma, and breast cancer. *Am J Hum Genet* 62(2):334–345
- Stephens M, Donnelly P (2003) A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* 73(5):1162–1169
- Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68(4):978–989
- Szabo CI, Schutte M, Broeks A, Houwing-Duistermaat JJ, Thorsten-son YR, Durocher F, Oldenburg RA, Wasielewski M, Odefrey F, Thompson D, Floore AN, Kraan J, Klijn JG, van den Ouweland AM, Wagner TM, Devilee P, Simard J, van 't Veer LJ, Goldgar DE, Meijers-Heijboer H (2004) Are ATM mutations 7271T->G and IVS10-6T->G really high-risk breast cancer-susceptibility alleles? *Cancer Res* 64(3):840–843
- Tischkowitz M, Xia B, Sabbaghian N, Reis-Filho JS, Hamel N, Li G, van Beers EH, Li L, Khalil T, Quenneville LA, Omeroglu A, Poll A, Lepage P, Wong N, Nederlof PM, Ashworth A, Tonin PN, Narod SA, Livingston DM, Foulkes WD (2007) Analysis of PALB2/FANCN-associated breast cancer families. *Proc Natl Acad Sci USA* 104(16):6788–6793
- Ueda K, Nishijima M, Inui H, Watatani M, Yayoi E, Okamura J, Yasutomi M, Nakamura Y, Miyoshi Y (1998) Infrequent mutations in the PTEN/MMAC1 gene among primary breast cancers. *Jpn J Cancer Res* 89(1):17–21
- Vahteristo P, Yliannala K, Tamminen A, Eerola H, Blomqvist C, Nevanlinna H (2006) BACH1 Ser919Pro variant and breast cancer risk. *BMC Cancer* 6:19
- Vezina H, Durocher F, Dumont M, Houde L, Szabo C, Tranchant M, Chiquette J, Plante M, Laframboise R, Lepine J, Nevanlinna H, Stoppa-Lyonnet D, Goldgar D, Bridge P, Simard J (2005) Molecular and genealogical characterization of the R1443X BRCA1 mutation in high-risk French-Canadian breast/ovarian cancer families. *Hum Genet* 117(2–3):119–132
- Weale ME, Depondt C, Macdonald SJ, Smith A, Lai PS, Shorvon SD, Wood NW, Goldstein DB (2003) Selection and evaluation of tagging SNPs in the neuronal-sodium-channel gene SCN1A: implications for linkage-disequilibrium gene mapping. *Am J Hum Genet* 73(3):551–565
- Yu CE, Oshima J, Fu YH, Wijsman EM, Hisama F, Alisch R, Matthews S, Nakura J, Miki T, Ouais S, Martin GM, Mulligan J, Schellenberg GD (1996) Positional cloning of the Werner's syndrome gene. *Science* 272(5259):258–262
- Yu X, Chini CC, He M, Mer G, Chen J (2003) The BRCT domain is a phospho-protein binding domain. *Science* 302(5645):639–642