

# Comparison of *BRCA1* polymorphisms, rare sequence variants and/or missense mutations in unaffected and breast/ovarian cancer populations

Francine Durocher<sup>1</sup>, Donna Shattuck-Eidens<sup>2</sup>, Melody McClure<sup>2</sup>, Fernand Labrie<sup>1</sup>, Mark H. Skolnick<sup>2,3</sup>, David E. Goldgar<sup>3</sup> and Jacques Simard<sup>1,\*</sup>

<sup>1</sup>Medical Research Council Group in Molecular Endocrinology, CHUL Research Center and Laval University, Quebec City, G1V 4G2, Canada, <sup>2</sup>Myriad Genetics, 390 Wakara Way, Salt Lake City, Utah, USA and <sup>3</sup>Department of Medical Informatics, University of Utah School of Medicine, Salt Lake City, UT 84108, USA

Received March 7, 1996; Revised and Accepted March 18, 1996

Inherited mutations in the *BRCA1* gene are known to confer a predisposition to breast and ovarian cancer. We have first characterized 19 sequence variants in the *BRCA1* gene during mutation screening by direct sequencing using DNA samples from breast/ovarian cancer patients or obligate carriers. The frequencies of these sequence variants were then compared with those found in control populations of women. Among the 10 sequence variants showing an estimated frequency of the less common allele above 0.05, Q/R356, L/P871, E/G1038, K/R1183 and S/G1613 result in a change of amino acids, 2201C/T, 2430T/C and 4427C/T are silent mutations and the two others, 4209–141C/A and 5272+66A/G, are intronic polymorphisms. These frequent polymorphisms, with the exception of Q/R356, were in complete or significant pairwise linkage disequilibrium as evaluated in our control populations. With one exception (L/P871), none of these variants had statistically significant ( $P < 0.05$ ) differences in allele frequency between breast/ovarian cancer patients or obligate carriers and our control populations. Four rare sequence variants designated 710C→T, D693N, R841W and S1040N were found in both unaffected and breast/ovarian cancer populations, while the missense mutations M1008I, E1219D, R1347G, T1561I and M1628V were detected only once in our patient population. When a functional test is available, it will be important to determine the consequence on the *BRCA1* activity of these rare sequence variants and missense mutations.

## INTRODUCTION

Breast cancer is the most common malignancy in women, affecting one in eight women and is their second leading cause of death from cancer (1), whereas ovarian cancer, although less frequent, is the leading cause of death from gynecologic malignancies in North

America. It is estimated that 5–10% of breast cancer cases may be due to inherited autosomal dominant susceptibility genes (2,3). Ovarian cancer is also known to have a familial component (4,5). These predictions were confirmed by the mapping of *BRCA1*, which is responsible for an autosomal dominant syndrome of increased susceptibility to breast and/or ovarian cancer (6–13). Germline mutations in *BRCA1* are estimated to be responsible for ~45% of early onset familial breast cancer, for almost all families with multiple cases of breast and ovarian cancer (9–12). Female *BRCA1* mutation carriers are estimated to have an 87% lifetime risk of developing breast cancer and a 44% risk of ovarian cancer in breast/ovarian families (9–14). Moreover, *BRCA1* carriers have a 4-fold increased risk of colon cancer, whereas male carriers face a 3-fold increased risk of prostate cancer (14).

Recently, the structure of the *BRCA1* gene was elucidated (15). The *BRCA1* gene contains 5592 bp of coding sequence encompassed within 22 exons spread over more than ~70 kb encoding a protein of 1863 amino acids. The only recognized homology with other known genes is limited to the NH<sub>2</sub>-terminus of the protein which contains a RING finger motif similar to that found in other proteins interacting with nucleic acid and/or forming protein complexes (16), thus suggesting a role in the regulation of gene transcription. The functional significance of this motif is supported by the observation that it is highly conserved in the mouse homolog gene (17,18) and by the detection of the two missense mutations C61G and C64G in *BRCA1* families (19–21).

*BRCA1* is thought to act as a tumor suppressor gene, as strongly suggested by the loss of the wild-type allele in breast tumors from linked families (20,22–24). Although the 15 mutations in the *BRCA1* gene detected in breast or ovarian carcinomas were all germline alterations (25–27), five somatic *BRCA1* mutations were recently found in sporadic ovarian tumors (28,29), thus supporting its putative role as a tumor suppressor. It has been recently reported that *BRCA1* expression is decreased during the transition from carcinoma *in situ* to invasive cancer, and that inhibition of *BRCA1* expression with antisense oligonucleotides increased growth of normal and malignant mammary cells, thus suggesting that *BRCA1* may normally serve as a negative regulator of mammary epithelial cell growth (30). There are also data supporting the role of *BRCA1* in growth regulatory function in mouse mammary gland (18,31).

\*To whom correspondence should be addressed

**Table 1.** Sequence variants in *BRCA1*<sup>a</sup>

Name <sup>b,c</sup>	Nucleotide change	Frequency <sup>e</sup>		Location			Effect on coding sequence	Residue in mouse homolog <sup>i</sup>
		Controls	Patients <sup>h</sup>	Nucleotide	Codon	Exon		
710C→T	C→T	2/94 <sup>f</sup>	1/216	710	197	9	Silent, Cys	Cys
Q/R356 <sup>d</sup>	A or G	10/166 <sup>g</sup>	20/346	1186	356	11	Gln or Arg	Pro
D693N	G→A	3/78 <sup>f</sup>	25/364	2196	693	11	Asp→Asn	Ala
2201C/T	C or T	19/78 <sup>f</sup>	111/352	2201	694	11	Silent, Ser	Ser
2430T/C	T or C	28/94 <sup>f</sup>	46/144	2430	771	11	Silent, Leu	Leu
R841W	C→T	1/156 <sup>g</sup>	1/302	2640	841	11	Arg→Trp	Gln
L/P871 <sup>d</sup>	T or C	63/226 <sup>f,g</sup>	90/214	2731	871	11	Leu or Pro	Leu
M1008I	G→A	0/88 <sup>f</sup>	1/194	3143	1008	11	Met→Ile	Ser
E/G1038	A or G	20/84 <sup>f</sup>	80/234	3232	1038	11	Glu or Gly	Glu
S1040N	G→A	3/82 <sup>f</sup>	7/242	3238	1040	11	Ser→Asn	Gly
K/R1183 <sup>d</sup>	A or G	50/156 <sup>g</sup>	57/182	3667	1183	11	Lys or Arg	Arg
E1219D	G→C	0/64 <sup>f</sup>	1/144	3776	1219	11	Glu→Asp	Glu
R1347G	A→G	0/232 <sup>f,g</sup>	1/144	4158	1347	11	Arg→Gly	Met
4209-141C/A	C or A	52/168 <sup>g</sup>	N.D.	4209-141		Intron 11	Non-coding	
4427C/T <sup>d</sup>	C or T	55/166 <sup>g</sup>	72/238	4427	1436	13	Silent, Ser	Pro
T1561I	C→T	0/240 <sup>f,g</sup>	1/308	4801	1561	16	Thr→Ile	Thr
S/G1613 <sup>d</sup>	A or G	51/164 <sup>g</sup>	113/356	4956	1613	16	Ser or Gly	Ala
M1628V	A→G	0/92 <sup>f</sup>	1/376	5001	1628	16	Met→Val	Val
5272+66A/G	A or G	55/178 <sup>g</sup>	60/156	5272+66		Intron 18	Non-coding	

<sup>a</sup>Sequence variants are listed in order of location from the 5' to 3' end of *BRCA1*.

<sup>b</sup>Designation of sequence variant according to Beaudet and Tsui (54).

<sup>c</sup>Nucleotides refer to the *BRCA1* cDNA sequence in GenBank under accession number U-14680.

<sup>d</sup>Q/R356, 4427C/T, S/G1613, 4209-141C/A, 5272+66A/G, L/P871 and K/R1183 were previously designated PM1, PM2, PM3, PM4, PM5, PM6 and PM7, respectively (15,25).

<sup>e</sup>Frequency of the least common allele on the number of chromosomes typed.

<sup>f</sup>Québec population screening analysis performed by direct sequencing.

<sup>g</sup>Utah population screening analysis performed by ASO hybridization analysis.

<sup>h</sup>Breast/ovarian cancer patients or *BRCA1* obligate carriers were screened by direct sequencing. The mutations M1008I, E1219D and M1628V were first detected by dideoxy fingerprinting screening and thereafter confirmed by direct sequencing.

<sup>i</sup>From (17,18).

N.D. Not determined.

The isolation of this tumor suppressor gene offered the possibility for genetic testing of asymptomatic women in families with a high incidence of breast and/or ovarian cancer (3,32). Despite the fact that mutation screening for this gene is a challenging task due to its structure, >70 distinct germline mutations have already been found through a complete screen of the *BRCA1* gene (15,19-21,25-27,33-40). Frameshift, non-sense, splice or regulatory mutations accounted for ~85% of *BRCA1* mutations, while the remainder are due to missense mutations. Until a functional test is available, general population frequency analysis of sequence variants in the *BRCA1* coding region is useful to support the putative role of missense mutations. Moreover, analysis of polymorphisms in the *BRCA1* coding sequence has been useful in detecting regulatory mutations (15,19), to study the allele-specific expression in sporadic breast tumors (30) and to determine the common origin of frequent mutations (41). The present study was designed to compare the allele frequency of 19 sequence variants in the *BRCA1* gene detected by our group in DNA samples from breast/ovarian cancer patients or *BRCA1* obligate carriers versus that seen in an unselected population of women, thus avoiding bias related to a strong family history of breast/ovarian cancer. We have also

studied whether the frequent polymorphisms are in pairwise linkage disequilibrium in our control populations.

## RESULTS AND DISCUSSION

The 19 sequence variants described in the present study (Table 1) were first detected during screening by direct sequencing for *BRCA1* mutations using DNA samples from 200 patients with breast and/or ovarian cancer or known obligate carriers by haplotype analysis, drawn primarily from high-risk families as well as from isolated patients with early-onset breast cancer as described (15,33,34). Mutation screening was performed by direct sequencing of PCR fragments covering the entire *BRCA1* coding region and intron-exon junctions. The differences observed in the number of patients tested for each sequence variant are due to variability in sequencing gel quality which is frequently associated with difficulty in determining the sequence of some regions in the *BRCA1* gene. The frequency of these sequence variants found in the latter population was compared with that observed in an unselected group of anonymous women from Québec city and/or in marry-in-spousal controls from families being studied for a variety of hereditary cancers in the Utah

population by direct sequencing or ASO hybridization assay as indicated in Table 1.

### Frequent polymorphisms

Among the 10 sequence variants showing an estimated frequency of the less common allele above 0.05, Q/R356, L/P871, E/G1038, K/R1183 and S/G1613 result in a change of amino acids, while 2201C/T, 2430T/C and 4427C/T are silent mutations and the two others, 4209–141C/A and 5272+66A/G are intronic polymorphisms (Table 1). With the exception of 4209–141C/A, these frequent polymorphisms are due to C→T or A→G substitutions, which is in agreement with the recognized excess of transition-type mutations in vertebrate genomes thought to be the result of hypermutability of the methylated dinucleotide CpG (42,43). Eight of the other sequence variants indicated in Table 1 were also caused by transition-type mutations.

The allele frequencies of the polymorphisms Q/R356, 4427C/T, S/G1613, 4209–141C/A, 5272+66A/G, L/P871 and K/R1183, initially designated PM1, PM2, PM3, PM4, PM5, PM6 and PM7, respectively (15,25), were previously reported in our

Utah control population. In the present study, we extend our initial characterization by providing their observed versus expected heterozygosity and deviation from Hardy–Weinberg equilibrium (Table 2) not only in control populations but also in a large number of breast/ovarian cancer patients or *BRCA1* obligate carriers. The characteristics of the polymorphisms 2201C/T, 2430T/C and E/G1038 are also included in Table 2. The allele frequencies of the polymorphisms 2201C/T, 2430T/C, E/G1038, K/R1183 and S/G1613 found in our control populations are in close agreement with those recently reported by King's group, all with frequencies of about  $q = 0.33$  based on 42 control chromosomes (19). However, King's group reported a  $q = 0.18$  for Q/R356 in their population sample compared with 0.06 in our control Utah population ( $P = 0.01$ ), whereas Merajver *et al.* (25) observed this missense polymorphism in four out of 80 (0.05) chromosomes. As far as we know, all three populations tested are Caucasian individuals of Northern European origin. The recent observation that the polymorphisms E/G1038, K/R1183 and S/G1613 have similar  $q$  values in a control Japanese population (37) argues against their recent origin in the populations studied.

**Table 2.** Characteristics of frequent polymorphisms in *BRCA1*<sup>a</sup>

Name	Population screened	n <sup>b</sup>	q <sup>c</sup>	Heterozygosity		P value (HWE) <sup>e</sup>
				Observed	Expected <sup>d</sup>	
Q/R356	Utah control <sup>f</sup>	166	0.06	0.12	0.11	NS
Q/R356	Br/Ov patients <sup>g</sup>	346	0.06	0.10	0.11	NS
2201C/T	Québec control	100	0.27	0.34	0.39	NS
2201C/T	Br/Ov patients	352	0.32	0.43	0.43	0.01
2430T/C	Québec control <sup>h</sup>	94	0.30	0.30	0.42	0.048
2430T/C	Br/Ov patients	144	0.32	0.31	0.43	0.01
L/P871	Utah control	144	0.31	0.53	0.42	0.039
L/P871	Québec control	82	0.23	0.22	0.36	0.014
L/P871	Pooled control <sup>i</sup>	226	0.28	0.44	0.42	NS
L/P871	Br/Ov patients	214	0.42	0.45	0.49	NS
E/G1038	Québec control	84	0.24	0.29	0.36	NS
E/G1038	Br/Ov patients	234	0.34	0.34	0.45	0.009
K/R1183	Utah control	156	0.32	0.54	0.44	0.037
K/R1183	Br/Ov patients	182	0.32	0.27	0.44	0.0003
4209–141C/A	Utah control	166	0.31	0.52	0.43	0.047
4427C/T	Utah control	166	0.33	0.53	0.44	0.06
4427C/T	Br/Ov patients	238	0.30	0.42	0.42	NS
S/G1613	Utah control	164	0.30	0.51	0.42	0.06
S/G1613	Br/Ov patients	356	0.33	0.43	0.44	NS
5272+66A/G	Utah control	174	0.30	0.52	0.42	0.039
5272+66A/G	Br/Ov patients	156	0.38	0.38	0.47	0.10

<sup>a</sup>Sequence variants with a frequency of the less common variant ( $q$ ) >5%.

<sup>b</sup>Number of chromosomes typed.

<sup>c</sup>Estimated allele frequency of the less common variant.

<sup>d</sup>Expected heterozygosity =  $2q(1-q)$ .

<sup>e</sup>Significance level of  $\chi^2$  test of Hardy–Weinberg equilibrium at the specified locus.

<sup>f</sup>Utah control population screening analysis done by ASO hybridization.

<sup>g</sup>Breast/ovarian cancer patient (or *BRCA1* obligate carriers) population screening done by direct sequencing.

<sup>h</sup>Québec control population screening analysis done by direct sequencing.

<sup>i</sup>Pooled control samples.

**Table 3.** Pairwise linkage disequilibrium analysis between pairs of polymorphisms in *BRCA1*<sup>a</sup>

Polymorphism pairs <sup>b</sup>	<i>n</i>	<i>D</i>	<i>r</i>	<i>r</i> <sub>max</sub>	% <i>r</i> <sub>max</sub>
L/P871×2201C/T	54	0.11	0.78	0.78	1.00
L/P871×2430T/C	66	0.13	0.86	0.86	1.00
L/P871×E/G1038	58	0.12	0.73	0.95	0.77
2201C/T×2430T/C	56	0.12	0.78	0.89	0.86
2201C/T×E/G1038	48	0.13	0.82	0.82	1.00
2430T/C×E/G1038	64	0.12	0.71	0.80	0.89
Q/R356×L/P871	138	-0.013	-0.139	0.300	
Q/R356×K/R1183	148	-0.013	-0.138	0.306	
Q/R356×4209-141C/A	156	-0.012	-0.131	0.305	
Q/R356×4427C/T	156	-0.012	-0.135	0.300	
Q/R356×S/G1613	152	-0.012	-0.131	0.313	
Q/R356×5272+66A/G	156	-0.012	-0.131	0.305	

<sup>a</sup>The polymorphisms L/P871, K/R1183, 4209-141C/A, 4427C/T, S/G1613 and 5272+66A/G are in complete linkage disequilibrium ( $r = 1.0$ ) in the Utah population sample of 64-78 chromosomes typed by ASO hybridization assay with *D* values ranging from 0.06 to 0.10. Data for these loci were not indicated in the table due to space limitation.

<sup>b</sup>The analysis between the Q/R356 and the other indicated five polymorphisms was typed by ASO hybridization assay in the control Utah population sample, while the other data indicated in the table were obtained from the control Québec city population sample typed by direct sequencing.

Differences in allele frequency of these frequent polymorphisms between breast/ovarian cases and controls were assessed by  $\chi^2$  analysis. With one exception, none of the variants had statistically significant ( $P < 0.05$ ) differences in allele frequency between the patients and the control population (pooled Utah/Québec samples for L/P871). The single exception was L/P871, in which a frequency of 0.42 of the T variant (Leu) was found in the 107 cases compared with 0.28 in the 113 pooled control individuals ( $P < 0.002$ ). Because the breast/ovarian (or *BRCA1* obligate carriers) came from an ill-defined set of samples ascertained from many different centers for *BRCA1* testing, this result could very well be due to population artifact. However, it is of sufficient interest to warrant a follow-up examination in a formal case control study.

Sequence alignment analysis of the mouse homolog of *BRCA1* compared with the human *BRCA1* protein revealed that the residues Cys197, Ser694 and Leu771, which are the sites of the three silent mutations 710C→T (described below), 2201C/G and 2430T/C, respectively, are conserved in the mouse *BRCA1* protein (Table 1) (17,18). Moreover, the residues Leu871, Glu1038 and Arg1183, which are frequently substituted in the polymorphisms L/P871, E/G1038 and K/R1183, are also conserved in the mouse *BRCA1* protein.

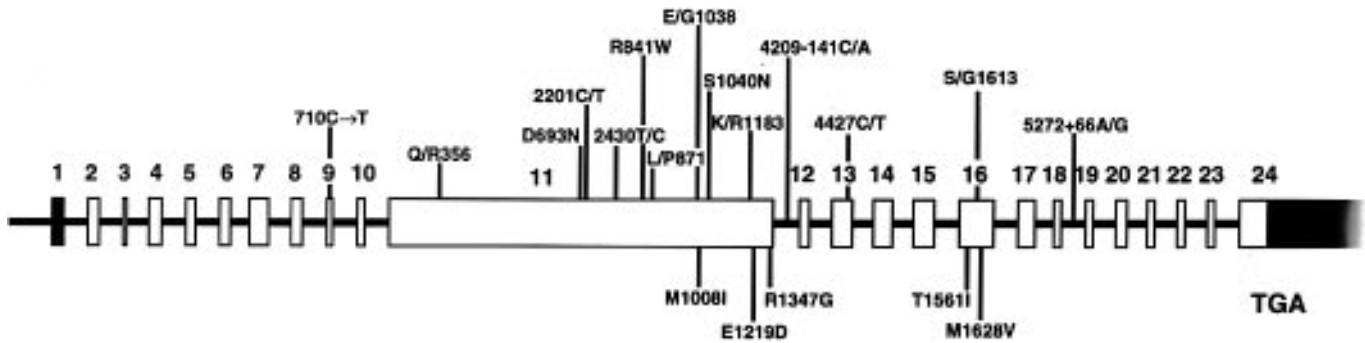
The polymorphisms L/P871, K/R1183, 4209-141C/A, 4427C/T, S/G1613 and 5272+66A/G are in complete pairwise linkage disequilibrium in a population sample of 64-78 chromosomes typed by ASO hybridization assay in our Utah control population with *D* values ranging from 0.06 to 0.10 and an *r* value of 1.0. As indicated in Table 3, no significant linkage disequilibrium was observed between Q/R356 and the latter six frequent polymorphisms. The polymorphism L/P871 has also been typed by direct sequencing in another unselected population of 48-66 chromosomes in parallel with 2201C/T, 2430T/C and E/G1038. As indicated in Table 3, the

pairs of loci L/P871×2201C/T, L/P871×2430T/C and 2201C/T×E/G1038 had significant pairwise linkage disequilibrium in which  $D = D_{\max}$ , which occurs when at least one haplotype combination is not observed. Moreover, the pairs of loci L/P871×E/G1038, 2201C/T×2430T/C and 2430T/C×E/G1038 were in highly significant ( $P < 0.001$  in all cases) but not complete pairwise linkage disequilibrium (Table 3). These data thus indicate that, with the exception of Q/R356 which is the most 5' polymorphism, all these frequent polymorphisms are in complete or highly significant pairwise linkage disequilibrium in our control populations.

The polymorphism Q/R356 did not show evidence of linkage disequilibrium with other polymorphisms tested. It should be noted, however, that this polymorphism was comparatively rare, with an estimated allele frequency of 0.06 of the Arg variant. Thus the linkage disequilibrium assessment between this polymorphism and the six other polymorphisms described above is based on 10 heterozygotes and no rare homozygotes. The evidence against close pairwise linkage disequilibrium comes from the observation of five individuals who are homozygous for the common variant at this locus and homozygous for the less common variant at the other six tested loci, and from six individuals who are heterozygous for Q/R356 and have the common homozygous genotype for the other polymorphisms. Thus, there are at least eight discordant haplotypes between Q/R356 and the other six polymorphisms tested in these individuals. The absence of detectable linkage disequilibrium between Q/R356 and the other six common polymorphisms is based entirely on analysis of the Utah control population. The Utah population consists largely of descendants of relatively recent immigrants from Scandinavia and England. Because of a large number of founders and continuous inward and outward migration, the control population studied has levels of inbreeding similar to other Northern European populations (44). It is of interest that these six polymorphisms, which are dispersed over a distance of ~30 kb of genomic DNA (Fig. 1) and are in complete linkage disequilibrium, all have allele frequencies of 0.28-0.33 of the less common allele in our control population, and have borderline deviation from Hardy-Weinberg equilibrium ( $P$  values 0.04-0.06). Although one can postulate a number of different possibilities as to how these polymorphisms arose, one scenario consistent with the observed data is that the six polymorphisms arose through relatively recent admixture between two populations each non-polymorphic for different variants at these nucleotide positions. If the Q/R356 mutation was present in both these populations prior to admixture, then this polymorphism could be in linkage equilibrium with respect to the other six polymorphisms. It is noteworthy that deviations from Hardy-Weinberg equilibrium are due to an increase in the number of observed heterozygotes compared with the expected; this is also consistent with admixture. It should be stated that this hypothesis is only one of many possibilities, each of which is essentially untestable given the available data. It is likely that different molecular mechanisms may be involved in creating/maintaining linkage disequilibrium within these small intragenic regions than those responsible over large physical distances.

#### Rare sequence variants and/or putative missense mutations

The sequence variant 710C→T at nucleotide position 710 is due to a change at the codon 197 (TGC) into TGT, both encoding a Cys



**Figure 1.** Schematic representation of the *BRCA1* gene showing the location of the 19 sequence variants analyzed. The sequence variants changing an amino acid and which were not detected in the normal populations studied are indicated below the diagram. Exons are numbered above the diagram and are represented by boxes in which the open boxes demarcate the coding regions, while black boxes represent the non-coding regions. Introns are represented by solid lines.

at the end of exon 9. This variant was found only once in 108 breast/ovarian cancer patients or obligate carriers (Table 1). This variant was detected in two out of 94 control chromosomes from anonymous women from Québec city. Although this synonymous mutation changes the nucleotide at position -3 of the donor splice site sequence, which is usually a C or A in the consensus sequence for vertebrate genes, it is not predicted that this mutation gives rise to aberrant splicing (45,46). This sequence variant was also detected independently in the kindred 178 recently described (20). No aberrant splicing was found by sequencing the cDNA sample obtained from transformed lymphoblastoid cells from this patient (F. Couch and B. Weber, pers. comm.). However, it cannot be ruled out that this rare sequence variant does decrease the splicing efficiency. Thereafter, this rare sequence variant was also found in one out of 115 unselected cases of ovarian carcinoma (26).

The sequence variant S1040N resulting from a G→A transition that converts codon 1040 (AGC) encoding Ser into AAC encoding the polar residue Asn was detected in seven of the 121 cases in which this variant was sequenced. We have found this rare sequence variant in three out of 82 control chromosomes from anonymous women from Québec city (Table 1). This variant was first reported as a missense mutation because this mutation did segregate with the disease in family 14 and was absent in 120 control chromosomes (19). In support of our data indicating that this mutation is rather a rare sequence variant, this alteration did not segregate with the disease in family 62, and was found by another group in three out of 232 control chromosomes (20).

We have also detected in our patient population the novel point mutation D693N that is caused by a G→A transition converting codon 693 (AGC) encoding the acidic residue Asp into AGT encoding the polar residue Asn (Table 1). This sequence variant was detected in three out of 78 control chromosomes from anonymous women from Québec city, whereas it was detected in 25 chromosomes out of 364 chromosomes from breast/ovarian cancer patients or obligate carriers. Moreover, S1040N and D693N had no statistically significant differences in allele frequency between breast/ovarian cancer patients or obligate carriers and control populations as tested by G-test of independence, Fisher's exact test or normal approximation to Poisson distribution.

The variant R841W was detected once in 151 cases tested for this mutation which is caused by a conversion of codon 841 (CGG) encoding the basic residue Arg into TGG encoding the non-polar amino acid Trp. This mutation was found in only one

out of 156 chromosomes from our Utah control population of women. It will be essential to verify the consequence of these rare sequence variants on the activity of *BRCA1* when a functional test is available, because it cannot be excluded that a significant abnormality could potentially be identified in our control populations.

The missense mutation R1347G caused by an A→G substitution at nucleotide 4158 changing codon 1347 (AGA) encoding Arg, a basic residue, into GGA encoding the non-polar residue Gly was first detected by us in the Utah kindred K2039 (34). This Arg was not conserved in the mouse *BRCA1* protein (17,18). This mutation was not found in the 71 additional breast/ovarian cancer patients or obligate carriers screened for this variant by sequencing. This missense mutation was not found in 156 chromosomes from our Utah control population screened by ASO hybridization assay as well as in another population of 76 chromosomes from our other Québec city control population analyzed by direct sequencing. However, this alteration was found in a patient who also had a severe frameshift mutation (34) but it was not determined if both mutations were on the same allele (M.-C. King, pers. comm.). Thus it cannot be excluded that this patient was a compound heterozygote and that R1347G may be a pathological allele.

The missense mutation T1561I is caused by a C→T transition at nucleotide 4801 in exon 16 changing codon 1561 (ACC) encoding the polar residue Thr, which is conserved in the mouse *BRCA1* protein, into ATC encoding the non-polar residue Ile. This point mutation was found in only one individual out of the 154 cases tested by sequencing for this mutation. However, this missense mutation was not found by direct sequencing in 90 control chromosomes from anonymous women from Québec city and in 150 chromosomes from our Utah control population tested by ASO analysis. However, this variant was originally found as a germline mutation but was absent in a breast tumor from the same patient (Roger Wiseman, pers. comm.). Because *BRCA1* is thought to act as a tumor suppressor gene, as strongly suggested by the loss of the wild-type allele in breast tumors from linked families (20,22-24), the latter observation could suggest the benign effect of this missense mutation on its function. On the other hand, it has been preliminarily reported that this mutation was recently found in a breast cancer patient from the family 930681 (47). It will be important to determine the real repercussion of these two missense mutations as well as of the R1347G

mutation on the *BRCA1* activity when a functional test is available.

Three additional novel sequence variants were characterized in our study (Table 1, Fig. 1). The M1008I missense mutation was found in a patient who developed a breast cancer at the age of 41. This mutation was not found in the 96 other cases tested for this mutation (Table 1). Moreover, this conservative mutation due to a G→A transition at nucleotide 3143 in exon 11 changing codon 1008 (ATG) encoding the non-polar Met into ATA encoding Ile, another non-polar residue, was not found by direct sequencing in 88 control chromosomes. This Met is changed for a Ser in the mouse *BRCA1* protein. The E1219D substitution was detected in a patient who developed a breast cancer at the age of 50 and she has a sister who developed a breast cancer at the age of 45. This missense mutation was not detected in the 71 additional cases tested for this substitution (Table 1). This conservative mutation results from a G→C transversion at nucleotide 3776, located in exon 11, converting codon 1219 (GAG) encoding the acidic residue Glu, which is conserved in the mouse *BRCA1* protein, into GAC encoding Asp, another acidic residue. The E1219D mutation was not found in 64 control chromosomes. In fact, the missense mutation M1628V was found in only one patient out of the 187 breast/ovarian cancer patients (or *BRCA1* obligate carriers) tested for this substitution. This latter patient developed a breast cancer at the age of 27. This point mutation was not found in 92 control chromosomes by direct sequencing. The M1628V mutation is also conservative and is due to an A→G transition at nucleotide 5001 in exon 16 changing codon 1628 (ATG) encoding the non-polar residue Met into GTG encoding Val, which is also a non-polar residue. It is of interest to note that the frequent polymorphism M1628T has been recently reported in the Japanese population (37), thus suggesting that the functional significance of the M1628V substitution could be minor, if any. In agreement with a limited effect, if any, of this substitution, there is a Val residue at this position in the mouse *BRCA1* protein.

In conclusion, we have observed that nine of the 10 frequent polymorphisms were in complete or significant pairwise linkage disequilibrium in our control populations, while the polymorphism Q/R356 did not show evidence of such a phenomenon with these nine other variants. We have also observed that, with one exception, L/P871, none of these frequent polymorphisms had statistically significant differences in allele frequency between control populations and breast/ovarian cancer patients (or *BRCA1* obligate carriers). We have also provided information suggesting that the point mutations 710C→T, D693N and S1040N are rare sequence variants showing frequencies much higher in our control populations than the estimated frequency of all *BRCA1* mutations together in the general population (13). No statistically significant difference in allele frequency of the rare sequence variants D693N and S1040N was found between control populations and breast/ovarian cancer patients (or *BRCA1* obligate carriers). It cannot be excluded that a significant abnormality could potentially be identified in our control populations; it will be essential to verify the consequence on *BRCA1* activity when a functional test is available. This is especially true for the point mutation R841W which was first detected in only one patient, and thereafter in one woman from our Utah control population. Furthermore, additional functional studies will also be required to determine the significance of the three conservative mutations M1008I, E1219D and M1628V as

well as of the two other non-conservative missense mutations R1347G and T1561I.

## MATERIALS AND METHODS

### Patient samples

The sequence variants described in the present study were detected during screening for *BRCA1* mutations using DNA samples isolated from peripheral blood lymphocytes, or Epstein–Barr virus-transformed lymphoblastoid cell lines or cDNA from such cell lines from 200 patients with breast and/or ovarian cancer or known carriers by haplotype analysis, drawn primarily from high-risk families as well as from isolated patients with early-onset breast cancer as described (15,33,34). All subjects signed appropriate informed consent forms.

### Control samples

The genomic DNA samples were isolated from an unselected group of women from Québec city or marry-in-spousal controls from families being studied for a variety of hereditary cancers in the Utah population. The genomic DNA samples from anonymous unselected women from Québec city were purified from whole blood collected from hemoglobin/hematocrit or complete blood count leftover samples as previously described (48,49) (kindly provided by Dr François Rousseau, St-François-D'Assise Hospital). The Québec city population consists almost exclusively of French-Canadians. The number of French immigrants to Canada in the 17th century has been estimated to be 8000, but not more than 3380 pioneers settled permanently in the St. Lawrence Valley. The molecular studies conducted so far indicate that both populations are genetically diversified (44,50,51).

### Detection of sequence variants by direct DNA sequencing

The sequence variants 710C→T, D693N, 2201C/T, 2430T/C, L/P871, M1008I, E/G1038, S1040N, E1219D, R1347G, T1561I and M1628V were screened in 32–54 anonymous control individuals from a population of Québec. PCRs were performed using the appropriate sets of primer as described (33), in a 50 µl volume containing 10 mM Tris–HCl, pH 8.3, 50 mM KCl, 1.5 mM MgCl<sub>2</sub>, 50 µl dNTPs, 0.5 µM of each primer, 0.1% Triton X-100, 2% dimethylsulfoxide (DMSO) and 100 ng of genomic DNA. The reaction was first heated at 96°C for 5 min, followed by an incubation at 94°C for 5 min during which time 1 U of *Taq* polymerase (Perkin-Elmer) was added. The reactions were carried out using a Perkin-Elmer Cetus thermal cycler with a two-step temperature cycle consisting of 40 s of denaturation at 94°C and 30 s of annealing at 60°C; the elongation step was considered to be the time taken by the thermal cycler to reach 94°C after the annealing step. After 35 cycles, a final extension at 72°C for 10 min was performed. The primers were removed by selective precipitation and the PCR products were then subjected to a 35-cycle asymmetric amplification under the same conditions, except for the use of 10 nM limiting primer. After selective precipitation, the asymmetric PCR products were sequenced directly. Single-stranded DNA produced by asymmetric PCR was sequenced by the dideoxy method using the limiting PCR primer or sequence-specific primers with the T7 sequencing kit (Pharmacia LKB Biothechnologies). The single-stranded

DNA and the primers were mixed together in the annealing buffer and denatured at 95°C for 5 min. The samples were then quickly frozen on ice/water and the sequencing reactions were performed following the protocol provided by the manufacturer using [<sup>35</sup>S]dATP or [<sup>35</sup>S]dCTP as the labeled nucleotide.

All the sequence variants indicated in Table 1, with the exception of 4209–141C/A, were also screened by direct sequencing of PCR fragments covering the entire *BRCA1* coding region and intron–exon junctions as described (15,33,34) using DNA samples from ~200 patients with breast and/or ovarian cancer or known carriers by haplotype analysis. The differences observed in the number of patients tested for each sequence variant are due to sequencing gel variability and difficulty in sequencing some *BRCA1* gene regions.

#### Allele-specific oligonucleotide hybridization assay

The sequence variants Q/R356, R841W, L/P871, K/R1183, R1347G, 4209–141C/A, 4427C/T, S/G1613 and 5272+66A/G were screened using DNA samples purified as previously described (15) from at least 78 control individuals from our control population of Utah (Table 1). Pairs of oligonucleotides were designed for each sequence variant to include the variant nucleotide in the middle, when possible. The PCR products were generated as described (15) and quantified by electrophoresis on 2% agarose gels stained with ethidium bromide by comparison with DNA standards. PCR product (10 µl) was added to 110 µl of denaturing solution (0.067 M NaOH, 0.042 mM EDTA, 0.002% bromophenol blue) and incubated for 10 min at room temperature. Samples (30 µl) were then blotted onto Hybond-N membrane (Amersham) with a dot-blotting apparatus (Gibco-BRL). The DNA was immobilized by UV cross-linking. Pre-hybridization was performed at 45°C in 5× SSPE (0.75 M NaCl, 0.05 M NaH<sub>2</sub>PO<sub>4</sub>·H<sub>2</sub>O, 0.005 M EDTA) and 2% SDS. Wild-type and variant ASOs were end-labeled by incubation at 37°C for 10 min with 5 µCi [<sup>γ</sup>-<sup>32</sup>P]ATP, 100 ng of ASO, 10 U of T4 polynucleotide kinase (Boehringer-Mannheim) and kinase buffer. Twenty ng of labeled ASO was used in overnight hybridization reaction in the same buffer used for pre-hybridization. Each blot was washed twice in 5× saline sodium citrate and 0.1% SDS for 10 min at room temperature and then for 30 min at progressively higher temperatures until non-specific signals were eliminated. Blots were exposed to X-ray film for 40 min without intensifying screen.

#### Statistical analysis

Hardy–Weinberg equilibrium was tested by  $\chi^2$  analysis comparing observed genotype frequencies with those expected under Hardy–Weinberg equilibrium assuming the estimated allele frequencies. Pairwise linkage disequilibrium between each pair of the sequence variants indicated in Table 2 was assessed through analysis of joint genotype distribution. The standard linkage disequilibrium coefficient  $D = P_{11} - p_1q_1$  was calculated where  $P_{11}$  is the observed frequency of the 1–1 haplotype and  $p_1q_1$  is the corresponding expected frequency based on the estimated allele frequencies calculated from the sample of individuals on whom both polymorphisms were typed. Double heterozygotes whose haplotype status could not be determined were excluded from the analysis. These exclusions represented 3–20% of the data for the individual analyses. The values of  $D$  were transformed into a correlation coefficient,  $r$ , as in (52) by  $r = D/(p_1p_2q_1q_2)^{1/2}$  and

compared with the theoretical maxima based on the allele frequencies  $D_{\max}$  and corresponding  $r_{\max}$  where  $D_{\max} = \min(p_1q_2, p_2q_1)$  (53). The significance of the linkage disequilibrium coefficients was assessed by  $\chi^2$  analysis.

#### ACKNOWLEDGEMENTS

We would like to thank Dr Lynn Jorde from the Department of Human Genetics, University of Utah School of Medicine for helpful discussion on linkage disequilibrium. We would like to also thank Dr François Rousseau for providing DNA samples from anonymous Quebec city women. This work was supported by Medical Research Council of Canada grant to J. S., National Institute of Health grant CA-55914 to M. S., by Myriad Genetics Inc., and Endorecherche. J. S is a MRC scholar and F. D. is recipient of a studentship from the MRC.

#### REFERENCES

- Boring, C.C., Squires, T.S., Tong, T. and Montgomery, S. (1994) *Ca, Cancer J. Clin.*, **44**, 7–26.
- Claus, E.B., Risch, N. and Thompson, W.D. (1991) *Am. J. Hum. Genet.*, **48**, 232–241.
- Hoskins, K.F., Stopfer, J.E., Calzone, K.A., Merajver, S.D., Rebbeck, T.R., Garber, J.E. and Weber, B. (1995) *J. Am. Med. Assoc.*, **273**, 577–585.
- Schildkraut, J.M., Risch, N. and Thompson, W.D. (1988) *Am. J. Hum. Genet.*, **45**, 521–529.
- Cannon-Albright, L., Thomas, A., Goldgar, D., Gholami, K., Rowe, K., Jacobsen, M., McWorther, W.P. and Skolnick, M.H. (1994) *Cancer Res.*, **54**, 2378–2385.
- Hall, J.M., Lee, M.K., Newman, B., Morrow, J.E., Anderson, L.A., Huey, B. and King, M.C. (1990) *Science*, **250**, 1684–1689.
- Narod, S., Feunteun, J., Lynch, H.T., Watson, P., Conway, T., Lynch, J. and Lenoir, G. (1991) *Lancet*, **338**, 82–83.
- Simard, J., Feunteun, J., Lenoir, G., Tonin, P., Normand, T., Luu-The, V., Vivier, A., Lasko, D., Morgan, K., Rouleau, G.A., Lynch, H., Labrie, F. and Narod, S. (1993) *Hum. Mol. Genet.*, **2**, 1193–1199.
- Easton, D.F., Bishop, D.T., Ford, D., Crockford, G.P. and the Breast Cancer Linkage Consortium (1993) *Am. J. Hum. Genet.*, **52**, 678–701.
- Steichen-Gersdorf, E., Gallion, H.H., Ford, D., Girodet, C., Easton, D.F., DiCioccio, R.A., Evans, G., Ponder, M.A., Pye, C., Mazoyer, S., Noguchi, T., Kereguen, F., Sobol, H., Hardouin, A., Bignon, Y.-J., Piver, M.S., Smith, S.A. and Ponder, B.A.J. (1994) *Am. J. Hum. Genet.*, **55**, 870–875.
- Narod, S.A., Ford, D., Devilee, P., Bakardotir, R.B., Lynch, H.T., Smith, S.A., Ponder, B.A.J., Weber, B.L., Garber, J.E., Birch, J.M., Cornelis, R.S., Kelsell, D.P., Spurr, N.K., Smith, E., Haines, N., Sobol, H., Bignon, Y.-J., Claude, J.-C., Hamann, U., Linblom, A., Borg, A., Piver, M.S., Gallion, H.H., Struwing, J.P., Whittemore, A., Tonin, P., Goldgar, D., Easton, D.F. and the Breast Cancer Linkage Consortium (1995) *Am. J. Hum. Genet.*, **56**, 254–264.
- Easton, D. F., Ford, D., Bishop, D. T. and the Breast Cancer Linkage Consortium (1995) *Am. J. Hum. Genet.*, **56**, 265–271.
- Ford, D., Easton, D.F. and Peto, J. (1995) *Am. J. Hum. Genet.*, **57**, 1457–1462.
- Ford, D., Easton, D.F., Bishop, D.T., Narod, S. and Goldgar, D.E. (1994) *Lancet*, **343**, 692–695.
- Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P.A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L.M., Ding, W., Bell, R., Rosenthal, J., Hussey, C., Tran, T., McClure, M., Frye, C., Hattier, T., Phelps, R., Haugen-Strano, A., Katcher, H., Yakumo, K., Gholami, Z., Shaffer, D., Stone, S., Bayer, S., Wray, C., Bogden, R., Dayananth, P., Ward, J., Tonin, P., Narod, S., Bristow, P.K., Norris, F.H., Helvering, L., Morrison, P., Rosteck, P., Lai, M., Barrett, J.C., Lewis, C., Neuhausen, S., Cannon-Albright, L., Goldgar, D., Wiseman, R., Kamb, A. and Skolnick, M.H. (1994) *Science*, **266**, 66–71.
- Barlow, P.N., Luisi, B., Milner, A., Elliott, M. and Everett, R. (1994) *J. Mol. Biol.*, **237**, 201–211.
- Bennett, L.M., Haugen-Strano, A., Cochran, C., Brownlee, H.A., Fiedorek, F.T., Jr. and Wiseman, R.W. (1995) *Genomics*, **29**, 576–581.
- Lane, T.F., Deng, C., Elson, A., Lyu, M.S., Kozak, C.A. and Leder, P. (1995) *Gene Dev.*, **9**, 2712–2722.
- Friedman, L.S., Ostermeyer, E.A., Szabo, C.I., Dowd, P., Lynch, E.D., Rowell, S.E. and King, M.-C. (1994) *Nature Genet.*, **8**, 399–404.

20. Castilla, L.H., Couch, F., Erdos, M., Hoskins, K.F., Calzone, K., Garber, J., Boyd, J., Lubin, M.B., Deshano, M.L., Brody, L.C., Collins, F.S. and Weber, B. (1994) *Nature Genet.*, **8**, 387–391.
21. Friedman, L.S., Szabo, C.I., Ostermeyer, E.A., Dowd, P., Butler, L., Park, T., Lee, M.K., Goode, E.L., Rowell, S.E. and King, M.C. (1995) *Am. Hum. Genet.*, **57**, 1284–1297.
22. Smith, S.A., Easton, D.F., Evans, D.G.R. and Ponder, B.A.J. (1992) *Nature Genet.*, **2**, 128–131.
23. Kelsell, D.P., Black, D., Bishop, D.T. and Spurr, N.K. (1993) *Hum. Mol. Genet.*, **2**, 1823–1828.
24. Neuhausen, S.L. and Marshall, C.J. (1994) *Cancer Res.*, **54**, 6069–6072.
25. Futreal, P.A., Liu, Q., Shattuck-Eidens, D., Cochran, C., Harshman, K., Tavtigian, S., Bennett, L.M., Haugen-Strano, A., Swensen, J., Miki, Y., McClure, M., Frye, C., Weaver-Feldhaus, J., Ding, W., Gholami, Z., Soderkvist, P., Terry, L., Jhanwar, S., Berchuck, A., Iglehart J.D., Marks, J., Ballinger, D.G., Barrett, J.C., Skolnick, M.H., Kamb, A. and Wiseman, R. (1994) *Science*, **266**, 120–122.
26. Takahashi, H., Behbakht, K., McGovern, P.E., Chiu, H.-C., Couch, F.J., Weber, B.L., Friedman, L.S., King, M.-C., Furusato, M., LiVolsi, V.A., Menzin, A.W., Liu, P.C., Benjamin, I., Morgan, M. A., King, S.A., Rebane, B.A., Cardonick, A., Mikuta, J.J., Rubin, S.C. and Boyd, J. (1995) *Cancer Res.*, **55**, 2998–3002.
27. Matsushima, M., Kobayashi, K., Emi, M., Saito, H., Saito, H., Suzumori, K. and Nakamura, Y. (1995) *Hum. Mol. Genet.*, **4**, 1953–1956.
28. Merajver, S.D., Pham, T.H., Caduff, R.F., Chen, M., Poy, E.L., Cooney, K.A., Weber, B., Collins, F.S., Johnston, C. and Frank, T.S. (1995) *Nature Genet.*, **9**, 439–443.
29. Hosking, L., Trowsdale, J., Nicolai, H., Solomon, E., Foulkes, W., Stamp, G., Signer, E. and Jeffreys, A. (1995) *Nature Genet.*, **9**, 343–344.
30. Thompson, M.E., Jensen, R.A., Obermiller, P.S., Page, D.L. and Holt, J.T. (1995) *Nature Genet.*, **9**, 444–450.
31. Marquis, S.T., Rajan, J.V., Wynshaw-Boris, A. Xu, J., Yin, G.-Y, Abel, K.J., Weber, B.L. and Chodosh, L.A. (1995) *Nature Genet.*, **11**, 17–26.
32. American Society of Human Genetics (1994) *Am. J. Hum. Genet.*, **55**, i–iv.
33. Simard, J., Tonin, P., Durocher, F., Morgan, K., Rommens, J., Gingras, S., Samson, C., Leblanc, J.-F., Bélanger, C., Dion, F., Liu, Q., Skolnick, M., Goldgar, D., Shattuck-Eidens, D., Labrie, F. and Narod, S. (1994) *Nature Genet.*, **8**, 392–398.
34. Shattuck-Eidens, D., McClure, M., Simard, J., Labrie, F., Narod, S., Couch, F., Hoskins, K., Weber, B., Castillas, L., Erdos, M., Brody, L., Friedman, L., Ostermeyer, E., Szabo, C., King, M.-C., Jhanwar, S., Offit, K., Norton, L., Gilewski, T., Lubin, M., Osborne, M., Black, D., Boyd, M., Steel, M., Ingles, S., Haile, R., Lindblom, A., Olsson, H., Borg, A., Bishop, D.T., Solomon, E., Radice, P., Spatti, G., Gayther, S., Ponder, B., Warren, W., Stratton, M., Liu, Q., Fujimura, F., Lewis, C., Skolnick, M. and Goldgar, D. (1995) *J. Am. Med. Assoc.*, **273**, 535–541.
35. Struewing, J.P., Brody, L.C., Erdos, M.R., Kase, R.G., Giambarresi, T.R., Smith, S.A., Collins, F.S. and Tucker, M.A. (1995) *Am. J. Hum. Genet.*, **57**, 1–7.
36. Hogervorst, F.B.L., Cornelis, R.S., Bout, M., van Vliet, M., Oosterwijk, J.C., Olmer, R., Bakker, B., Klijn, J.G.M., Vasen, H.F.A., Meijers-Heijboer, H., Menko, F.H., Cornelisse, C.J., den Dunnen, J.T., Devilee, P. and van Ommen, G.-J.B. (1995) *Nature Genet.*, **10**, 208–212.
37. Inoue, R., Fukutomi, T., Ushijima, T., Matsumoto, Y., Sugimura, T. and Nagao, M. (1995) *Cancer Res.*, **55**, 3521–3524.
38. Neuhausen, S.L., Mazoyer, S., Friedman, L., Stratton, M., Offit, K., Colegio, A., Tomlinson, G., Cannon-Albright, L., Bishop, T., Kelsell, D., Weber, B., Couch, F., Struewing, J., Tonin, P., Durocher, F., Narod, S., Skolnick, M., Lenoir, G., Serova, O., Ponder, B., Stoppa-Lyonnet, D., Easton, D., King, M.C. and Goldgar, D. (1996) *Am. J. Hum. Genet.*, **58**, 271–280.
39. Plummer, S.J., Anton-Culver, H., Webster, L., Noble, B., Liao, S., Kennedy, A., Belinson, J. and Casey, G. (1995) *Hum. Mol. Genet.*, **4**, 1989–1991.
40. Boyd, M., Harris, F., McFarlane, R., Davidson, H.R. and Black, D.M. (1995) *Nature*, **375**, 541–542.
41. Durocher, F., Bélanger, C., Labrie, F., Tonin, P., Morgan, K., Narod, S., Shattuck-Eidens, D., Neuhausen S.L., Goldgar, D.E. and Simard, J. (1995) *Proceedings of The 2nd Joint Clinical Genetic Meeting*, Los Angeles, 108, Abstract # 64.
42. Cooper, D.N. and Krawczak (1990) *Hum. Genet.*, **85**, 55–74.
43. Krawczak, M., Smith-Sorensen, B., Schmidtke, J., Kakkur, V.V., Cooper, D.N. and Hovig, E. (1995) *Hum. Mutat.*, **5**, 48–57.
44. McLellan, T., Jorde, L.B. and Skolnick, M.H. (1984) *Am. J. Hum. Genet.*, **36**, 836–837.
45. Padgett, R.A., Grabowski, P.J., Kornasza, M.M., Seiler, S. and Sharp, P.A. (1986) *Annu. Rev. Biochem.*, **55**, 1119–1150.
46. Krawczak, M., Reiss, J. and Cooper, D.N. (1992) *Hum. Genet.*, **90**, 41–54.
47. Butler, L.M., Newman, B., Friedman, L.S., Ostermeyer, E.A., Morrow, J.E., Dowd, P., Moorman, P.G., Milikan, R. and King, M.-C. (1995) *Am. J. Hum. Genet.*, **suppl 57**, A61, Abstract #319.
48. Rousseau, F., Rouillard, P., Morel, M.-L., Khandjian, E.W. and Morgan, K. (1995) *Am. J. Hum. Genet.*, **57**, 1006–1018.
49. Rousseau, F., Réhel, R., Rouillard, P., Degranpré, P. and Khandjian, E.W. (1994) *Hum. Mutat.*, **4**, 51–54.
50. Charbonneau, H. and Robert, N. (1987) The French origins of the Canadian population, 1608–1759. In Harris, R.C. (ed.), *Historical Atlas of Canada. Vol 1: From the Beginning to 1800*. University of Toronto Press, Plate 45.
51. Bouchard, G. and De Braekeleer, M. (1990) *Social History*, **23**, 325–361.
52. Hill, W.G. and Robertson, A. (1968) *Theor. Appl. Genet.*, **38**, 226–231.
53. Litt, M. and Jorde, L.B. (1986) *Am. J. Hum. Genet.*, **39**, 166–178.
54. Beaudet, A.L. and Tsui, L.-C. (1993) *Hum. Mutat.*, **2**, 245–248.