Mutation analysis and characterization of HSD17B2 sequence variants in breast cancer cases from French Canadian families with high risk of breast and ovarian cancer

Marie Plourde¹, Caroline Manhes¹, Gilles Leblanc¹, Francine Durocher¹, Martine Dumont¹, Olga Sinilnikova², INHERIT BRCAs* and Jacques Simard¹,³

¹Cancer Genomics Laboratory, Oncology and Molecular Endocrinology Research Center, Centre Hospitalier Universitaire de Québec and Laval University, Quebec G1V 4G2, Canada
²Plate-Forme Mixte de Génétique Constitutionnelle des Cancers Fréquents, Hospices Civil de Lyon/centre Léon Bérard, 69373 Lyon Cedex 08, France
³Canada Research Chair in Oncogenetics

Abstract

Estrogen exposure is a risk factor for breast cancer. Given that HSD17B2 gene encodes an enzyme that catalyses estradiol inactivation, it appears as a good candidate breast cancer susceptibility gene. This study was designed to screen for HSD17B2 germline mutations potentially involved in breast cancer predisposition. Our re-sequencing analysis did not identify any deleterious germline mutations, and therefore mutations in HSD17B2 do not explain the clustering of breast cancer cases in non-BRCA1/2 high-risk French Canadian families. However, six sequence variants were identified, including two novel missense variants. Expression assays revealed that p.Ala111Asp and p.Gly160Arg did not alter the catalytic properties of 17β-hydroxysteroid dehydrogenase type 2 enzyme, although p.Ala111Asp appears to affect protein stability resulting in significant decreases in the protein levels, providing valuable information on structure–function relationship.

Journal of Molecular Endocrinology (2008) 40, 161–172

Introduction

Most of the familial aggregation in breast cancer results predominantly from inherited susceptibility (Lichtenstein et al. 2000, Peto & Mack 2000). Germline mutations in high-penetrance cancer susceptibility genes such as breast cancer 1 gene (BRCA1), breast cancer 2 gene (BRCA2), tumor protein p53 (TP53), phosphatase and tensin homolog (PTEN), and ataxia telangiectasia mutated gene (ATM) are associated with some of the familial breast cancer cases, although this association accounts for ~25% of the familial component of breast cancer risk (Pharoah et al. 2004, Thompson & Easton 2004, Antoniou & Easton 2006).

It has been proposed that a complex polygenic model may be the best explanation for this missing genetic risk (Antoniou et al. 2001). Under the common variant/common disease (CV/CD) model, disease susceptibility is suggested to result from the joint action of several CVs, with unrelated affected individuals sharing a substantial proportion of disease alleles (Cargill & Daley 2000, Reich & Lander 2001, Lohmueller et al. 2003, Easton et al. 2007). The alternative is the heterogeneity hypothesis, which maintains that genetic susceptibility to CD is caused by many different rare genetic variants with a relatively large effect produced by each allele (Pritchard 2001, Pritchard & Cox 2002, Fearnhead et al. 2004). This hypothesis will be more easily detectable by the candidate gene re-sequencing approach (Renwick et al. 2006, Seal et al. 2006, Rahman et al. 2007).

Both epidemiological and cell biology experiments have documented the contribution of estrogens in the regulation of cellular growth, differentiation and proliferation of the normal mammary gland as well as in hormone-sensitive breast carcinomas, while androgens rather exert an anti-proliferative action in these tissues (Doisneau-Sixou et al. 2003, Labrie et al. 2003).

There is a large body of evidence showing that in humans, the local intracrine formation of active estrogens and androgens from inactive steroid precursors secreted by the adrenals, namely, dehydroepiandrosterone and its sulfate, regulates the growth and function of peripheral target tissues, including the breast (Labrie et al. 2003).

Among the enzymes modulating local sex steroid availability, 17β-hydroxysteroid dehydrogenase type 2...
enzyme (17β-HSD) that is responsible for the inactivation of estradiol (the more potent estrogen) into estrone, is located on chromosome 16q23.3-q24.1, a genomic region showing frequent loss of heterozygosity in breast cancer (Suzuki et al. 1996, Sato et al. 1998, Cleton-Jansen et al. 2001). For oxidative reactions in intact cells, 17β-HSD type 2 prevails in its kinetic parameters over other forms of 17β-HSDs (Wu et al. 1993, Labrie et al. 1995, Mindnich et al. 2004). HSD17B2 is expressed both in normal breast tissue and in malignant breast tumors and it has been suggested that it may protect normal breast cells from the effects of excessive estradiol production (Casey et al. 1994, Miettinen et al. 1999, Gunnarsson et al. 2001, Oduwole et al. 2004). It should be noted that tumors frequently lack detectable levels of 17β-HSD type 2 and in cases where its expression is retained, there is evidence of a decreased risk for late relapse of breast cancer (Gunnarsson et al. 2001). Among estrogen receptor-positive patients, those with a low expression of 17β-HSD type 2 had a significantly higher recurrence rate when compared with patients who expressed normal levels (Gunnarsson et al. 2005). An increase or decrease in this enzyme activity may therefore alter the levels of endogenousestro gens, thereby influencing breast cancer susceptibility.

Studies investigating the role of genetic variants in the HSD17B1 gene, which gives rise to 17β-HSD type 1 enzyme that predominantly catalyses the reduction of estrone to estradiol, have been conducted in relation to breast cancer risk with non-conclusive results (Mannermaa et al. 1994, Feigelson et al. 2001, 2006, Wu et al. 2003, Setiawan et al. 2004, Plourde et al. 2007). On the other hand, when this study began, there were no studies investigating the presence of HSD17B2 germ line mutations or polymorphisms in families with high risk of breast cancer. Based on the pivotal role of HSD17B2 in modulating local sex steroid availability in the breast tissue, HSD17B2 represents an attractive candidate gene to potentially explain a fraction of the remaining familial component of breast cancer risk. We therefore performed comprehensive re-sequencing of the HSD17B2 gene in women affected with breast cancer from non-related BRCA1- and BRCA2-negative French Canadian families with a high risk of breast or ovarian cancer.

Materials and methods

Ascertainment of high-risk families and DNA extraction

The recruitment of French Canadian families with a high risk of breast and ovarian cancer started in 1996 through a research project, which thereafter evolved in a large ongoing interdisciplinary research program designated INHERIT BRCAs. More details regarding ascertainment criteria, experimental and clinical procedures, as well as the INHERIT BRCAs research program have been described elsewhere (Simard et al. 2007). A major component was to identify and characterize the prevalence and penetrance of BRCA1 and BRCA2 mutations in French Canadian high-risk families (Antoniou et al. 2006, Simard et al. 2007). Subsequently, another component was designed for the ‘localization and identification of new breast cancer susceptibility loci/genes’. Ethics approval was obtained from the different institutions participating in this research component and each participant had to sign an informed consent to participate in this latter project (Durocher et al. 2006). A subset of 50 high-risk French Canadian breast/ovarian cancer families were recruited in the present study according to the presence of multiple cases of breast cancer, among which 45 included at least 3 individuals with breast cancers within second degree relatives, while 5 families included 3 or more individuals within third degree relatives. All participants had to be at least 18 years of age and mentally capable. The diagnoses of breast and/or ovarian cancer were confirmed by a pathology report, and when two or more subjects were available within a family, the youngest subject was systematically chosen for this study. The mean age at diagnosis of these 50 subjects affected with breast cancer was 47.7 years (34–69 years), while 35 of them have been diagnosed before 50 years of age and 8 had bilateral breast cancer.

The BRCA1/2 status of each participant was assessed previously (Simard et al. 2007). The evidence of the absence of genomic rearrangements in BRCA1/2 genes was thereafter investigated by multiplex ligation-dependent probe amplification (MLPA) for 44 out of the 50 subjects (BRCA1/2 Southern analyses were done for 32 out of these 50 subjects). For four of the remaining subjects, MLPA was performed on another individual of the family (Moisan et al. 2006), while for two subjects this analysis was not performed in their family.

Genomic DNA from 73 healthy unrelated French populations was obtained from Dr Damian Labuda at the Centre de Cancérologie Charles Bruneau, Hôpital Ste-Justine, Montreal, Canada. The individuals who provided these samples were recruited on a non-nominative basis, in the framework of long-term studies aiming the characterization of the genetic variability in human populations, approved by the Institutional Ethic Review Board. The mean age of these individuals was 43.3 years: 4 (5.5%), 24 (32.9%), 27 (37.0%), 13 (17.8%), and 5 (6.8%) of them were between 25–29, 30–39, 40–49, 50–59, and 60–69 years of age respectively.

PCR amplification, mutation analysis, and sequence variant identification

The complete sequence of the five coding exons, flanking intronic sequences, 5′ and 3′ non-coding regions and 1000 bp of the promoter region of the HSD17B2 gene was
analyzed in DNA samples obtained from Epstein–Barr virus (EBV)-transformed B-lymphoblastoid cell lines of 50 affected individuals. Five primer pairs were used to amplify the fragments that were sequenced in both directions with primers indicated in Table 1. The frequency of variants was also ascertained in 73 healthy unrelated French Canadians. After the completion of this study, Jansson et al. (2007) published their results from the HSD17B2 mutation screening among breast cancer cases from Sweden, which reported a previously unknown missense variant (p.Met226Val) not observed among our affected individuals. The presence of this variant was subsequently ascertained in 69 out of our 73 healthy controls. HSD17B2 direct sequencing was performed on an ABI Prism 3730 DNA Analyser automated sequencer using version 3.1 of the Big Dye fluorescent method according to the manufacturer’s instructions (Applied Biosystems, Foster City, CA, USA). Sequence data were analyzed using the Staden preGap4 and Gap4 programs. The GenBank accession number for the region of the HSD17B2 gene including the polymorphic CA repeat (single nucleotide polymorphism, SNP 1) was amplified using the following primers: forward, 5′-GTAGGAGTCTTGAACTATGGG-3′ and reverse, 5′-GGTTTTCTGAACCTGATGGAGATA-ACAG-3′ of which the forward primer was labeled with Applied Biosystems fluorescent dye (5-carboxy-fluorescein; FAM). Typically, 20 ng genomic DNA was used in each PCR of 15 μl. The PCR protocol involved a denaturation step at 95 °C for 5 min and then 35 cycles of 95 °C for 30 s followed by 40 s at 57 °C, with a final elongation cycle of 72 °C for 8 min. Each reaction contained 2.3 mM MgCl₂, 1 U platinum Taq polymerase (Invitrogen), 1× PCR buffer without MgCl₂, 0·1 mM dNTP and 2 pmol of each primer. PCR products were diluted to obtain a 1:250 ratio. From this dilution, 2 μl were added to 10 μl Applied Biosystems HiDi-formamide containing Genescan 500 LIZ (3:1000 dilution). Samples were resolved on an ABI Prism 3100 Genetic Analyzer and data were analyzed using the GeneMapper Version 2 software (Applied Biosystems). Several samples were sequenced to confirm the repeat length as established by the ABI 3100 Genotyper.

### LD analysis, haplotype estimation, and in silico analysis tools

To estimate the pattern of linkage disequilibrium (LD), we used the LDA program (http://www.chgb.org.cn/lda/lda.htm; Ding et al. 2003), which calculates pairwise LD for the common sequence variant pairs, Lewontin’s standardized disequilibrium coefficient $D^j$ (Lewontin 1964, Devlin & Risch 1995) and the square of correlation coefficient $r^2$ (Hill & Robertson 1968) were used as measures of LD between variants. Haplotype reconstructions and frequency estimations were performed using the PHASE 2.1.1 software (http://www.stat.washington.edu/stephens/; Stephens et al. 2001). This program (PHASE) estimates haplotype frequencies with a Bayesian-based algorithm and then uses a permutation test to determine the significance of differences in the inferred haplotypes between cases and controls. Computations were made using the software’s default parameters.

Splice site prediction scores were evaluated using splice site prediction by neural network (SSPN; http://www.fruitfly.org/seq_tools/splice.html) and Alex Dong Li’s splice site finder (http://www.genet.sickkids.on.ca/ali/splicesitefinder.html; Shapiro & Senapathy 1987), while the effect of amino acid substitutions was evaluated using PolyPhen (http://genetics.bwh.harvard.edu/pph/; Ramensky et al. 2002).

### Site-directed mutagenesis

The following oligonucleotide sequences were designed such that the desired mutation was in the

---

**Table 1** Oligonucleotide primers used for amplification and sequence analysis of the HSD17B2 gene

<table>
<thead>
<tr>
<th>Amplified regiona</th>
<th>Forward primer (5′–3′)</th>
<th>Reverse primer (5′–3′)</th>
<th>Product length (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Promoter and exon 1</td>
<td>GCATTATGAGGTCTTCTTACTGb</td>
<td>CAAGTTCTCAGATCATACCCCTAGTb</td>
<td>1846</td>
</tr>
<tr>
<td>Exon 2</td>
<td>GAATGCCTAGGGGGCATACAGe</td>
<td>GCACCTTGATCTGTCATAGe</td>
<td></td>
</tr>
<tr>
<td>Exon 3</td>
<td>AGACCCGTTCTAGGGGCTTCTG</td>
<td>GTACTCAGGAGTCTTCTAGG</td>
<td>491</td>
</tr>
<tr>
<td>Exon 6</td>
<td>TGCTGTGCTCCAGGTATTG</td>
<td>AACCCTGCCCTTGTGCTCGG</td>
<td>371</td>
</tr>
<tr>
<td>Exon 7</td>
<td>AGGCCGCCATCTCTTTCCAGAGACAG</td>
<td>TGAGGGGTCAATGGTGGTGCCTG</td>
<td>307</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CCTCAATTCTCAGTGAACACTAG</td>
<td>699</td>
</tr>
</tbody>
</table>

---

*a Annealing temperature at 60 °C.
*b Used as amplification and sequencing primers.
*e Used as sequencing primer only.
middle of the primer with about 10–15 bases of specific sequence on either side. The primer sequences were as follows. For the C to A mutation of c.332G>A: forward (5'-GCTTCACGGTATTTTGAGGATTTTGAATG-3’) and reverse (5'-CATTTAAAAACTCCGTCAAATACCGTGGAAGC-3’); for the G to C mutation of c.478G>C: forward (5'-GCAGGACAGACGACTGTGGGCG-3’) and reverse (5'-GCCCACAGTCTGTCTCTCGC-3’). Site-directed mutagenesis was performed on an expression vector containing the full-length cDNA fragment encoding the complete amino acid sequence for human 17β-HSD type 2 (kindly provided by Dr Van Luu-The, CHUL Research Center, Quebec, Canada) inserted in the pcDNA3.1 vector (Invitrogen) using the QuickChange Site-directed mutagenesis kit from Stratagene (Stratagene Cloning Systems, La Jolla, CA, USA) according to the supplier’s protocol. Insertion of the mutation was confirmed by direct sequencing using Big Dye Terminator chemistry on an ABI Prism 3730 automated sequencer from Applied Biosystems.

Transcription/translation

Transcription/translation was performed using the TNT Quick coupled transcription/translation system from Promega according to the manufacturer’s instructions, as described previously (Plourde et al. 2007).

Cell culture and transfection

Human embryonic kidney 293 cells (HEK293) were obtained from the American Type Culture Collection (ATCC; Manassas, VA, USA) and were cultured in Dulbecco’s modified Eagle’s medium (DMEM)/high glucose from Invitrogen Life Technologies Inc. supplemented with 10% fetal bovine serum (FBS), 1% glutamine, 100 IU/ml penicillin, and 50 mg/ml streptomycin from WISENT Inc. (St-Bruno, Quebec, Canada) until confluence. The cells were passaged and plated in 175 cm² culture dishes at a density of 9·1×10⁶ cells. The cells were allowed to settle overnight after which the medium was changed to DMEM without FBS just before transfection. Transient transfection was performed using ExGen 500 cationic polymer transfection reagent (MBI Fermentas Inc., Ontario, Canada) according to the supplier’s protocol. The cells were also transfected with the pcDNA3 vector as a negative control. One day after transfection, the cells were scrapped to assess enzyme activity, as described below.

Western analysis

Western analysis of proteins was performed by SDS-PAGE on discontinuous acrylamide gels, as described previously (Plourde et al. 2007). Briefly, total proteins were extracted from HEK293 cells in PBS and cell lysis was performed using standard procedures. Total proteins (15 µg) were separated on a 4% stacking and 12% resolving gel and transferred onto a nitrocellulose membrane (Hybond-ECL, Amersham Pharmacia Biotech Inc.) and thereafter hybridized to a polyclonal antibody directed against human 17β-HSD type 2 (kindly provided by Dr Van Luu-The) at a dilution of 1:1000 and subsequently incubated with a donkey anti-rabbit immuno-globulin G (IgG) peroxidase-conjugated secondary antibody (Amersham Pharmacia Biotech Inc.) at a dilution of 1:10 000. Membranes were washed and proteins were visualized using the SuperSignal West Dura Extended Duration Substrate kit (PIERCE, Rockford, IL, USA), followed by exposure of the membranes to X-ray films for 1–10 min. The autoradiographic film was scanned and the ImageJ program (NIH, Bethesda, MD, USA) was used to quantify the density of the autoradiographic bands.

Assay of 17β-HSD type 2 enzymatic activity

In order to measure 17β-HSD type 2 activity, the amount of translated mutant recombinant proteins have been adjusted to be equivalent to 10 µg total protein of normal 17β-HSD type 2 enzyme in each corresponding cell homogenate preparation, as measured previously by immunoblot analysis. The cell homogenates were incubated for 0·5–4 h at 37°C in the presence of 10 nM [4-¹⁴C]-estradiol (52·0 mCi/mmol) from Perkin–Elmer Life Sciences Inc. (Boston, MA, USA), 1 mM NADH (NAD⁺) cofactor, and a buffer containing 50 mM NaH₂PO₄ (pH 7·4), 20% glycerol (vol/vol) and 1 mM EDTA. After the indicated time intervals, steroids were extracted by the addition of 2 volumes of diethyl ether and the incubation mixture was chilled in a dry ice/ethanol bath. Steroids were separated by thin layer chromatography using a mobile phase of toluene:acetone (4:1) and analyzed using phosphoimaging, Storm 860 (Molecular Dynamics Inc., Sunnyvale, CA, USA). All results are expressed as the mean±S.E.M. of at least three separate transfection experiments performed in triplicate.

Results

HSD17B2 mutation analysis and variant characterization

Although no truncating mutation was found in the HSD17B2 coding region of our French Canadian breast cancer cases, we identified six variants in HSD17B2 exonic, flanking intronic, and promoter sequences (Fig. 1A).
These included a CA repeat located in the promoter region (arbitrarily defined as a 1000 bp genomic segment upstream exon 1), a common nucleotide substitution in the 5’-untranslated region (c.-133C>T), two rare non-synonymous variants in exon 2 (c.332C>A (p.Ala111Asp) and c.478G>C (p.Gly160Arg)), one common substitution in intron 6 (c.803-57C>G) and one rare synonymous sequence variant located in exon 7 (c.936G>A; Table 2). Two were novel, while four were reported in the SNP databases (dbSNP; www.ncbi.nlm.nih.gov/SNP). Genotype and minor allele frequencies (MAF) were determined in cases and controls, both of French Canadian origin. Frequencies were similar with frequencies reported in dbSNP database. As indicated in Table 2, the distribution of all genotypes was in accordance with Hardy-Weinberg equilibrium.

**Effect of missense substitutions p.Ala111Asp and p.Gly160Arg on 17\(\beta\)-HSD activity and expression levels**

Two non-synonymous SNPs were observed in the coding sequence of HSD17B2 (SNP #3 and SNP #4), and both were only observed once at a heterozygous state (Table 2). These two cSNPs resulted in changes in encoded amino acids, p.Ala111Asp and p.Gly160Arg. Segregation of the missense variants with breast cancer was analyzed in...
<table>
<thead>
<tr>
<th>SNP ID</th>
<th>dbSNP ID</th>
<th>Location</th>
<th>Series</th>
<th>Genotype frequencies</th>
<th>MAF</th>
<th>$\chi^2 P$ value</th>
<th>Reported MAF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>g.35682736(CA)4-9</td>
<td>rs10595118</td>
<td>Promoter</td>
<td>Cases</td>
<td>0.37 (0.37)</td>
<td>0.49 (0.48)</td>
<td>0.14 (0.15)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Controls</td>
<td>0.39 (0.36)</td>
<td>0.43 (0.48)</td>
<td>0.18 (0.16)</td>
</tr>
<tr>
<td>2</td>
<td>g.35683097C&gt;T</td>
<td>rs4445895</td>
<td>5'UTR of exon 1</td>
<td>Cases</td>
<td>0.36 (0.37)</td>
<td>0.50 (0.48)</td>
<td>0.14 (0.15)</td>
</tr>
<tr>
<td></td>
<td>c.-133C&gt;T</td>
<td></td>
<td></td>
<td>Controls</td>
<td>0.40 (0.41)</td>
<td>0.48 (0.46)</td>
<td>0.12 (0.13)</td>
</tr>
<tr>
<td>3</td>
<td>g.35716040G&gt;A</td>
<td>N/A</td>
<td>Exon 1</td>
<td>Cases</td>
<td>0.98 (0.98)</td>
<td>0.02 (0.02)</td>
<td>0.00 (0.00)</td>
</tr>
<tr>
<td></td>
<td>c.332C&gt;A</td>
<td></td>
<td></td>
<td>Controls</td>
<td>1.00 (1.00)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
</tr>
<tr>
<td>4</td>
<td>g.35716186G&gt;C</td>
<td>N/A</td>
<td>Exon 2</td>
<td>Cases</td>
<td>0.98 (0.98)</td>
<td>0.02 (0.02)</td>
<td>0.00 (0.00)</td>
</tr>
<tr>
<td></td>
<td>c.478G&gt;C</td>
<td></td>
<td></td>
<td>Controls</td>
<td>1.00 (1.00)</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
</tr>
<tr>
<td>5</td>
<td>g.35745822C&gt;G</td>
<td>rs8191243</td>
<td>Intron 6</td>
<td>Cases</td>
<td>0.40 (0.44)</td>
<td>0.52 (0.45)</td>
<td>0.08 (0.12)</td>
</tr>
<tr>
<td></td>
<td>c.803-57C&gt;G</td>
<td></td>
<td></td>
<td>Controls</td>
<td>0.48 (0.50)</td>
<td>0.46 (0.41)</td>
<td>0.06 (0.08)</td>
</tr>
<tr>
<td>6</td>
<td>g.35746012G&gt;A</td>
<td>rs8191245</td>
<td>Exon 7</td>
<td>Cases</td>
<td>0.98 (0.98)</td>
<td>0.02 (0.02)</td>
<td>0.00 (0.00)</td>
</tr>
<tr>
<td></td>
<td>c.936G&gt;A</td>
<td></td>
<td></td>
<td>Controls</td>
<td>0.99 (0.98)</td>
<td>0.01 (0.02)</td>
<td>0.00 (0.00)</td>
</tr>
</tbody>
</table>

N/A, information not available for this SNP (not reported in dbSNP). MAF, minor allele frequency.

---

1. According to the nomenclature of the Human Genome Variation Society, for cDNA numbering +1 corresponds to the A of the ATG translation initiation codon in the reference sequence.
2. Observed among 100 chromosomes from 50 breast cancer cases and 146 chromosomes from 73 women controls DNA samples.
3. As expected under Hardy–Weinberg equilibrium.
4. $P$ value for deviation from Hardy–Weinberg equilibrium (Pearson’s $\chi^2$).
5. Minor allele frequencies are reported from www.hapmap.org using the CEPH population and www.ensembl.org with the Perlegen European panel.
6. (CA)$_p$: common allele, (CA)$_b$: rare allele; (CA)$_p$ was observed once in controls while (CA)$_b$ was observed twice in cases, these data are not included in calculated frequencies.
respective families in order to determine their potential involvement in the modification of breast cancer risk (data not shown). In the p.Ala111Asp variant carrier’s family, the variant does not segregate with breast cancer, in fact, the participant’s sister was also diagnosed with breast cancer at 42 years of age and does not carry the mutation. The same observation applies to the p.Gly160Arg variant carrier’s family in which the participant’s sister has also developed breast cancer at an early age (38 years), while she is a common homozygote Gly/Gly. These observations thus do not support the implication of these sequence variants in the clustering of breast cancer in these high-risk families.

In order to obtain more information on structure–function relationships, we performed several analyses for these missense variants. The two amino acid substitutions are located outside the known sequence motifs of the short chain dehydrogenase/reductase family of which 17β-HSD type 2 protein is a member (Fig. 1A; Oppermann et al. 2001). However, in order to obtain a more representative prediction of the importance of specific residues on protein function, comparison of missense substitutions was performed across relevant species. Alignment of 17β-HSD type 2 ortholog sequences revealed that p.Ala111 and p.Gly160 are conserved amino acids in vertebrates thus suggesting that these positions are under strong functional constraint (Fig. 1B). Furthermore, for the p.Gly160Arg variant, unlike arginine, glycine is not a charged residue and has a very different structure, while for the p.Ala111Asp variant, alanine and asparagine are both hydrophobic and charged amino acids. In silico analysis revealed that for both variants, the change in amino acid is predicted to be damaging to protein structure (Ng & Henikoff 2001, Ramensky et al. 2002). Moreover, p.Gly160Arg (c.478G>C) is located at the boundary of an exon–intron splice recognition sequence. We evaluated in silico the potential consequence of this variant on splicing by calculating the splicing efficiencies in the normal and mutant sequences with two splice prediction programs: the SSPNN from the Berkeley Drosophila Genome Project and Alex Dong Li’s Splice Site Finder (Shapiro & Senapathy 1987). Both programs predicted that this SNP abolished the normal donor splicing site of exon 2, suggesting a loss of the splicing-site recognition property. In order to further investigate whether the p.Gly160Arg (p.478G>C) variant may lead to alternative splicing on mRNA, we performed a PCR on cDNA from immortalized cell lines from the heterozygous individual as well as from wild-type individuals, in the region surrounding the p.Gly160Arg (c.478G>C) variant. No transcript of wild-type HSD17B2 could be detected in immortalized cell lines and therefore the effect of the p.Gly160Arg (c.478G>C) variant on splicing could not be confirmed. Note that such in silico analysis was also carried out to investigate the potential effect of SNP #5 (c.803-57C>G) located in intron 6, on splicing. This analysis predicted that this variant had no effect on splicing-site recognition properties.

Functional assays were performed to investigate the effect of the amino acid substitutions resulting from the two non-synonymous variants on 17β-HSD type 2 recombinant products. In vitro transcription/translation assays show normal expression levels of both wild-type and mutant recombinant p.Ala111Asp and p.Gly160Arg proteins (Fig. 2A). A 43 kDa band corresponding to 17β-HSD type 2 protein was detected by western blot analysis using cells expressing the wild-type and p.Gly160Arg variant, whereas a very weak signal was detectable for the mutant recombinant protein p.Ala111Asp (Fig. 2B). Similar expression levels were observed in at least five independent experiments, as well as in the presence of protease and protease inhibitors (data not shown). We also investigated the time course formation of [14C]estrone from [14C]-estradiol in transfected homogenated HEK293 cells. As illustrated in Fig. 2C, the apparent activities of p.Ala111Asp and p.Gly160Arg enzymes were similar to that of the wild-type enzyme. These results for p.Ala111Asp indicate that this substitution may probably yield a much less stable enzyme that could be rapidly degraded, but that otherwise does not affect its enzymatic capacities when the amount of translated mutant recombinant protein is corrected to be equivalent to the wild-type.

LD and haplotype analyses

Tagging SNP analysis is a way of maximizing the study of genetic variability irrespective of the functionality of individual SNPs. Indeed, haplotype prediction and tSNP selection using common SNPs (MAF ≥5%) provides good power to capture common variations in a gene and therefore allows the use of these tSNPs in studies involving other populations. Furthermore, since all genotyped SNPs are in strong LD with one of the tSNPs, although the causal variant may not be genotyped, it will be indirectly represented by one of the tSNPs. The estimation of haplotypes may also allow the observation of a phenotype associated with a combination of SNPs rather than with a single SNP.

In order to do so, we must first determine LD values between SNPs and infer major haplotypes, to determine SNPs that will tag these haplotypes. Pairwise LD values involving the two CVs, SNPs 2 and 5 (excluding the CA repeat), were calculated, while LD values for pairs with variants 3, 4, and 6 were not since each one was observed in only a single heterozygous case. When measured by Lewontin’s $D^r$, these sequence variants were in strong LD with each other ($D^r=0.61$ for breast cancer cases and $D^r=1.00$ for controls) even with an inter-marker distance of ~63 kb, suggesting that LD
Figure 2 Comparison of expression levels and stability of mutant recombinant 17β-HSD type 2 proteins.
(A) Representation of an in vitro transcription/translation (TNT) rabbit reticulocyte lysate assay showing that each pcDNA3 construct is adequately translated into a [35S]-labeled-43 kDa protein, indicative of normal expression levels of mutant recombinant 17β-HSD type 2 proteins. Translation was assessed by separation on a 12% SDS-PAGE gel. (B) Western blot analysis of homogenates purified from the corresponding HEK293 cells transiently transfected with the indicated expression vectors. A 43 kDa band corresponding to 17β-hydroxysteroid dehydrogenase type 2 protein is detected in homogenate preparations from HEK293 transfected cells expressing wild-type and mutant recombinant protein p.Gly160Arg, while a weak signal is detected in cells transfected with the mutant recombinant protein p.Ala111Asp. No 43 kDa protein is observed in mock pcDNA3 vector homogenates. The nonspecific band observed may be used as an internal control for loading. (C) Comparison of the time course of enzymatic conversion of [14C]-estradiol into [14C]-estrone in HEK293 cell homogenates transfected with the indicated expression vectors after normalization of the amount of immunoreactive mutant recombinant proteins to equivalent normal 17β-HSD type 2 protein expression for each cell homogenate preparation. The results are presented as the mean ± S.E.M. (n=3) and when the S.E.M. overlaps with the symbol used, only the symbol is illustrated. The cells were transfected with the pcDNA3 vector alone to show the absence of endogenous 17β-HSD type 2 mRNA expression.
at the HSD17B2 locus does not decrease significantly with distance. Seven haplotypes were estimated by PHASE from the six sequence variants found in the French Canadian population and haplotype frequencies were similar in cases and controls (data not shown). Our results, however, should be taken with caution since our study sample is of modest size and large sample sizes are required to obtain reliable results. Out of the seven haplotypes, only three exhibited a frequency of ≥5%, representing 93.1% of all observed haplotypes. Variants #2 and #5 can be used as tSNPs to predict these common haplotypes in future large-scale association studies.

Discussion

Since the discovery of susceptibility genes, such as BRCA1 and BRCA2, that harbor many rare high-risk variants, failure to find strong evidence of new linkages has led some to argue that few genes harboring high-risk breast cancer susceptibility alleles remain to be identified, although there could still be individually rare variants in a number of genes that confer high risk. Homogenous family selection through close attention to tumor phenotype and/or from more homogenous populations, such as the French Canadian founder population, increases the likelihood of identifying genes harboring true high-risk variants. Based on the pivotal role of HSD17B2 gene product in the estrogen metabolism pathway, it may be considered as a putative candidate gene that could possibly explain a fraction of the remaining familial breast cancer risk. However, no deleterious germline mutations leading to a premature termination of the protein were identified in the coding region. Therefore, HSD17B2 is unlikely to play a role as a high penetrance gene in breast cancer predisposition.

On the other hand, it is likely that much of the unexplained familial risk is due to alleles of low to moderate penetrance. The genetic association study is a powerful tool for identifying such alleles. Although six sequence variants were identified in this study, among which two are novel missense variants, the possible involvement of variants or haplotypes observed in cases compared with those found in controls would need to be analyzed in a much larger sample set (Pharoah et al. 2002, Antoniou & Easton 2006, Khoury et al. 2007). Indeed, until recently, most studies lacked the power to detect moderate relative risks at stringent levels of statistical significance, thus large sample sizes are needed to detect and confirm genetic variants that are associated with modest increases in risk. The ability to identify such genetic variants can be further improved by careful selection of both candidate gene and candidate polymorphism (Cox et al. 2007, Easton et al. 2007, Khoury et al. 2007). We thus sought to identify tagging SNPs that could be useful to other studies and populations. Our $|D'|$ values of LD should be interpreted with caution because a small sample size tends to show higher levels of LD based on the mean of $|D'|$ (Teare et al. 2002). On the other hand, variants #2 ($c.-133C>T$) and #5 ($c.803-57C>G$) are frequent variants and are in strong LD, even if they are distantly spaced over 63 kb. These findings suggest that HSD17B2 gene is a region of strong LD with limited haplotype diversity and these sequence variants may be used as markers in case–control association studies using large cohorts to investigate their potential association with breast cancer risk (Pharoah et al. 2004).

Recently, after the completion of this study, Jansson et al. (2007) published their results from the HSD17B2 mutation screening among breast cancer cases from Sweden. No mutation was detected, but they did identify a previously unknown missense variant (p.Met226Val). They observed 5 and 15 heterozygotes for this variant among the 122 ‘hereditary’ breast cancer cases and 457 controls respectively. This missense variant was not observed among the 50 affected individuals of the present study. In light of this very recent report, the presence of this variant was also ascertained in 69 French Canadian healthy controls and only one heterozygote was found. Further studies using larger sample sets will be needed to confirm that the p.Met226Val rare sequence variant is less frequent among the French Canadian population compared with the Swedish population.

In conclusion, no deleterious germline mutation potentially involved in breast cancer susceptibility has been identified in this French Canadian sample set. Indeed, data obtained from our functional assays taken together with the absence of co-segregation with breast cancer in family members do not support the association of p.Gly160Arg and p.Ala111Asp sequence variants with clustering of breast cancer in these families. Nevertheless, the possibility that a common sequence variant might have a differential effect in another ethnic group via gene–gene or gene–environment interactions, or that a predisposing sequence variant might be present exclusively in another population can not be excluded. The present study also demonstrates that p.Ala111Asp sequence variant seems to affect protein stability, a situation reminiscent of mutations reported in 3β-hydroxysteroid dehydrogenase/Δ5-Δ4 isomerase type 2 gene (Moisan et al. 1999, Simard et al. 2005). The in vivo functional implications of this observation remain to be determined, as are the possible functional consequences of the other observed sequence variants. Considering the importance of HSD17B2 gene, the complexities of the susceptibility to breast cancer, and
the limitations of our study, additional population studies using larger cohorts are warranted to define the possible association of HSD17B2 gene sequence variants with breast cancer susceptibility.

Acknowledgements

The authors are indebted to the participants and their families for their generosity and providing DNA samples. We thank Dr Damian Labuda from Centre de cancerologie Charles Bruneau, Ste-Justine Hospital, for control DNA samples. We would like to thank Martine Tranchant for his skillful technical assistance as well as Tina Babineau, Nathalie Bolduc, Claire Brousseau, Marie-Andrée Lajoie, Pascale Léger, Hélène Malouin, Andrée McMillan, and Josée Rhéaume for their genetic counseling and clinical data management at the Cancer Genomics Laboratory. We also thank Geneviève Ouellette for establishment of EBV-transformed B-lymphoblastoid cell lines and RNA and genomic DNA extractions. Expression vector containing the cDNA fragment encoding the human 17β-HSD type 2 and the antibody against the 17β-HSD type 2 protein were kindly provided by Dr Van Luu-The from the CHUL Research Center. We thank Claudia Moreau at the Centre de Recherche de l’Hôpital Ste-Justine for help with control DNA samples.

Disclosure

This work was supported by the Canadian Institutes of Health Research (CIHR) through the INHERIT BRCA1s research program, the Fonds de la Recherche en Santé du Québec (FRSQ)/Réseau de Médecine Génétique Appliquée (RMGA), the Canadian Breast Cancer Research Alliance and the CURE Foundation. J S is a Chairholder of the Canada Research Chair in Oncogenetics. F D is a recipient of a chercheur-boursier from the Fonds de la Recherche en Santé québécoise (RMGA), the Canadian Breast Cancer Research Program, the Fonds de la Recherche en Santé du Québec (CRIUQ), and a recipient of a chercheur-boursier from the Fonds de la Recherche en Santé du Québec (CRIUQ) and a recipient of a chercheur-boursier from the Fonds de la Recherche en Santé du Québec (CRIUQ).

References

Alex Dong Li’s SpliceSiteFinder. http://www.genet.sickkids.on.ca/ali/splicesitefinder.html.


Antoniou AC, Pharoah PD, McMullan G, Day NE, Ponder BA & Easton D 2001 Evidence for further breast cancer susceptibility genes in addition to BRCA1 and BRCA2 in a population-based study. *Genetic Epidemiology* 21 1–18.


Cargill M & Daly GQ 2000 Mining for SNPs: putting the common variants–common disease hypothesis to the test. *Pharmacogenomics* 1 27–37.


Received in final form 28 November 2007
Accepted 21 January 2008

Made available online as an Accepted Preprint 21 January 2008
Appendix

Other members of INHERIT BRCAs involved in clinical aspects of the program.

Paul Bessette: Department of Obstetrics and Gynecology, Centre Hospitalier Universitaire de Sherbrooke, Fleurimont, QC, Canada
Jocelyne Chiquette: Clinique des maladies du sein Deschênes-Fabia, Hôpital du Saint-Sacrement, Québec, QC, Canada
Rachel Laframboise: Medical Genetics Division, Centre Hospitalier Universitaire de Québec, CHUL, Laval University, Québec, QC, Canada
Jean Lépine: Haemato-Oncology Service, Centre Hospitalier Régional de Rimouski, Rimouski, QC, Canada
Bernard Lespérance, Roxanne Pichette: Department of Haemato-Oncology, Hôpital du Sacré-Coeur de Montréal, Montréal, QC, Canada
Marie Plante: Gynecology Oncology Division, Hôtel-Dieu de Québec, Centre Hospitalier Universitaire de Québec, Laval University, Québec, QC, Canada