Journal of Steroid Biochemistry & Molecular Biology 116 (2009) 134–153

Contents lists available at ScienceDirect

Journal of Steroid Biochemistry and Molecular Biology

journal homepage: www.elsevier.com/locate/jsbmb

Analysis of 17β-hydroxysteroid dehydrogenase types 5, 7, and 12 genetic sequence variants in breast cancer cases from French Canadian Families with high risk of breast and ovarian cancer

Marie Plourde[a,1], Alexandra Ferland[a,1], Penny Soucy[a], Yosr Hamdi[a], Martine Tranchant[a], Francine Durocher[a], Olga Sinilnikova[b], Van Luu The[c], Jacques Simard[a,d,*], INHERIT BRCAs[2]

[a] Cancer Genomics Laboratory, Endocrinology and Genomics Division/CHUQ Research Center and Laval University, Quebec, Canada
[b] Unité Mixte de Génétique Constitutionnelle des Cancers Fréquents, Hospices Civil de Lyon/Centre Léon Bérard, Lyon, France
[c] Endocrinology and Genomics Division/CHUQ Research Center and Laval University, Quebec, Canada
[d] Canada Research Chair in Oncogenetics, Canada

ARTICLE INFO

Article history:
Received 27 December 2008
Received in revised form 6 May 2009
Accepted 8 May 2009

Keywords:
Breast cancer
Genetic susceptibility
17β-Hydroxysteroid dehydrogenase
Polymorphisms
Estrogen metabolism

ABSTRACT

A family history and estrogen exposure are well-known risk factors for breast cancer. Members of the 17β-hydroxysteroid dehydrogenase family are responsible for important steps in the metabolism of androgens and estrogens in peripheral tissues, including the mammary gland. The crucial biological function of 17β-HSDs renders these genes good candidates for being involved in breast cancer etiology. This study screened for mutations in *HSD17B7* and *HSD17B12* genes, which encode enzymes involved in estradiol biosynthesis and in *AKR1C3*, which codes for 17β-HSD type 5 enzyme involved in androgen and progesterone metabolism, to assess whether high penetrance allelic variants in these genes could be involved in breast cancer susceptibility. Mutation screening of 50 breast cancer cases from non-*BRCA1/2* high-risk French Canadian families failed to identify germline likely high-risk mutations in *HSD17B7*, *HSD17B12* and *AKR1C3* genes. However, 107 sequence variants were identified, including seven missense variants. Assessment of the impact of missense variants on enzymatic activity of the corresponding enzymes revealed no difference in catalytic properties between variants of 17β-HSD types 7 and 12 and wild-type enzymes, while variants p.Glu77Gly and p.Lys183Arg in 17β-HSD type 5 showed a slightly decreased activity. Finally, a haplotype-based approach was used to determine tagging SNPs providing valuable information for studies investigating associations of common variants in these genes with breast cancer risk.

1. Introduction

Breast cancer is the most common malignant form of cancer among occidental women. Although the major susceptibility genes *BRCA1* and *BRCA2* have been identified as high-penetrance alleles, these alleles only account for approximately 15–20% of the familial component of breast cancer [1–3]. A recent study by our group has demonstrated that among French Canadians, about two-thirds of the high-risk families tested for the presence of mutations in the *BRCA1* and *BRCA2* genes will yield an inconclusive result [4], suggesting the existence of either other high-penetrance genes or multiple alleles of low to moderate penetrance [5].

Sex steroid hormone signaling regulates the development and functioning of the normal mammary gland and plays a role in the etiology of breast cancer. The contribution of estrogens in the regulation of cellular growth, differentiation and proliferation of the mammary gland as well as in hormone-sensitive breast carcinomas is now well-documented [6,7]. Recent evidence also suggests that progestin metabolites may play important roles in regulating the development of breast cancer [8]. Epidemiological studies of hormone replacement therapy in post-menopausal women demonstrated an increased risk of breast cancer under combined estrogen/progestin therapy. However, evidence tends to show that the nature of the progestin component in combined hormone therapy is of importance regarding breast cancer risk [8,9]. Increasing evidence also indicates that androgens exert inhibitory effects on the proliferation of the breast epithelial cells and play a protective role in the pathogenesis of breast cancer (for review see [7]). The intracellular concentration of active sex steroid hormones in the

* Corresponding author at: Cancer Genomics Laboratory, CHUQ Research Center, 2705 Laurier Boulevard, T3-57, Quebec City, Quebec, Canada G1 V 4G2. Tel.: +1 418 654 2264; fax: +1 418 654 2278.
E-mail address: Jacques.simard@crchul.ulaval.ca (J. Simard).
[1] Joint first authors.
[2] Other members of INHERIT BRCAs involved in clinical aspects of the program are listed in Appendix A.

0960-0760/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jsbmb.2009.05.005

breast is regulated by several enzymes, and changes in the expression pattern of these enzymes may play a pathophysiological role in malignant transformation, by significantly altering the intracellular steroid content. Indeed, several recent studies have reported altered expression levels of several steroid hormone metabolizing enzymes in human breast carcinoma [10,11] and therefore the analysis of genetic variations potentially responsible of the modulation of either expression levels or activity of these enzymes in the mammary gland, is an interesting venue to possibly explain an increased susceptibility to breast cancer.

Human 17β-hydroxysteroid dehydrogenase (17β-HSD) type 5 belongs to the aldo-keto reductase (AKR) superfamily [12] and is widely expressed in human tissues including the prostate, endometrium and mammary gland [13]. It participates in the biosynthesis and metabolism of a variety of substrates including androgens [14], estrogens [13] and progestins [14]. The enzyme possesses a strong 17-keto reductase activity towards androgens, converting the C19 steroid precursor Δ4-androstenedione to testosterone, and also shows a weak 17-keto reductase activity towards estrogens, converting estrone to the more potent estradiol [14,15]. The enzyme also possesses a strong 20α-HSD activity converting progesterone to 20α-OH-progesterone. Studies by Wiebe et al. [16] have demonstrated that 4-pregnene and 5α-pregnane metabolites, such as 20α-OH-progesterone, formed in non-tumor and tumor breast tissues have opposite effects on breast cell proliferation and adhesion. These results, coupled with altered *AKR1C3* (*HSD17B5*) expression levels in breast tumors as compared to normal tissue [10,11], renders this gene an attractive candidate that might explain a fraction of the inherited susceptibility to breast cancer.

17β-hydroxysteroid dehydrogenase type 7 enzyme (17β-HSD type 7) converts estrone into estradiol, while inactivating dihydrotestosterone through its additional 3-keto-reductase activity [17]. A recent study by Seth et al. has shown that overexpression of *HSD17B7* reverts cholesterol auxotrophy of NSO cells [18]. Indeed, through its 3-keto-reductase activity, it participates in the postsqualene cholesterol biosynthesis pathway by converting zymosterone to zymosterol, therefore providing evidence pointing to a dual functionality of *HSD17B7*, being involved in both steroidogenesis and cholesterol biosynthesis [19]. 17β-HSD type 7 is expressed in a variety of human tissues, including normal breast tissue and malignant breast tumors [20,21].

HSD17B12 encodes 17β-hydroxysteroid dehydrogenase type 12 enzyme (17β-HSD type 12), which also catalyzes the last step in the formation of estradiol from estrone [22]. Furthermore, the enzyme possesses a ketoacyl-coenzyme A reductase activity and is involved in fatty acid metabolism, more specifically in the microsomal fatty acyl two-carbon elongation cascade [23]. 17β-HSD type 12 is highly expressed in organs related to lipid metabolism and hormone-sensitive tissues such as normal mammary gland and breast carcinoma [21,23,24]. A recent study by Sinilnikova et al. suggested a possible role of acetyl-CoA carboxylase α common sequence variants in susceptibility to breast cancer [25], further rendering *HSD17B12*, and other genes involved in fatty acid synthesis, good breast cancer candidate genes. Recently, Luu-The et al. found that 17β-HSD type 12 mRNA was expressed at higher levels than 17β-HSD types 1 and 7 in the mammary gland, a ratio which differs according to tissues and which also suggests an implication of 17β-HSD type 12 in estrogen synthesis in the mammary gland [22]. It was also demonstrated that the expression of 17β-HSD type 12 was significantly higher in breast carcinoma specimens than in normal tissue [21], which is consistent with its potential role in the development and/or progression of breast cancer. It is also noteworthy that a recent genome-wide linkage analysis as well as a genome-wide association study for expression levels indicate that *HSD17B12* is among the genes for which there is strongest evidence of linkage between expression levels and *cis*-regulatory elements [26,27]. These conclusions were based on marked associations between expression level and certain haplotypes, directly demonstrating differential allelic expression [27].

To date, studies investigating the role of genetic variations in *HSD17B1* [28–34], and to a lesser extent in *HSD17B2* [35,36], have been conducted in relation to breast cancer risk with unconvincing results, but to our knowledge no studies have investigated germline mutations in *AKR1C3, HSD17B7* and *HSD17B12* genes with regard to breast cancer, in spite of the evidences mentioned above. The current study sought to identify germline potentially pathogenic variations in these genes using resequencing of the promoter region, the exonic and flanking intronic sequences in individuals affected with breast cancer from non-related *BRCA1/2*-negative high-risk French Canadian breast/ovarian cancer families. In addition, a haplotype-based analysis has also been done to characterize and to facilitate further study of common patterns of genetic variation in these genes.

2. Materials and methods

2.1. Ascertainment of high-risk families and DNA extraction

The recruitment of French Canadian families with high-risk of breast and ovarian cancer started in 1996 through a research project, which thereafter evolved in a large interdisciplinary research program called INHERIT BRCAs. More details regarding ascertainment criteria, experimental and clinical procedures have been described elsewhere [4,37]. A component was designed for the "localization and identification of new breast cancer susceptibility loci/genes". In order to increase the likelihood of potentially identifying novel genetic variants associated with breast cancer risk, individuals from French Canadian high-risk breast cancer families without *BRCA1/2* mutations were selected [38]. Ethics approval was obtained from the different institutions participating in this research component and each participant signed an informed consent for their participation in this latter project [39,40]. 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 families included at least 3 individuals with breast cancers among 2nd degree relatives, while five families included three or more individuals among 3rd 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. 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 46.1 years old (32–59 years), 9 of them had bilateral breast cancer. The *BRCA1/2* status of each participant was assessed as described elsewhere [4]. Genomic rearrangements in *BRCA1/2* genes were investigated by multiplex ligation-dependant probe amplification (MLPA) for 42 of the 50 subjects (BRCA1/2 Southern analyses were done for 32 of these 50 subjects). For four of the remaining subjects, MLPA was performed on another individual of the family [41], while for two subjects this analysis was not performed in their family.

Genomic DNA samples from 70 healthy unrelated French Canadian individuals were provided by Dr Damian Labuda at the Centre de Cancérologie Charles Bruneau, Hôpital Ste-Justine, Montreal, Canada. These control individuals 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 47.1 years old.

2.2. PCR amplification, mutation analysis and sequence variant identification

The sequence of coding exons, flanking intronic sequences, 5′ and 3′ non-coding regions and approximately 1–1.5 kb of the promoter region of *AKR1C3*, *17BHSD7*, *17BHSD12* genes was analyzed in DNA samples obtained from Epstein–Barr virus (EBV)-transformed B-lymphoblastoid cell lines of 50 affected individuals, as well as in genomic DNA samples from 70 healthy unrelated French Canadians. Primer pairs used to amplify the fragments are indicated in Supplemental Table 1. Direct sequencing was performed on an ABI Prism 3730xl 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. GenBank accession numbers for *AKR1C3*, *17BHSD7* and *17BHSD12* reference sequences used in these experiments were **NT_077567.3** and **NM_003739.4**, **NT_004487.47** and **NM_016371.2**, **NC_0000118.8** and **NM_016142.1**, respectively.

2.3. LD analysis, haplotype estimation and in silico analysis tools

The r^2 statistics of the Haploview program [42] were used to calculate the pairwise linkage disequilibrium (LD) for each sequence variant pair in the whole case-control set. Haplotype reconstructions and frequency estimations were performed using Phase 2.1.1 software [43]. This program estimates haplotype frequencies with a Bayesian-based algorithm. Haplotypes were estimated using SNPs with minor allele frequencies (MAF) ≥ 5% in control individuals. Genotype and marker data from control individuals were loaded into the Haploview software (http://www.broad.mit.edu/mpg/haploview) [42] for haplotype block identification. The default algorithm of block definition based on work of Gabriel et al. [44] was selected. Tagging SNPs (tSNPs) from each LD block, efficiently tagging all the known common variants (MAF ≥ 5%), were then identified using the same software.

Analysis of transcription factor binding sites in the promoter region was performed using MatInspector from Genomatix [45]. Splice site prediction scores were evaluated using splice site prediction by neural network (SSPNN; 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, [46]), while exonic splicing enhancers were analyzed using the ESEfinder program (available at http://rulai.cshl.edu//cgi-bin/tools/ESE3/esefinder.cgi, [47]). The effect of amino acid substitutions was evaluated using sorting intolerant from tolerant (SIFT; http://blocks.fhcrc.org/sift/SIFT.html, [48]) and PolyPhen (http://genetics.bwh.harvard.edu/pph/, [49]).

2.4. Site-directed mutagenesis

Site-directed mutagenesis was performed on an expression vector containing the full-length cDNA fragment encoding the complete amino acid sequence for human 17β-HSDs type 5, 7 and 12 (kindly provided by Dr Van Luu-The, CHUL Research Center, Quebec, Canada) inserted in the pcDNA3.1 vector (Invitrogen, Carlsbad, CA, USA) (17β-HSD type 5 and 7) or the pCMV vector (Invitrogen, Carlsbad, CA, USA) (17β-HSD type 12), 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 desired mutation was confirmed by direct sequencing of both strands using Big Dye Terminator chemistry on an ABI Prism 3730xl automated sequencer from Applied Biosystems (Applied Biosystems, Foster City, CA, USA). The primers used for insertion of the c.15C>G, c.230A>G, c.638C>T and c.648A>G mutations in 17β-HSD type 5, c.32G>C and c.961A>G in 17β-HSD type 7 and c.839C>T in 17β-HSD type 12 are indicated in Supplemental Table 2.

2.5. Cell culture and transfection

Human embryonic kidney 293 cells (HEK293) were cultured in Dulbeccos' modified Eagles medium (DMEM)/low glucose from Invitrogen Life Technologies, Inc. (Grand Island, NY) supplemented with 10% fetal bovine serum (FBS), 1% glutamine, 100 IU/ml penicillin and 50 mg/ml streptomycin from Wisent, Inc. (St-Bruno, Québec, Canada). Cells were then seeded in 6-well culture dishes and transient transfection was performed using ExGen 500 cationic polymer transfection reagent (MBI Fermentas Inc, Ontario, Canada) according to the supplier's protocol. Briefly, cells were transfected with 3 μg of constructs containing the wild-type and mutants and 1 μg of pSV-β-galactosidase DNA (Promega Corporation, Madison, WI, USA) as a control for transfection efficiency. The cells were also transfected with mock vectors as a negative control. All activities were normalized to β-galactosidase activity. β-galactosidase was measured using the β-galactosidase enzyme assay system with reporter lysis buffer (Promega Corporation, Madison, WI, USA).

2.6. Functional assays for non-synonymous variants

2.6.1. Transcription/translation

Transcription/translation for 17β-HSD type 5 and 7 wild-type and recombinant constructs was performed using the TNT® T7 Quick Coupled Transcription/Translation System from Promega (Promega Corporation, Madison, WI, USA) according to the manufacturer's instructions, as previously described [31].

2.6.2. Western analysis

Western analysis of proteins was performed by sodium dodecyl sulfate (SDS)–PAGE on discontinuous acrylamide gels. HEK293 cells were washed in phosphate buffered saline 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. Picastaway, NJ, USA). Thereafter, the membrane was hybridized to a polyclonal antibody directed against human 17β-HSD type 5 and 7 (kindly provided by Dr Van Luu-The) at dilutions of 1:2000 and 1:4000 respectively, and subsequently incubated with a donkey anti-rabbit IgG peroxidase conjugated secondary antibody (Amersham Pharmacia Biotech Inc. Picastaway, NJ) at a dilution of 1:10,000. Membranes were washed and proteins were visualized using the Western lightning™ chemiluminescence Reagent Plus (PerkinElmer), followed by exposure of the membranes to X-ray films for 1–10 min. The audiographic film was scanned and the ImageJ program (NIH, USA) was used to quantify the density of the autoradiographic bands.

2.6.3. Assay of 17β-HSD type 5, 7 and 12 enzymatic activity

For 17β-HSD type 5, transfected cells were incubated for indicated time periods with either 10 nM [4-^{14}C]-androst-4-ene-3,17-dione (56 mCi/mmol) (American Radiolabeled Chemicals inc., St-Louis, MO) (17β-HSD activity) or [4-^{14}C]-progesterone (55, 40 mCi/mmol) (NEN/ PerkinElmer Life Sciences, Inc., Boston, MA) (20α-HSD activity), while for 17β-HSD type 7 and 12, cells were incubated with 10 nM and 1 μM [4-^{14}C]-estrone (54 mCi/mmol) (American Radiolabeled Chemicals inc., St-Louis, MO) respectively, using well-established methods [14,22]. After the indicated time intervals, steroids were extracted by 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 using a mobile phase of toluene:acetone (4:1) and analyzed using phosphorimaging, Storm 860 (Molecular Dynamics, Inc., Sunnyvale, CA). All results

Table 1
Observed sequence variants and genotype frequencies in AKR1C3 gene among familial breast cancer cases and controls.

SNP	SNP ID[a]	dbSNP ID	Location	Series[b]	Genotype frequencies			MAF	χ^2 *P*-value[§§]	Reported MAF
					Common homozygotes (expected)[*]	Heterozygotes (expected)[*]	Rare homozygotes (expected)[*]			
1	g.5075410_5075413delATAA	N/A	Promoter	Cases	1.00 (1.00)	0.00 (0.00)	0.00 (0.00)	0.000		N/A
				Controls	0.99 (0.99)	0.01 (0.01)	0.00 (0.00)	0.007	0.9520	N/A
2	g.5075582G > T	rs34747896	Promoter	Cases	0.98 (0.98)	0.02 (0.02)	0.00 (0.00)	0.010	0.9431	N/A
				Controls	1.00 (1.00)	0.00 (0.00)	0.00 (0.00)	0.000		N/A
3	g.5076041C > A	rs11818810	Promoter	Cases	0.98 (0.98)	0.02 (0.02)	0.00 (0.00)	0.010	0.9431	N/A
				Controls	0.99 (0.99)	0.01 (0.01)	0.00 (0.00)	0.007	0.9520	0 [76]
4	g.5076148A > G	rs1937845	Promoter	Cases	0.52 (0.49)	0.36 (0.42)	0.12 (0.09)	0.300	0.3124	N/A
				Controls	0.39 (0.36)	0.43 (0.48)	0.19 (0.16)	0.400	0.3700	0.383 [76], 0.368 [77]
5	g.5076203C > T	rs10904415	Promoter	Cases	0.32 (0.35)	0.54 (0.48)	0.14 (0.17)	0.410	0.4114	N/A
				Controls	0.44 (0.45)	0.46 (0.44)	0.10 (0.11)	0.329	0.7628	0.383 [76], 0.392 [77]
6	g.5076293T > G	rs36201177	Promoter	Cases	0.94 (0.94)	0.06 (0.06)	0.00 (0.00)	0.030	0.8269	N/A
				Controls	0.97 (0.97)	0.03 (0.03)	0.00 (0.00)	0.014	0.9035	0 [77]
7	g.5076360C > A	rs2518047	Promoter	Cases	0.68 (0.67)	0.28 (0.30)	0.04 (0.03)	0.180	0.7158	N/A
				Controls	0.69 (0.71)	0.31 (0.26)	0.00 (0.02)	0.157	0.1188	0.143 [76], 0.145 [77]
8	g.5076499A > G	rs3763676	Promoter	Cases	0.32 (0.35)	0.54 (0.48)	0.14 (0.17)	0.410	0.4114	N/A
				Controls	0.44 (0.45)	0.46 (0.44)	0.10 (0.11)	0.329	0.7628	0.386 [76], 0.395 [77]
9	g.5076651C > G c.15C > G p.His5Gln	rs12529	Exon 1	Cases	0.52 (0.49)	0.36 (0.42)	0.12 (0.09)	0.300	0.3124	N/A
				Controls	0.39 (0.36)	0.43 (0.48)	0.19 (0.16)	0.400	0.3700	0.400 [76], 0.368 [77]
10	g.5076915A > G c.84 + 195A > G	rs1937843	Intron 1	Cases	0.32 (0.35)	0.54 (0.48)	0.14 (0.17)	0.410	0.4114	N/A
				Controls	0.43 (0.44)	0.47 (0.45)	0.10 (0.11)	0.336	0.6336	0.383 [76], 0.395 [77]
11	g.5078607G > A c.90G > A	rs7741	Exon 2	Cases	0.32 (0.35)	0.54 (0.48)	0.14 (0.17)	0.410	0.4114	N/A
				Controls	0.44 (0.45)	0.46 (0.44)	0.10 (0.11)	0.329	0.7628	0.390 [76], 0.395 [77]
12	g.5078747A > G c.230A > G p.Glu77Gly	rs41306308	Exon 2	Cases	0.82 (0.81)	0.16 (0.18)	0.02 (0.01)	0.100	0.4321	N/A
				Controls	0.81 (0.80)	0.16 (0.19)	0.03 (0.01)	0.107	0.1350	0.048 [67]
13	g.5079003T > C c.252 + 234T > C	rs2801883	Intron 2	Cases	0.52 (0.49)	0.36 (0.42)	0.12 (0.09)	0.300	0.3124	0.393 [78]
				Controls	0.39 (0.36)	0.43 (0.48)	0.19 (0.16)	0.400	0.3700	0.423 [76], 0.592 [77], 0.395 [78]
14	g.5079685A > G c.312A > G	rs12387	Exon 3	Cases	0.66 (0.67)	0.32 (0.30)	0.02 (0.03)	0.180	0.5525	N/A
				Controls	0.69 (0.70)	0.30 (0.27)	0.01 (0.03)	0.164	0.4388	0.167 [76], 0.132 [77]
15	g.5079815C > A c.369 + 73C > A	rs2245191	Intron 3	Cases	0.60 (0.59)	0.34 (0.35)	0.06 (0.05)	0.230	0.7768	N/A
				Controls	0.59 (0.58)	0.36 (0.36)	0.06 (0.06)	0.236	0.9414	0.242 [76], 0.264 [77]
16	g.5079886C > A c.369 + 144C > A	rs2298305	Intron 3	Cases	0.74 (0.74)	0.24 (0.24)	0.02 (0.02)	0.140	0.9813	N/A
				Controls	0.81 (0.82)	0.19 (0.17)	0.00 (0.01)	0.093	0.3918	0.100 [76], 0.059 [77]
17	g.5080980T > C c.370 − 14T > C	rs34140485	Intron 3	Cases	0.82 (0.81)	0.16 (0.18)	0.02 (0.01)	0.100	0.4321	N/A
				Controls	0.81 (0.80)	0.16 (0.19)	0.03 (0.01)	0.107	0.1350	0.106 [77]
18	g.5081137A > G c.447 + 66A > G	rs10508293	Intron 4	Cases	0.86 (0.86)	0.14 (0.13)	0.00 (0.00)	0.070	0.5946	N/A
				Controls	0.71 (0.70)	0.24 (0.27)	0.04 (0.03)	0.164	0.3336	0.142 [76], 0.100 [77]
19	g.5081262G > T c.447 + 191G > T	N/A	Intron 4	Cases	0.82 (0.81)	0.16 (0.18)	0.02 (0.01)	0.100	0.4321	N/A
				Controls	0.81 (0.80)	0.16 (0.19)	0.03 (0.01)	0.107	0.1350	N/A

Table 1 (*Continued*)

SNP	SNP ID[a]	dbSNP ID	Location	Series[b]	Genotype frequencies			MAF	χ^2 *P*-value[§§]	Reported MAF
					Common homozygotes (expected)[*]	Heterozygotes (expected)[*]	Rare homozygotes (expected)[*]			
20	g.5081289G > A c.448 – 230G > A	rs1937841	Intron 4	Cases Controls	0.74 (0.76) 0.81 (0.82)	0.26 (0.23) 0.19 (0.17)	0.00 (0.02) 0.00 (0.01)	0.130 0.093	0.2907 0.3918	0.107 [78] 0.100 [76], 0.109 [78]
21	g.5081307C > G c.448 – 212C > G	rs1937840	Intron 4	Cases Controls	0.52 (0.49) 0.39 (0.36)	0.36 (0.42) 0.43 (0.48)	0.12 (0.09) 0.19 (0.16)	0.300 0.400	0.3124 0.3700	N/A 0.383 [76], 0.361 [77]
22	g.5081530G > A c.459G > A	rs1937839	Exon 5	Cases Controls	0.98 (0.98) 0.99 (0.99)	0.02 (0.02) 0.01 (0.01)	0.00 (0.00) 0.00 (0.00)	0.010 0.007	0.9431 0.9520	N/A 0 [76]
23	g.5081609C > T c.638C > T p.Pro180Ser	rs34186955	Exon 5	Cases Controls	1.00 (1.00) 0.96 (0.96)	0.00 (0.00) 0.04 (0.04)	0.00 (0.00) 0.00 (0.00)	0.000 0.021	0.8546	0.081 [77]
24	g.5081619A > G c.648A > G p.Lys183Arg	N/A	Exon 5	Cases Controls	1.00 (1.00) 0.99 (0.99)	0.00 (0.00) 0.01 (0.01)	0.00 (0.00) 0.00 (0.00)	0.000 0.007	0.9520	N/A N/A
25	g.5083935A > C c.571 – 358A > C	rs2154307	Intron 5	Cases Controls	0.60 (0.59) 0.57 (0.54)	0.34 (0.35) 0.33 (0.39)	0.06 (0.05) 0.10 (0.07)	0.230 0.264	0.7768 0.1945	N/A 0.250 [76]
26	g.5084037T > G c.571 – 256T > G	rs4881400	Intron 5	Cases Controls	0.68 (0.69) 0.61 (0.58)	0.30 (0.28) 0.30 (0.36)	0.02 (0.03) 0.09 (0.06)	0.170 0.236	0.6556 0.1614	0.223 [78] 0.250 [76], 0.227 [78]
27	g.5084063C > G c.571 – 230C > G	rs12242350	Intron 5	Cases Controls	0.60 (0.59) 0.57 (0.54)	0.34 (0.35) 0.33 (0.39)	0.06 (0.05) 0.10 (0.07)	0.230 0.264	0.7768 0.1945	N/A 0.263 [76]
28	g.5084073A > T c.571 – 220A > T	rs11252940	Intron 5	Cases Controls	0.60 (0.59) 0.57 (0.54)	0.34 (0.35) 0.33 (0.39)	0.06 (0.05) 0.10 (0.07)	0.230 0.264	0.7768 0.1945	N/A 0.250 [76]
29	g.5084238T > A c.571 – 55T > A	N/A	Intron 5	Cases Controls	0.98 (0.98) 1.00 (1.00)	0.02 (0.02) 0.00 (0.00)	0.00 (0.00) 0.00 (0.00)	0.010 0.000	0.9431	N/A N/A
30	g.5084562insG c.681 – 115insG	N/A	Intron 6	Cases Controls	0.32 (0.35) 0.44 (0.45)	0.54 (0.48) 0.46 (0.44)	0.14 (0.17) 0.10 (0.11)	0.410 0.329	0.4114 0.7628	N/A N/A
31	g.5084599C > T c.681 – 78C > T	rs10904419	Intron 6	Cases Controls	0.68 (0.67) 0.69 (0.71)	0.28 (0.30) 0.31 (0.26)	0.04 (0.03) 0.00 (0.02)	0.180 0.157	0.7158 0.1188	N/A 0.143 [77]
32	g.5084621C > T c.681 – 56C > T	rs33921818	Intron 6	Cases Controls	0.68 (0.69) 0.60 (0.57)	0.30 (0.28) 0.31 (0.37)	0.02 (0.03) 0.09 (0.06)	0.170 0.243	0.6556 0.2238	N/A 0.181 [77]
33	g.5084635T > C c.681 – 42T > C	rs4347280	Intron 6	Cases Controls	0.60 (0.59) 0.57 (0.54)	0.34 (0.35) 0.33 (0.39)	0.06 (0.05) 0.10 (0.07)	0.230 0.264	0.7768 0.1945	N/A 0.292 [77]
34	g.5084638C > G c.681 – 39C > G	rs33979906	Intron 6	Cases Controls	0.68 (0.69) 0.59 (0.56)	0.30 (0.28) 0.33 (0.38)	0.02 (0.03) 0.09 (0.06)	0.170 0.250	0.6556 0.3003	N/A 0.181 [77]
35	g.5084657C > G c.681 – 20C > G	rs11252941	Intron 6	Cases Controls	0.34 (0.36) 0.46 (0.46)	0.52 (0.48) 0.44 (0.44)	0.14 (0.16) 0.10 (0.10)	0.400 0.321	0.5557 0.8988	N/A 0.400 [77]
36	g.5084969C > T c.846 + 127C > T	rs12769666	Intron 7	Cases Controls	0.60 (0.59) 0.57 (0.54)	0.34 (0.35) 0.33 (0.39)	0.06 (0.05) 0.10 (0.07)	0.230 0.264	0.7768 0.1945	N/A 0.276 [77]
37	g.5087579A > G c.847 – 208A > G	rs17155732	Intron 7	Cases Controls	0.58 (0.57) 0.57 (0.54)	0.34 (0.35) 0.33 (0.39)	0.06 (0.05) 0.10 (0.07)	0.235 0.264	0.8108 0.1945	N/A N/A
38	g.5087610delT c.847 – 177delT	N/A	Intron 7	Cases Controls	0.98 (0.98) 1.00 (1.00)	0.02 (0.02) 0.00 (0.00)	0.00 (0.00) 0.00 (0.00)	0.010 0.000	0.9431	N/A N/A
39	g.5087670A > G c.847 – 117A > G	rs35768949	Intron 7	Cases Controls	0.98 (0.98) 1.00 (1.00)	0.02 (0.02) 0.00 (0.00)	0.00 (0.00) 0.00 (0.00)	0.010 0.000	0.9431	N/A 0.026 [77]

40	g.5087909A > G c.929 + 40A > G	rs2275928	Intron 8	Cases Controls	0.38 (0.41) 0.36 (0.35)	0.52 (0.46) 0.47 (0.48)	0.10 (0.13) 0.17 (0.17)	0.360 0.407	0.3636 0.8444	N/A 0.375 [76], 0.329 [77]
41	g.5087944T > C c.929 + 75T > C	N/A	Intron 8	Cases Controls	1.00 (1.00) 0.99 (0.99)	0.00 (0.00) 0.01 (0.01)	0.00 (0.00) 0.00 (0.00)	0.000 0.007	 0.9520	N/A N/A
42	g.5088012C > T c.929 + 143C > T	rs35983920	Intron 8	Cases Controls	0.90 (0.88) 0.94 (0.94)	0.08 (0.11) 0.06 (0.06)	0.02 (0.00) 0.00 (0.00)	0.060 0.029	0.0398 0.8056	N/A 0 [77]
43	g.5089303T > C c.930 − 350T > C	rs9329316	Intron 8	Cases Controls	0.60 (0.56) 0.57 (0.50)	0.30 (0.38) 0.27 (0.41)	0.10 (0.06) 0.16 (0.09)	0.250 0.293	0.1573 0.0039	N/A 0.250 [76]
44	g.5089330A > G c.930 − 323A > G	rs7917546	Intron 8	Cases Controls	0.84 (0.85) 0.77 (0.73)	0.16 (0.15) 0.17 (0.24)	0.00 (0.01) 0.06 (0.02)	0.080 0.143	0.5386 0.0121	N/A 0.138 [76]
45	g.5089385C > T c.930 − 268C > T	N/A	Intron 8	Cases Controls	0.90 (0.90) 0.97 (0.97)	0.10 (0.10) 0.03 (0.03)	0.00 (0.00) 0.00 (0.00)	0.050 0.014	0.7098 0.9035	N/A N/A
46	g.5089404G > A c.930 − 249G > A	N/A	Intron 8	Cases Controls	0.98 (0.96) 0.99 (0.97)	0.00 (0.04) 0.00 (0.03)	0.02 (0.00) 0.01 (0.00)	0.020 0.014	1.54×10^{-12} 5.93×10^{-17}	N/A N/A
47	g.5089413T > C c.930 − 240T > C	rs11252951	Intron 8	Cases Controls	0.60 (0.56) 0.57 (0.50)	0.30 (0.38) 0.27 (0.41)	0.10 (0.06) 0.16 (0.09)	0.250 0.293	0.1573 0.0039	N/A 0.263 [76]
48	g.5089427T > C c.930 − 226T > C	rs10160019	Intron 8	Cases Controls	0.60 (0.56) 0.57 (0.50)	0.30 (0.38) 0.27 (0.41)	0.10 (0.06) 0.16 (0.09)	0.250 0.293	0.1573 0.0039	N/A 0.250 [76]
49	g.5089439_5089444delCGTGTT c.930 − 215delcgtgtt	N/A	Intron 8	Cases Controls	1.00 (1.00) 0.99 (0.99)	0.00 (0.00) 0.01 (0.01)	0.00 (0.00) 0.00 (0.00)	0.000 0.007	 0.9520	N/A N/A
50	g.5089539A > G c.930 − 114A > G	rs17134355	Intron 8	Cases Controls	0.98 (0.98) 0.99 (0.99)	0.02 (0.02) 0.01 (0.01)	0.00 (0.00) 0.00 (0.00)	0.010 0.007	0.9431 0.9250	N/A 0 [76]
51	g.5089649T > G c.930 − 4T > G	rs34320249	Intron 8	Cases Controls	0.98 (0.96) 0.99 (0.97)	0.00 (0.04) 0.00 (0.03)	0.02 (0.00) 0.01 (0.00)	0.020 0.014	1.54×10^{-12} 5.93×10^{-17}	N/A 0.111 [77]
52	g.5089703A > G c.980A > G	rs3209896	Exon 9	Cases Controls	0.52 (0.50) 0.50 (0.46)	0.38 (0.41) 0.36 (0.44)	0.10 (0.08) 0.14 (0.10)	0.290 0.321	0.5850 0.1293	N/A 0.381 [76], 0.338 [77]
53	g.5089732G > A c.1009G > A	rs28943581	Exon 9	Cases Controls	0.98 (0.98) 0.99 (0.97)	0.02 (0.02) 0.00 (0.03)	0.00 (0.00) 0.01 (0.00)	0.010 0.014	0.9431 5.93×10^{-17}	N/A N/A

[a] 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.

[b] Observed among 100 chromosomes from 50 breast cancer cases and 140 chromosomes from 70 control DNA samples.

[*] As expected under Hardy–Weinberg equilibrium.

§§ *P*-value for deviation from Hardy–Weinberg equilibrium (Pearson's chi-square). N/A: information not available for this SNP (not reported in dbSNP). MAF: minor allele frequency.

are expressed as the mean ± SE of at least two separate transfection experiments performed in triplicate. Enzymatic activities were compared by a 2-tailed unpaired Student's *t*-tests. Statistical significance was established at $P < 0.05$.

3. Results and discussion

3.1. Mutation analysis and variant characterization

3.1.1. AKR1C3/HSD17B5

Mutation analysis failed to reveal the presence of truncating mutations in the *AKR1C3/HSD17B5* coding region of our French Canadian breast cancer cases. However, 53 sequence variants were identified in the exonic, flanking intronic and promoter sequences (Table 1), which included eight promoter variants (arbitrarily defined as a 1500 bp genomic segment upstream exon 1), seven coding variants, four of which resulted in amino acid changes, two variations located in the 3′-UTR region, while the remaining 36 were intronic sequence variations. Among the 53 variants, 49 are single nucleotide substitutions, while the four remaining variations consist of three short deletions and one 1 bp insertion. Ten of these variants were novel while forty-three were reported in the single nucleotide polymorphism database (dbSNP Build 129) (www.ncbi.nlm.nih.gov/SNP). Thirty-five of the observed variants are common polymorphisms with minor allele frequencies (MAF) ≥5%, 12 of which are very frequent (MAF ∼ 30–40%), while 18 of the identified variants are considered rare variants since they display frequencies ≤5% (Table 1).

3.1.2. In silico analysis of promoter variants

Promoter analysis using the MatInspector program of Genomatix [45] revealed that, of these variants, variant #3 (g.5076041C > A) created a novel GATA binding site in the promoter region of the *AKR1C3* gene. Interestingly, four consensus GATA elements are found within the first 1.5 kilobases of the *AKR1C3* promoter, rendering this gene a potential target for GATA factor regulation. The creation of an additional consensus element by SNP #3 represents a potential mechanism that may be involved in altered *AKR1C3* expression levels in individuals harboring this variation. Indeed, the GATA transcription factors are emerging as important regulators of steroidogenesis [50] and several key enzymes involved in steroid hormone metabolism are target genes for GATA factors, including *HSD17B1* [51], *HSD3B1* [52], *HSD3B2* [53], *CYP17* [54,55] and *CYP19* [56]. It is interesting to note that, within the GATA protein family, GATA-3 was identified as an essential regulator of mammary gland morphogenesis and is also involved in the regulation of tumor differentiation and suppression of tumor dissemination in breast cancer [57–59]. Further analyses, which are beyond the scope of the current study, will be necessary to determine the contribution of GATA factors to *AKR1C3* transcriptional activity as well as the potential variation of promoter activity induced by the presence of an additional GATA consensus element. However with regard to breast cancer, taking into account that SNP #3 is a rare variant (MAF ≤ 1%) observed at a heterozygous state in one breast cancer case and one control individual, its relevance in the development of breast cancer appears rather unlikely. A previously reported promoter SNP (SNP #8, g.5076499A > G), shown to significantly increase the transcriptional activity of the *AKR1C3* promoter by increasing the binding affinity for Sp1/Sp3 transcription factors [60], was also observed in our cases as well as in our controls albeit at similar frequencies in both samples sets (MAF 0.41 for cases vs 0.33 for controls, p value = 0.196).

In addition to those mentioned above, three other potentially interesting transcription factor binding sites are created by polymorphisms #2 and #4 in the promoter region of the *AKR1C3* gene, as estimated by the Genomatix software. Variant #2 (g.5075582G > T) creates a new binding site for Fork head domain factors, which are essential in a wide range of cellular and developmental processes. Of most importance in the context of the current study is the implication of the forkhead gene in tumorigenesis and particulary in cell-cycle control [61]. This same variant also creates a binding site for SOX/SOY-sex/testis determining and related HMG box factors. Members of the Sox protein family are transcription factors, with some being transcriptional activators while others are repressors, depending on the requisite partners for target specificity [62]. A member of this family, SOX4, is expressed in normal breast tissue and in breast cancer cells and is transcriptionally regulated by progesterone [63]. Variant #2 was observed only once at a heterozygote state in our breast cancer cases. Lastly, variant #4 (g.5076148A > G) creates a new binding site for cellular and viral myb-like transcriptional regulators. C-myb is known to be a proto-oncogene and Myb-related protein B was shown to be deregulated in breast cancer [64]. The associated variant is very frequent with a MAF ≤ 30% for cases and 40% for controls.

3.1.3. Assessment of the impact of missense substitutions on 17β-HSD type 5 enzymatic activity

Among the exonic variants, four resulted in amino acid changes (p.His5Gln, p.Glu77Gly, p.Pro180Ser, p.Lys183Arg). For each variant, *in silico* analyses as well as functional assays were performed to determine the impact of variants on protein translation (TNT transcription/translation), protein expression (Western blot analysis) and enzymatic activity. The first non-synonymous coding SNP, p.His5Gln, is a frequent variant (MAF 0.3 in cases and 0.4 in controls) located in exon 1. It codes for a histidine to glutamine change, which are basic and acidic residues respectively, but whose structures are quite similar. Comparison of ortholog and several paralog sequences revealed that this residue is poorly conserved across species as well as in other members of the aldo-keto reductase family (data not shown). *In silico* analyses revealed that the change from a histidine to a glutamine is predicted to be a tolerated change and therefore probably have little impact on protein structure [49,65]. In support of these predictions, functional analyses in non-steroidogenic human embryonic kidney cells revealed that the mutant Gln5 enzyme possesses similar expression (Fig. 1, Panel B) and activity to the wild-type enzyme, and this for its 17-keto-activity (Fig. 1, Panel C) as well as its 20-keto-activity (Fig. 1, Panel D). A previous study found that the His5Gln polymorphism was associated with lung cancer, suggesting functional impact on the oxidation of polycyclic aromatic hydrocarbons to catechol through its 3-keto-activity, however due to the small sample size, additional studies are warranted to confirm this association [66], as well as the true effect of this polymorphism on the enzyme's activity.

The second missense substitution (p.Glu77Gly) has a MAF of 10% in cases and controls and involves a more drastic change from an acidic, polar, hydrophilic amino acid to a very small, aliphatic, non-polar, hydrophobic residue. Alignment of *AKR1C3* ortholog and several paralog sequences revealed that this amino acid is conserved in all species, including more distant species like *Xenopus tropicalis* and *Takifugu rubripes*, as well as in other members of the AKR family. Given that this residue is invariant from human to frog, this could suggest that this position is under strong functional constraint. In support of this, analyses using Polyphen and SIFT softwares predicted this change to be damaging to protein structure. Functional analyses show that the Gly77 variant protein is properly translated (Fig. 1, Panel A) but that its 17-keto activity (Fig. 1, Panel C) appears to slightly decrease compared to the wild-type enzyme activity ($P < 0.005$ at 24 hr). On the other hand, its 20-keto-activity on progesterone appears unaltered (Fig. 1, Panel D). Jakobsson et al. reported an association between the p.Glu77Gly polymorphism and lowered serum testosterone levels in men [67]. Functional analysis

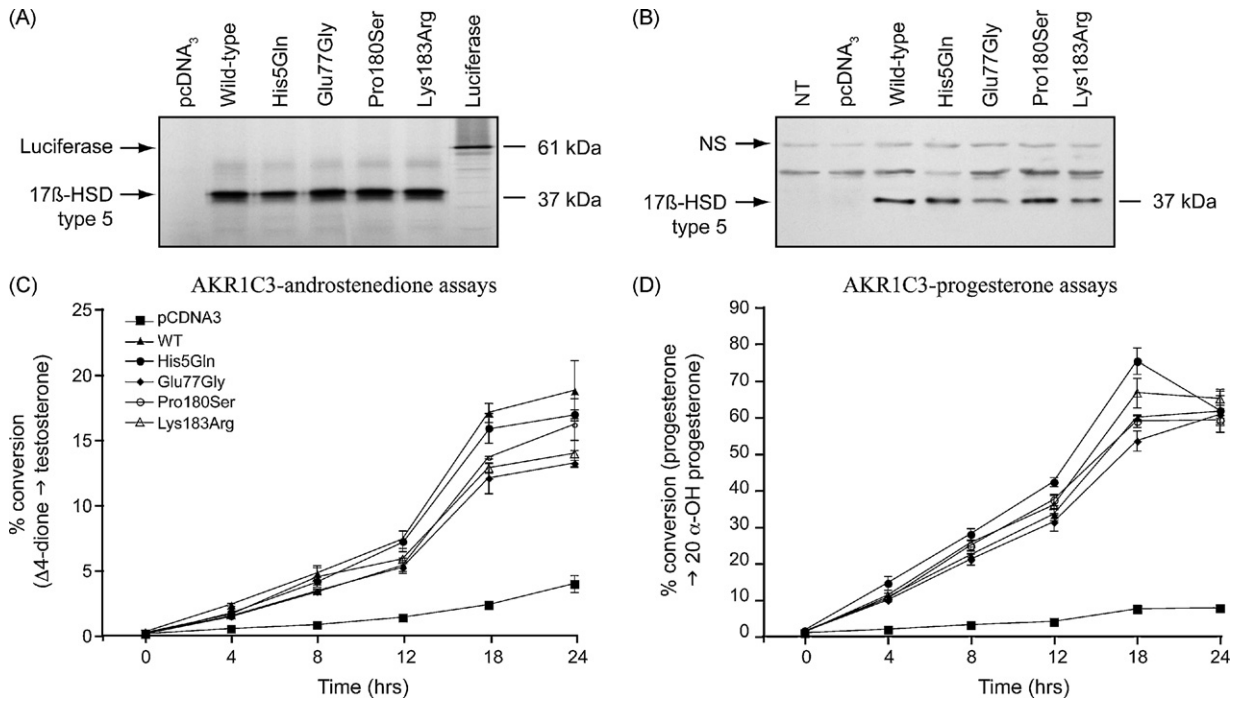


Fig. 1. Comparison of expression levels and activity of mutant recombinant 17β-HSD type 5 proteins. (A) Representation of an *in vitro* transcription/translation (TNT) rabbit reticulocyte lysate assay showing that each pcDNA3 construct is adequately translated into a [^{35}S]-labeled-37 kDa protein, indicative of normal expression levels of mutant recombinant 17β-HSD type 5 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 37 kDa band corresponding to 17β-HSD type 5 protein is detected in homogenate preparations from HEK293 transfected cells expressing wild-type and mutant recombinant proteins. The nonspecific band observed may be used as an internal control for loading. Enzymatic conversion of (C) [^{14}C]-androstenedione to [^{14}C]-testosterone and (D) progesterone to 20α-OH-progesterone in HEK293 cells transfected with the expression vectors after the indicated time periods. 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 5 mRNA expression.

performed in the study also revealed a tendency for decreased activity of the polymorphic Gly77 enzyme, although this difference was not significant. However, as stated by the authors, the crude bacterial cell lysates used in the study may not be sensitive enough to detect a modest decrease in activity. Furthermore, bacteria may not be an ideal system for investigating potential differences in activity resulting from post-translational events. Although a tendency toward reduced activity is observed in both studies, more refined enzymology experiments will be required to determine the precise impact of the variant on catalytic efficiency of this enzyme. Indeed, although the functional assays performed allow the detection of altered enzymatic activity, the exact nature of this alteration on the other hand, cannot be identified (e.g. if the decreased activity is caused by decreased affinity of the enzyme for the substrate, if it is due to improper folding of the protein etc.). Also, these assays are performed in presence of excess substrate which may not be the situation actually observed in the tissue and therefore with this in mind it cannot be assumed that our observations would be identical in an environment where the substrate is limiting. Finally, although the functional assays were performed using sex steroids, these enzymes are multifunctional enzymes with multiple activities and therefore it cannot be excluded that these variations do not have an effect on these activities and hence have an impact on other metabolic pathways.

The two remaining missense variants are rare polymorphisms only observed in control individuals. The first involves a change from a proline to a serine residue at position 180 located in exon 5. This variant involves a change from a cyclic, non-polar amino acid to a polar, hydrophilic residue. The second substitution involves two basic, polar, hydrophilic residues which are structurally fairly similar. Alignment of *AKR1C3* ortholog sequences revealed that both positions are highly conserved across species as well as in several members of the AKR superfamily (data not shown). However, *in silico* analyses predicted these variants to be benign or tolerated for protein folding. Furthermore, *in vitro* assessment of expression levels (Fig. 1, Panel B) and 20-keto enzymatic activities (Fig. 1, Panel C and D, respectively) of these two recombinant enzymes does not reveal a significant change as compared to wild-type. We do observe, however, a slightly significant decrease of the 17-keto-activity of Lys183Arg ($P < 0.05$ at 24 hr) but the actual impact of this variation needs to be further investigated.

Regarding the possible effect of p.His5Gln, p.Glu77Gly, p.Pro180Ser, p.Lys183Arg missense variants as well as that of the remaining exonic variants on exonic splicing enhancers (ESE), analysis of the scores revealed that none of these variants significantly altered the binding capacity of putative ESE elements (data not shown).

3.1.4. *In silico analysis of the effect of intronic variants on splicing*

The possible effect of all intronic variants on splicing consensus sequences was also assessed using *in silico* analysis with SSPNN program. Four of the genetic variations, namely g.5080980T > C (c.370 – 14T > C), g.5081137A > G (c.447 + 66A > G), g.5081307C > G (c.448 – 212C > G) and g.5081530G > A (c.459G > A), significantly altered the splicing score of either a physiological acceptor or donor site thereby potentially affecting pre-mRNA splicing. Variant c.370 – 14T > C abolishes intron 3 donor site, which could potentially cause skipping of exon 4. Variants c.447 + 66A > G and c.448 – 212C > G create new donor and acceptor sites predicted to result in the addition of 61 nucleotides to exon 4 and 208 nucleotides to exon 5, respectively, while variant c.459G > A increases the score of a pseudo-donor site in exon 5, potentially resulting in the exclusion of the last 109 nucleotides of exon 5 from mRNA. To further assess the

Table 2
Observed sequence variants and genotype frequencies in *HSD17B7* gene among familial breast cancer cases and controls.

SNP	SNP ID[a]	dbSNP D	Location	Series[b]	Genotype frequency						MAF	χ^2 *P*-value[**]	Reported MAF
					Common homozygote (expected)[*]		Heterozygote (expected)[*]		Rare homozygote (expected)[*]				
1	g.13169262G > A – –	N/A	Promoter	Cases Controls	1.00 0.99	(1.00) (0.99)	0.00 0.01	(0.00) (0.01)	0.00 0.00	(0.00) (0.00)	0.000 0.007	– 0.95	N/A N/A
2	g.13169453G > T c.-78G > T –	N/A	5′UTR of exon 1	Cases Controls	0.96 0.99	(0.96) (0.99)	0.04 0.01	(0.04) (0.01)	0.00 0.00	(0.00) (0.00)	0.020 0.007	0.89 0.95	N/A N/A
3	g.13169488T > C c.-40T > C –	rs1704754	5′UTR of exon 1	Cases Controls	0.70 0.48	(0.71) (0.53)	0.28 0.49	(0.27) (0.40)	0.02 0.03	(0.03) (0.08)	0.16 0.275	0.77 0.05	N/A N/A
4	g.13169505C > T c.-26C > T –	rs12118590	5′UTR of exon 1	Cases Controls	0.88 0.93	(0.88) (0.93)	0.12 0.07	(0.11) (0.07)	0.00 0.00	(0.00) (0.00)	0.060 0.036	0.65 0.75	N/A N/A
5	g.13169562G > C c.32G > C p.Ser11Thr	N/A	Exon 1	Cases Controls	0.96 0.99	(0.96) (0.99)	0.04 0.01	(0.04) (0.01)	0.00 0.00	(0.00) (0.00)	0.020 0.007	0.89 0.95	N/A N/A
6	g.13175416T > C c.332 + 9T > C –	rs1780019	Intron 3	Cases Controls	0.72 0.48	(0.72) (0.52)	0.26 0.48	(0.26) (0.41)	0.02 0.04	(0.02) (0.08)	0.150 0.283	0.89 0.14	N/A 0.158 [76]
7	g.13175426A > C c.332 + 19A > C –	N/A	Intron 3	Cases Controls	0.98 0.97	(0.98) (0.97)	0.02 0.03	(0.02) (0.03)	0.00 0.00	(0.00) (0.00)	0.010 0.014	0.94 0.90	N/A N/A
8	g.13176681C > G c.447 + 35C > G –	N/A	Intron 4	Cases Controls	0.96 0.94	(0.96) (0.94)	0.04 0.06	(0.04) (0.06)	0.00 0.00	(0.00) (0.00)	0.020 0.029	0.89 0.80	N/A N/A
9	g.13177095T > C c.447 + 449T > C –	rs4095366	Intron 4	Cases Controls	0.44 0.38	(0.42) (0.43)	0.42 0.55	(0.46) (0.45)	0.14 0.07	(0.12) (0.12)	0.350 0.348	0.59 0.08	N/A N/A
10	g.13181925A > G c.642 − 236A > G –	rs10917597	Intron 5	Cases Controls	0.58 0.46	(0.58) (0.50)	0.36 0.49	(0.37) (0.41)	0.06 0.04	(0.06) (0.08)	0.240 0.290	0.93 0.10	N/A 0.233 [76]
11	g.13182454G > C c.747 + 189G > C –	rs10917598	Intron 6	Cases Controls	0.48 0.46	(0.45) (0.50)	0.38 0.49	(0.44) (0.41)	0.14 0.04	(0.11) (0.08)	0.330 0.290	0.32 0.10	N/A 0.300 [76]
12	g.13182924T > C c.748 − 73T > C –	rs2665498	Intron 6	Cases Controls	0.94 0.91	(0.94) (0.92)	0.06 0.09	(0.06) (0.08)	0.00 0.00	(0.00) (0.00)	0.030 0.043	0.83 0.71	N/A N/A
13	g.13183176A > G c.804 + 123A > G –	rs11590943	Intron 7	Cases Controls	0.44 0.39	(0.42) (0.44)	0.42 0.54	(0.46) (0.45)	0.14 0.07	(0.12) (0.12)	0.350 0.341	0.59 0.11	N/A N/A
14	g.13183772T > C c.805 − 352T > C –	N/A	Intron 7	Cases Controls	0.98 1.00	(0.98) (1.00)	0.02 0.00	(0.02) (0.00)	0.00 0.00	(0.00) (0.00)	0.010 0.000	0.94 –	N/A N/A

Table 2 (*Continued*)

SNP	SNP ID[a]	dbSNP D	Location	Series[b]	Genotype frequency				MAF	χ^2 *P*-value[**]	Reported MAF
					Common homozygote (expected)[*]		Heterozygote (expected)[*]		Rare homozygote (expected)[*]			
15	g.13183989C > T	N/A	Intron 7	Cases	0.96	(0.96)	0.04	(0.04)	0.00 (0.00)	0.020	0.88	N/A
	c.805 − 135C > T			Controls	0.99	(0.99)	0.01	(0.01)	0.00 (0.00)	0.007	0.95	N/A
	–											
16	g.13184207C > T	rs12563263	Exon 8	Cases	0.48	(0.45)	0.38	(0.44)	0.14 (0.11)	0.330	0.32	N/A
	c.888C > T			Controls	0.46	(0.49)	0.48	(0.42)	0.06 (0.09)	0.297	0.23	N/A
	–											
17	g.13191085A > G	rs2684875	Exon 9	Cases	0.94	(0.94)	0.06	(0.06)	0.00 (0.00)	0.030	0.83	N/A
	c.961A > G			Controls	0.91	(0.92)	0.09	(0.08)	0.00 (0.00)	0.043	0.71	N/A
	p.Lys321Glu											

[a] 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.
[b] Observed among 100 chromosomes from 50 breast cancer cases and 140 chromosomes from 70 women controls DNA samples.
[*] As expected under Hardy–Weinberg equilibrium.
[**] *P*-value for deviation from Hardy–Weinberg equilibrium (Pearson's chi-square). N/A: information not available for this SNP (not reported in dbSNP). MAF: minor allele frequency.

possible effect of these intronic variants on splicing at the mRNA level, PCR analyses were performed on cDNA obtained from lymphoblastoid cell lines of carriers of these SNPs. These analyses failed to reveal evidence of aberrant splicing in lymphoblastoid cell lines from neither heterozygous nor homozygous (when observed) carriers of all these variants (data not shown).

3.1.5. *HSD17B7*

Seventeen variants were found in *HSD17B7* (Table 2), 7 of which were novel while 10 were reported in the single nucleotide polymorphism (dbSNP) database Build 129 (www.ncbi.nlm.nih.gov/SNP). Seven of the observed variants were very common polymorphisms with minor allele frequencies (MAF) ≥15%, while four of the identified variants had MAF around 5% and six were observed only once or twice (Table 2). These variants included one nucleotide substitution in the promoter region (SNP #1; g.13169262G > A) (arbitrarily defined as a 1000 bp genomic segment upstream exon 1) and 3 in the 5′-UTR of exon 1, one of which was not reported (SNP #2; c.-78G > T). Promoter analysis using the MatInspector program of Genomatix [45] revealed that none of these variants were found to lie in a known transcription factor binding site (data not shown). Three variants were found in the coding exonic regions, including two amino acid substitutions (SNP #5; c.32G > C (p.Ser11Thr) and SNP #17; c.961A > G (p.Lys321Glu)). P.Ser11Thr was observed in 2 cases and one control at a heterozygous state while p.Lys321Glu was observed in 3 and 6 heterozygous cases and controls, respectively. Both amino acids do not appear to be critical residues in the sequence motifs (cofactor binding and catalytic center) of the short chain dehydrogenase/reductase family, of which 17BHSD type 7 is a member. More precisely, p.Ser11 is a non-conserved amino acid in the cofactor binding motif ($T^8GxxxGxG^{15}$) while p.Lys321 is located outside the functional sequence motifs [68]. Alignment of 17BHSD type 7 ortholog sequences revealed that p.Ser11 is conserved among mammals while an asparagine residue is observed in more distant species such as *Danio rerio*, *X. tropicalis* and *Drosophilia melanogaster*, and p.Lys321 is a highly conserved residue in most mammals with the exception of *Pan troglodytes* in which a tryptophan is observed (data not shown). *In silico* analyses revealed that for both variants, the change in amino acid is predicted to be a tolerated change and therefore probably have little impact on protein structure [49,65]. Functional assays were performed to assess whether these substitutions possibly had an effect on enzymatic activity. *In vitro* transcription/translation assays showed normal expression levels of both wild type and mutant recombinant p.Ser11Thr and p.Lys321Glu proteins (Fig. 2, panel A). A 37 kDa band corresponding to 17β-HSD type 7 protein was detected by Western blot analysis using cells expressing the wild type and both variants (Fig. 2, panel B). We also investigated the time course formation of [^{14}C]-estradiol from [^{14}C]-estrone in transfected HEK293 cells. As illustrated in Fig. 2, panel C, the activities of p.Ser11Thr and p.Lys321Glu enzymes were similar to that of the wild-type enzyme. Taken together, these results indicate that these amino acid substitutions do not have a significant impact on the activity of the 17β-HSD type 7 enzyme.

The influence of these base substitutions on putative ESE sites revealed that ESE scores were not significantly different between the wild-type and the variant sequence for c.961A > G (p.Lys321Glu). Regarding variant c.32G > C (p.Ser11Thr), the score for SF2/ASF (2.68) was decreased below the threshold value (1.96), although this G-to-C transition also produced a new positive score for another SR protein, SRp40 (3.35), as well as an increase of the score for the SC35 site (3.74 → 4.65). Furthermore, three of the four silent exonic *HSD17B7* sequence variants identified in this study were predicted to be localized in ESE sites. Indeed, c.-78 G > T

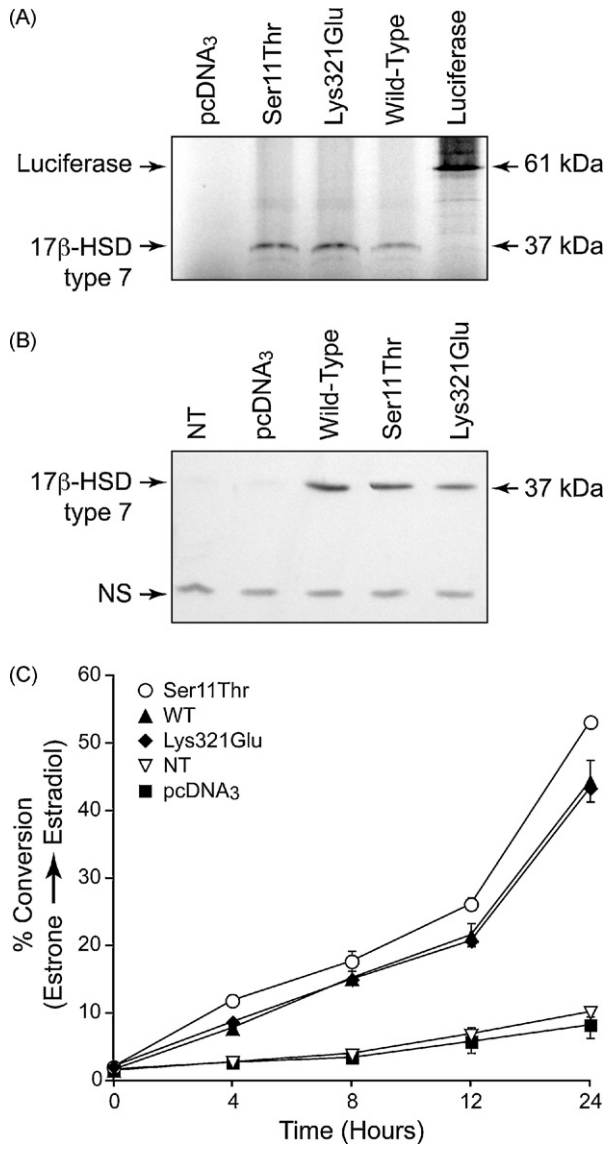


Fig. 2. Comparison of expression levels and activity of mutant recombinant 17β-HSD type 7 proteins. (A) Representation of an *in vitro* transcription/translation (TNT) rabbit reticulocyte lysate assay showing that each pcDNA3 construct is adequately translated into a [^{35}S]-labeled-37 kDa protein, indicative of normal expression levels of mutant recombinant 17β-HSD type 7 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 37 kDa band corresponding to 17β-HSD type 7 protein is detected in homogenate preparations from HEK293 transfected cells expressing wild-type and mutant recombinant proteins p.Ser11Thr and p.Lys321Glu. The non-specific band observed may be used as an internal control for loading. (C) Enzymatic conversion of [^{14}C]-estrone to [^{14}C]-estradiol in HEK293 cells transfected with the indicated expression vectors after a 24-hr incubation period. 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 7 mRNA expression.

resulted in the loss of one SRp55 site (3.18), c.-40T > C induced the gain of a SRp55 site (3.44) and also resulted in the prediction of two less efficient ESE sites (SF2/ASF 4.39 → 2.54 and SRp40 5.24 → 2.85). No significant difference was predicted for the c.888C > T variant.

Ten nucleotide substitutions were found in the intronic regions. The possible effect on splicing of all coding and intronic variants located in splice site junctions was assessed and predicted that none of the variants had a significant impact on either consensus donor and/or acceptor site scores (data not shown).

3.1.6. HSD17B12

A total of 37 variants were identified in the *HSD17B12* gene, twenty-two of which were novel (Table 3). Fourteen of the identified variants were very common polymorphisms with MAFs ≥ 18%. Four of the SNPs had MAFs around 5% and nineteen were observed only a few times (Table 3). Eight nucleotide substitutions and one deletion (SNP #3; g.43658300_43658301delAA) were identified in the promoter region, none of which seemed to significantly affect any known transcription factor binding sites or create a new one, as revealed by *in silico* analysis using the MatInspector program (data not shown) [45]. Eleven exonic variants were found, only one of which altered the amino acid sequence (SNP #29; p.Ser280Leu). Comparison of Ser280 across species revealed that this residue is conserved among mammals while a tryptophan is observed in other more distant species such as *D. rerio*, *Fugu rubripes*, *Tetraodon nigroviridis* and *X. tropicalis* (data not shown). Although p.Ser280Leu involves a change from an uncharged polar amino acid (serine) to a non-polar residue (leucine), this variation probably has a weak impact on protein structure since *in silico* analysis predicts this change to be benign [49,65]. Furthermore, this amino acid is located outside the conserved sequence motifs of the short chain dehydrogenase/reductase family [68]. In support of these analyses, assessment of the functional impact of this substitution revealed that the activity of the mutant Ser280Leu enzyme was similar to that of the wild-type enzyme, as measured by conversion of estrone to estradiol (data not shown).

The influence of this variant on putative ESE sites revealed that a reduced score for SF2/ASF (2.11) to a level below that of the threshold (1.96), suggesting a potential loss of the splicing-site recognition property. We further investigated the importance of the ten silent exonic variants and analyses predicted that variants c.615C > T and c.1349_1359delGTTTTT do not lie within putative ESEs. Four of these variants resulted in the loss of one or more ESE sites, three sequence variants induced gain of an ESE site, while three of them resulted in prediction of a less efficient ESE site, but still above the matrix score threshold level.

Finally, 15 nucleotide substitutions and 2 deletions were found in the intronic regions. The possible effect of all coding and intronic variants on splicing consensus sequences was assessed using *in silico* analysis. None of the observed genetic variants showed a significant change in the splicing score (data not shown).

For all three genes, genotypes and MAFs were determined in cases and controls, both of French Canadian origin. Screening in control individuals was performed to help assess the deleterious nature of the variants, as deleterious mutations are unlikely to be observed in this sample set. As mentioned above, it should be noted that a significant association with breast cancer susceptibility between variants found in cases compared to those found in controls would need much larger sample sets, especially for low penetrance genes [69,70]. As indicated in Tables 1–3, most genotype frequencies were consistent with those reported in the dbSNP databases for the CEPH cohort, with the exception of SNPs #15, 26, 27 and 29 in *HSD17B12* for which the frequencies observed in our sample sets of cases and controls are slightly but not significantly different from those reported in the NCBI database. Genotype frequency distributions were close to those expected under Hardy–Weinberg equilibrium except for a few rare SNPs (SNPs# 46, 51, 53 see Table 1) in *AKR1C3*. This was due to that for these rare variations, we observed the presence of one rare homozygote while no heterozygotes were detected.

To summarize, no likely deleterious or gain-of-function mutations were identified in either *AKR1C3*, *HSD17B7* or *HSD17B12*, and thus these genes are unlikely to play a role as high penetrance genes in breast cancer predisposition. An association with more modest risks cannot be excluded for the sequence variants identified in these genes through the current study. Careful choice of SNPs within these candidate genes is crucial to examine common gene

Table 3
Observed sequence variants and genotype frequencies in *HSD17B12* gene among familial breast cancer cases and controls.

SNP	SNP ID[a]	dbSNP D	Location	Series[b]	Genotype frequency						MAF	χ^2 *P*-value[**]	Reported MAF
					Common homozygote (expected)[*]		Heterozygote (expected)[*]		Rare homozygote (expected)[*]				
1	g.43658006G > T	N/A	Promoter	Cases	0.98	(0.98)	0.02	(0.02)	0.00	(0.00)	0.010	0.94	N/A
	–			Controls	1.00	(1.00)	0.00	(0.00)	0.00	(0.00)	0.000	–	N/A
	–												
2	g.43658275C > T	N/A	Promoter	Cases	0.94	(0.94)	0.06	(0.06)	0.00	(0.00)	0.030	0.83	N/A
	–			Controls	0.99	(0.99)	0.01	(0.01)	0.00	(0.00)	0.007	0.95	N/A
	–												
3	g.43658300_43658301delAA	rs35548313	Promoter	Cases	1.00	(1.00)	0.00	(0.00)	0.00	(0.00)	0.000	–	N/A
	–			Controls	0.91	(0.92)	0.09	(0.08)	0.00	(0.00)	0.043	0.71	N/A
	–												
4	g.43658360C > T	rs12224705	Promoter	Cases	0.60	(0.56)	0.30	(0.38)	0.10	(0.06)	0.250	0.16	N/A
	–			Controls	0.65	(0.66)	0.32	(0.31)	0.03	(0.04)	0.188	0.72	N/A
	–												
5	g.43658450G > C	N/A	Promoter	Cases	1.00	(1.00)	0.00	(0.00)	0.00	(0.00)	0.000	–	N/A
	–			Controls	0.99	(0.99)	0.01	(0.01)	0.00	(0.00)	0.007	0.95	N/A
	–												
6	g.43658479T > C	N/A	Promoter	Cases	0.84	(0.85)	0.16	(0.15)	0.00	(0.01)	0.080	0.54	N/A
	–			Controls	0.87	(0.87)	0.13	(0.12)	0.00	(0.00)	0.065	0.56	N/A
	–												
7	g.43658529C > T	N/A	Promoter	Cases	0.98	(0.98)	0.02	(0.02)	0.00	(0.00)	0.010	0.92	N/A
	–			Controls	0.99	(0.99)	0.01	(0.01)	0.00	(0.00)	0.007	0.93	N/A
	–												
8	g.43658581C > G	N/A	Promoter	Cases	0.98	(0.98)	0.02	(0.02)	0.00	(0.00)	0.010	0.94	N/A
	–			Controls	0.97	(0.97)	0.03	(0.03)	0.00	(0.00)	0.014	0.90	N/A
	–												
9	g.43658652G > T	N/A	Promoter	Cases	1.00	(1.00)	0.00	(0.00)	0.00	(0.00)	0.000	–	N/A
	–			Controls	0.99	(0.99)	0.01	(0.01)	0.00	(0.00)	0.007	0.95	N/A
	–												
10	g.43659106A > G	N/A	Exon 1	Cases	0.60	(0.56)	0.30	(0.38)	0.10	(0.06)	0.250	0.16	N/A
	c.153A > G			Controls	0.65	(0.66)	0.32	(0.31)	0.03	(0.04)	0.188	0.72	N/A
	–												
11	g.43659157A > G	rs4573669	Intron 1	Cases	0.60	(0.56)	0.30	(0.38)	0.10	(0.06)	0.250	0.16	N/A
	c.160 + 44A > G			Controls	0.58	(0.59)	0.38	(0.36)	0.04	(0.05)	0.232	0.63	0.317 [76]
	–												
12	g.43729125A > G	rs12801203	Intron 2	Cases	0.46	(0.41)	0.36	(0.46)	0.18	(0.13)	0.360	0.12	N/A
	c.207 + 42A > G			Controls	0.29	(0.30)	0.52	(0.50)	0.19	(0.20)	0.449	0.65	0.391 [79]
	–												
13	g.43729170G > A	N/A	Intron 2	Cases	0.98	(0.98)	0.02	(0.02)	0.00	(0.00)	0.010	0.94	N/A
	c.207 + 87G > A			Controls	1.00	(1.00)	0.00	(0.00)	0.00	(0.00)	0.000	–	N/A
	–												
14	g.43729189T > C	N/A	Intron 2	Cases	1.00	(1.00)	0.00	(0.00)	0.00	(0.00)	0.000	–	N/A
	c.207 + 106T > C			Controls	0.99	(0.99)	0.01	(0.01)	0.00	(0.00)	0.007	0.95	N/A
	–												
15	g.43732310G > A	rs4643069	Intron 3	Cases	0.46	(0.45)	0.42	(0.44)	0.12	(0.11)	0.330	0.72	N/A
	c.283 + 63G > A			Controls	0.45	(0.48)	0.49	(0.42)	0.06	(0.09)	0.304	0.17	0.250 [76]
	–												

Table 3 (*Continued*)

SNP	SNP ID[a]	dbSNP D	Location	Series[b]	Genotype frequency						MAF	χ^2 *P*-value[**]	Reported MAF
					Common homozygote (expected)[*]		Heterozygote (expected)[*]		Rare homozygote (expected)[*]				
16	g.43776332T > G c.284 – 114T > G –	N/A	Intron 3	Cases Controls	1.00 0.99	(1.00) (0.99)	0.00 0.01	(0.00) (0.01)	0.00 0.00	(0.00) (0.00)	0.000 0.007	– 0.95	N/A N/A
17	g.43776785C > A c.391 + 231C > A –	rs10838172	Intron 4	Cases Controls	0.60 0.61	(0.58) (0.58)	0.32 0.30	(0.37) (0.36)	0.08 0.09	(0.06) (0.06)	0.240 0.240	0.39 0.17	N/A 0.235 [76]
18	g.43793534A > G c.392 – 29A > G –	N/A	Intron 4	Cases Controls	1.00 0.99	(1.00) (0.99)	0.00 0.01	(0.00) (0.01)	0.00 0.00	(0.00) (0.00)	0.000 0.007	– 0.95	N/A N/A
19	g.43793711C > T c.456 + 84C > T –	N/A	Intron 5	Cases Controls	1.00 0.99	(1.00) (0.99)	0.00 0.01	(0.00) (0.01)	0.00 0.00	(0.00) (0.00)	0.000 0.007	– 0.93	N/A N/A
20	g.43793822T > C c.456 + 195T > C –	N/A	Intron 5	Cases Controls	0.98 0.97	(0.98) (0.97)	0.02 0.03	(0.02) (0.03)	0.00 0.00	(0.00) (0.00)	0.010 0.014	0.92 0.90	N/A N/A
21	g.43793921G > A c.456 + 294G > A –	N/A	Intron 5	Cases Controls	0.98 0.99	(0.98) (0.99)	0.02 0.01	(0.02) (0.01)	0.00 0.00	(0.00) (0.00)	0.010 0.007	0.94 0.95	N/A N/A
22	g.43794465delT c.457 – 8delT –	N/A	Intron 5	Cases Controls	0.98 0.99	(0.98) (0.99)	0.02 0.01	(0.02) (0.01)	0.00 0.00	(0.00) (0.00)	0.010 0.007	0.94 0.95	N/A N/A
23	g.43809067T > C c.502 – 35T > C –	rs6485471	Intron 6	Cases Controls	0.46 0.29	(0.41) (0.30)	0.36 0.52	(0.46) (0.50)	0.18 0.19	(0.13) (0.20)	0.360 0.449	0.12 0.65	N/A 0.425 [76]
24	g.43816521C > T c.615C > T –	N/A	Exon 8	Cases Controls	1.00 0.99	(1.00) (0.99)	0.00 0.01	(0.00) (0.01)	0.00 0.00	(0.00) (0.00)	0.000 0.007	– 0.93	N/A N/A
25	g.43816777delC c.618 + 253delC –	rs3839954	Intron 8	Cases Controls	0.46 0.45	(0.45) (0.48)	0.42 0.49	(0.44) (0.42)	0.12 0.06	(0.11) (0.09)	0.330 0.304	0.72 0.17	N/A N/A
26	g.43818324A > G c.684 + 134A > G –	rs10734518	Intron 9	Cases Controls	0.46 0.45	(0.45) (0.48)	0.42 0.49	(0.44) (0.42)	0.12 0.06	(0.11) (0.09)	0.330 0.304	0.72 0.17	N/A 0.263 [76]
27	g.43833011A > G c.834 + 21A > G –	rs3736505	Intron 10	Cases Controls	0.46 0.45	(0.45) (0.48)	0.42 0.49	(0.44) (0.42)	0.12 0.06	(0.11) (0.09)	0.330 0.304	0.72 0.17	N/A 0.258 [76]
28	g.43833116A > G c.834 + 126A > G –	rs11037681	Intron 10	Cases Controls	0.86 0.87	(0.87) (0.87)	0.14 0.13	(0.13) (0.12)	0.00 0.00	(0.01) (0.00)	0.070 0.065	0.59 0.56	N/A N/A
29	g.43833274C > T c.839C > T p.Ser280Leu	rs12576296	Exon 11	Cases Controls	0.46 0.45	(0.45) (0.48)	0.42 0.49	(0.44) (0.42)	0.12 0.06	(0.11) (0.09)	0.330 0.304	0.72 0.17	N/A 0.280 [76]

30	g.43833389A > C c.953A > C –	rs11037683	3′UTR of exon 11	Cases Controls	0.86 0.87	(0.87) (0.87)	0.14 0.13	(0.13) (0.12)	0.00 0.00	(0.01) (0.00)	0.070 0.065	0.59 0.56	N/A 0.052 [76]
31	g.43833405A > G c.969A > G –	rs11037684	3′UTR of exon 11	Cases Controls	0.86 0.87	(0.87) (0.87)	0.14 0.13	(0.13) (0.12)	0.00 0.00	(0.01) (0.00)	0.070 0.065	0.59 0.56	N/A 0.075 [76]
32	g.43833423G > A c.987G > A –	N/A	3′UTR of exon 11	Cases Controls	1.00 0.99	(1.00) (0.99)	0.00 0.01	(0.00) (0.01)	0.00 0.00	(0.00) (0.00)	0.000 0.007	– 0.95	N/A N/A
33	g.43833519A > G c.1083A > G –	N/A	3′UTR of exon 11	Cases Controls	0.98 1.00	(0.98) (1.00)	0.02 0.00	(0.02) (0.00)	0.00 0.00	(0.00) (0.00)	0.010 0.000	0.94 –	N/A N/A
34	g.43833785_43833791delGTTTTT c.1349_1354delGTTTTT –	N/A	3′UTR of exon 11	Cases Controls	0.46 0.45	(0.45) (0.48)	0.42 0.49	(0.44) (0.42)	0.12 0.06	(0.11) (0.09)	0.330 0.304	0.72 0.17	N/A N/A
35	g.43834024_43834025ins317 c.1588_1589ins317 Genbank: AK092938.1:c.1995_2311	N/A	3′UTR of exon 11	Cases Controls	0.52 0.46	(0.49) (0.50)	0.36 0.49	(0.42) (0.41)	0.12 0.04	(0.09) (0.08)	0.300 0.290	0.31 0.10	N/A N/A
36	g.43834510C > A c.2074C > A –	rs1061810	3′UTR of exon 11	Cases Controls	0.52 0.46	(0.49) (0.50)	0.36 0.49	(0.42) (0.41)	0.12 0.04	(0.09) (0.08)	0.300 0.290	0.31 0.10	N/A 0.242 [76]
37	g.43834511A > G c.2075A > G –	N/A	3′UTR of exon 11	Cases Controls	1.00 0.99	(1.00) (0.99)	0.00 0.01	(0.00) (0.01)	0.00 0.00	(0.00) (0.00)	0.000 0.007	– 0.95	N/A N/A

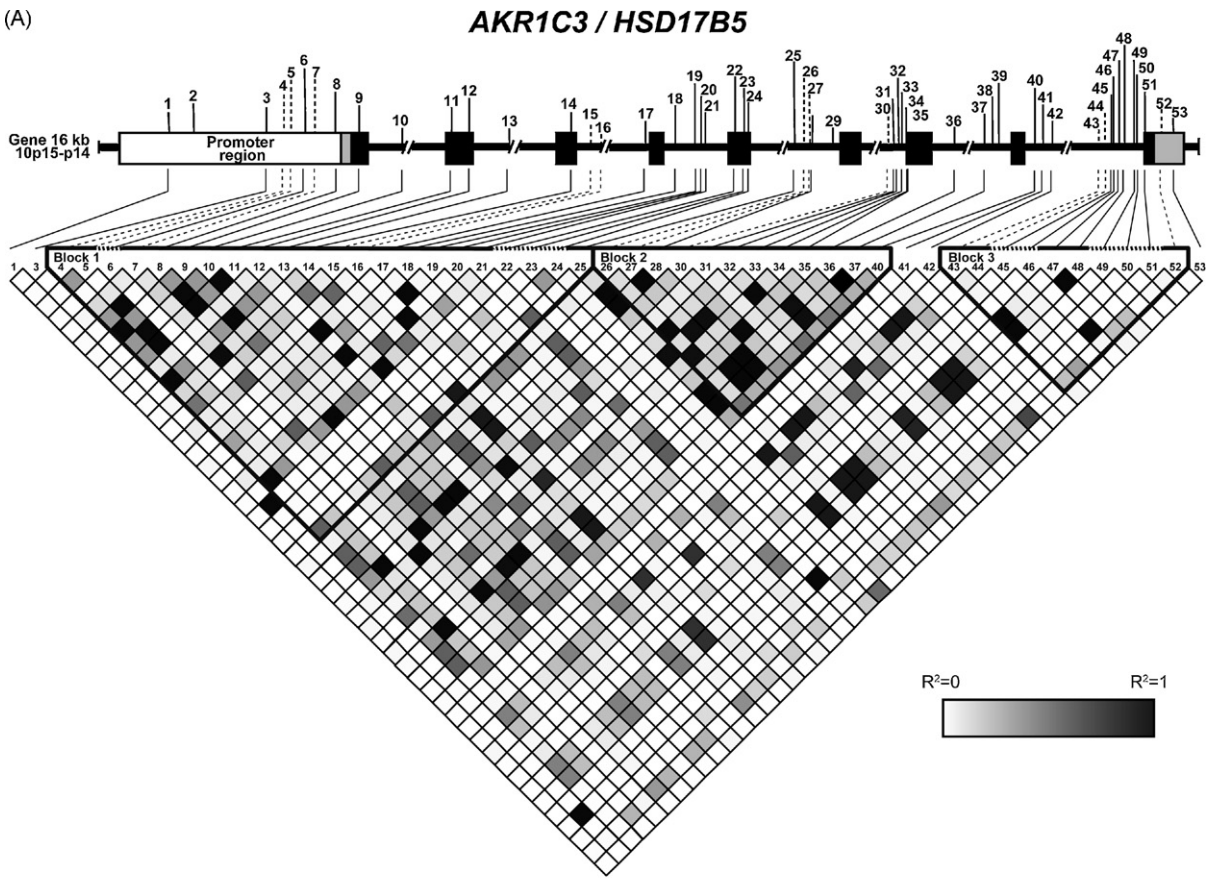
[a] According to the nomenclature of the Human Genome Variation Society, for cDNA numbering + 1 corresponds to the A of the ATG translation initiation codon in the reference sequence.

[b] Observed among 100 chromosomes from 50 breast cancer cases and 140 chromosomes from 70 women controls DNA samples.

[*] As expected under Hardy–Weinberg equilibrium.

[**] *P*-value for deviation from Hardy–Weinberg equilibrium (Pearson's chi-square). N/A: information not available for this SNP (not reported in dbSNP). MAF: minor allele frequency.

(A)



(B)

S4-5-7-8-9-10-11-12-13-14-15-16-17-18-19-20-21-25-26-27-28 30-31-32-33-34-35-36-37-40-43-44-47-48-52*	Frequency	Cumulative Frequency
ATCGCGAATACCTAGGCATCA+CCTCGCAATATTA	0.307	0.307
GCCAGAGACAACTAGGGCTGT-CCCCCTGACACCA	0.221	0.528
GCCAGAGACACCTGGGGAGCA-CTTGCCAGTGTTG	0.129	0.657
ACAACAGATGCATAGACATCA-TCTCCCAGTATTG	0.079	0.736
ACAACAGATGCCTAGGCATCA-TCTCCCAGTATTG	0.071	0.807
ACCACAGGTACCCATGCAGCA-CTTGCCAGTATTA	0.043	0.850

*SNP ID according to Table 1

(C)

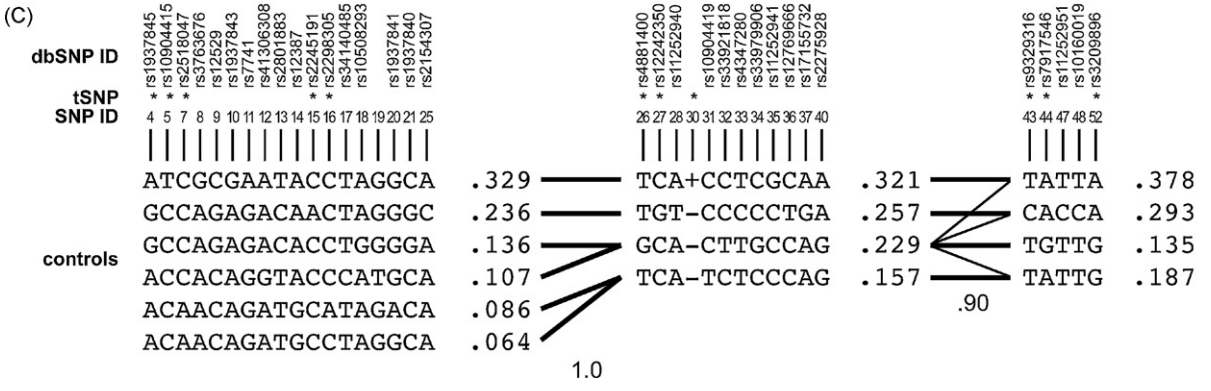
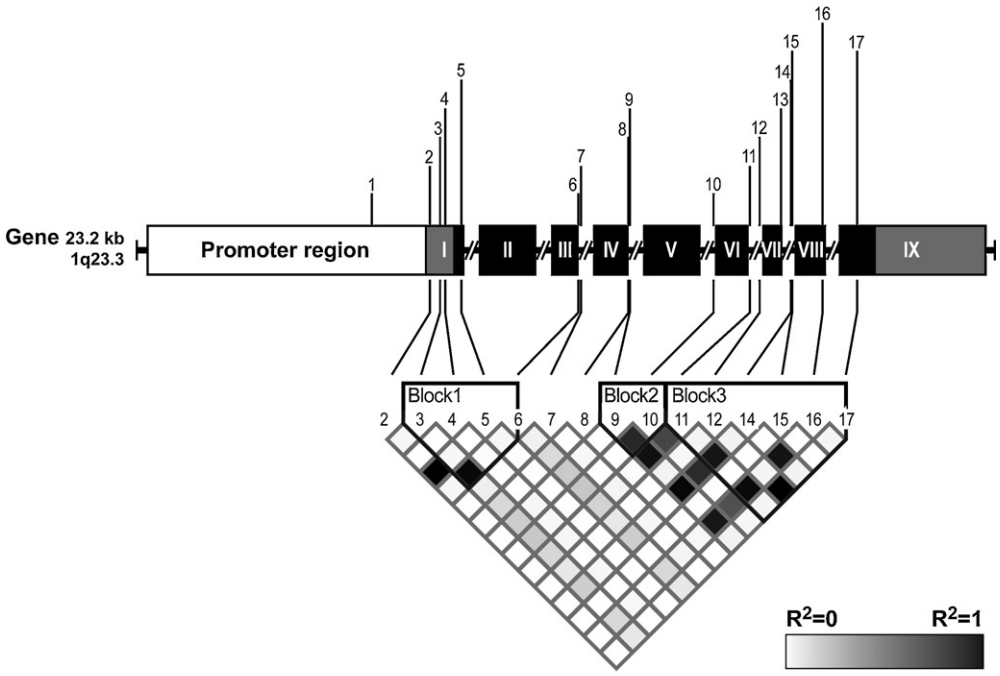


Fig. 3. (A) Pairwise linkage (LD) measures of r^2 for *AKR1C3* sequence variants identified in the control series. Sequence variants are represented by thin lines and are denoted numerically with reference to Table 1. (B) The table denotes haplotype frequencies estimated from *AKR1C3* sequence variants having a MAF ≥5% among the 70 controls. (C) Haplotype blocks predicted from variants identified in French Canadian controls showing a frequency ≥5%. tSNPs identified on a block-by-block basis are denoted with an asterisk (*) above the SNP number. Population haplotype frequencies are displayed on the right of each haplotype combination, while the level of recombination is displayed above the connections between the three blocks. Thick connections represent haplotypes with frequencies ≥5%, while frequencies below 5% are represented by thin lines. CGL: Cancer Genomics Laboratory.

(A)



(B)

s3-6-9-10-11-13-16*	Frequency	Cumulative Frequency
T T T A G A C	0.377	0.377
C C T A G A C	0.253	0.630
T T C G C G T	0.225	0.855
T T C A C G T	0.065	0.920
T T C G G G C	0.044	0.964
C C T G G A C	0.015	0.979
T T C G C A C	0.007	0.986
T C T A G A C	0.007	0.993
C C C A C G T	0.007	1.000

*SNP ID according to Table 2

(C)

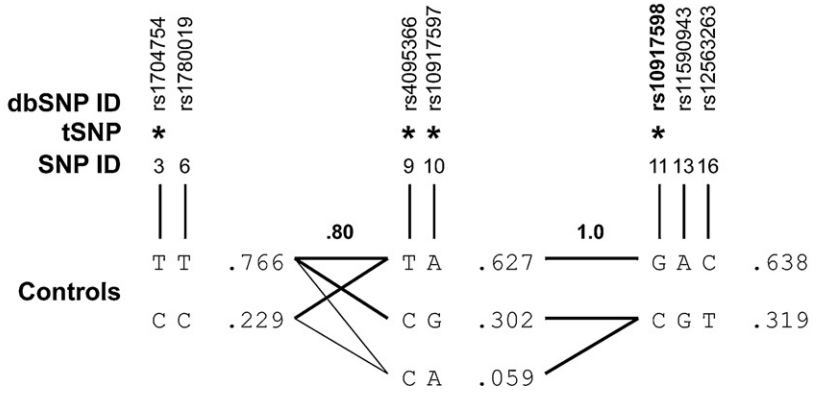
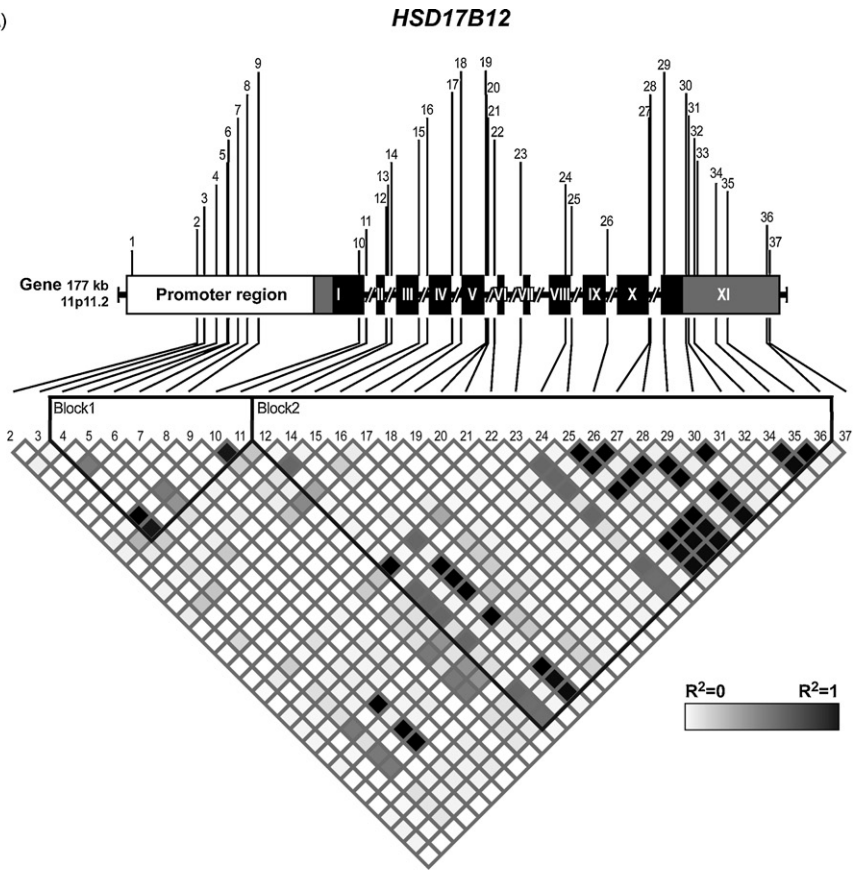


Fig. 4. (A) Pairwise linkage (LD) measures of r^2 for *HSD17B7* sequence variants identified in the control series. Sequence variants are represented by thin lines and are denoted numerically with reference to Table 2. (B) The table denotes haplotype frequencies estimated from *AKR1C3* sequence variants having a MAF ≥5% among the 70 controls. (C) Haplotype blocks predicted from variants identified in French Canadian controls showing a frequency ≥5%. tSNPs identified on a block-by-block basis are denoted with an asterisk (*) above the SNP number. Population haplotype frequencies are displayed on the right of each haplotype combination, while the level of recombination is displayed above the connections between the three blocks. Thick connections represent haplotypes with frequencies ≥5%, while frequencies below 5% are represented by thin lines. CGL: Cancer Genomics Laboratory.

(A)



(B)

s4-6-10-11-12-15-17-23-25-26-27-28-29-30-31-34-35-36*	Frequency	Cumulative Frequency
C T A A G G C C + A A A C A A + – C	0.428	0.428
C T A A A A C T – G G A T A A – + A	0.275	0.703
T T G G A G A T + A A A C A A + – C	0.094	0.797
T C G G A G C T + A A G C C G + – C	0.065	0.862
C T A A A G A T + A A A C A A + – C	0.051	0.913
C T A G A G A T + A A A C A A + – C	0.036	0.949
T T G G G G C C + A A A C A A + – C	0.022	0.971
C T A A A A C T – G G A T A A – – C	0.015	0.986
C T A G A A C T – G G A T A A – + A	0.007	0.993
T T G G A A C T – G G A T A A – + A	0.007	1.000

*SNP ID according to Table 3

(C)

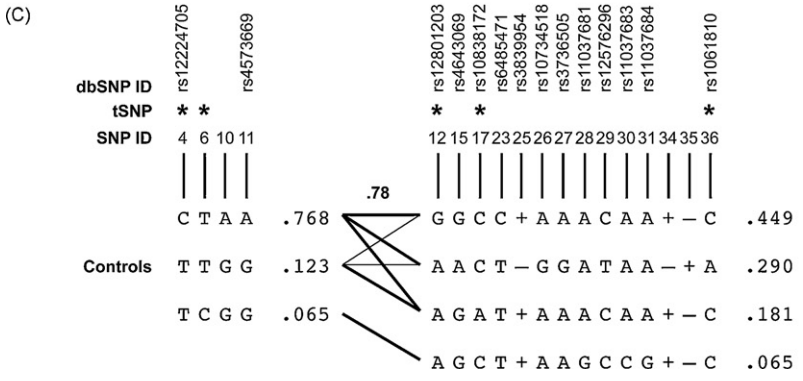


Fig. 5. (A) Pairwise linkage (LD) measures of r^2 for *HSD17B12* sequence variants identified in the control series. Sequence variants are represented by thin lines and are denoted numerically with reference to Table 3. (B) The table denotes haplotype frequencies estimated from *HSD17B12* sequence variants having a MAF ≥5% among the 70 controls. (C) Haplotype blocks predicted from variants identified in French Canadian controls showing a frequency ≥5%. tSNPs identified on a block-by-block basis are denoted with an asterisk (*) above the SNP number. Population haplotype frequencies are displayed on the right of each haplotype combination, while the level of recombination is displayed above the connections between the two blocks. Thick connections represent haplotypes with frequencies ≥5%, while frequencies below 5% are represented by thin lines. CGL: Cancer Genomics Laboratory.

variants of small risk [71–73]. The assessment and confirmation of risks potentially associated with these alleles will necessitate analyses in much larger sample sets [5,73,74].

3.2. LD, haplotype analysis and tSNP identification

Although no likely high penetrance mutations were identified in the *AKR1C3*, *HSD17B7* and *HSD17B12* genes which could be associated with breast cancer risk, the current study provides information on intragenic LD and haplotype diversity and allows an optimal selection of tSNPs, which will reduce genotyping costs and efforts in future association studies in the French-Canadian population without loss of power.

3.2.1. AKR1C3/HSD17B5

Pairwise linkage disequilibrium between all 53 variants identified in the control series and r^2 measures between all variants are illustrated in Fig. 3(panel A). As demonstrated, three LD blocks were identified in the *AKR1C3* gene. The first block encompasses a region from exon 1 to intron 5, the second block represents a region from intron 5 to intron 8, which includes exons 6–8, while the third block includes part of intron 8 and the last exon. Overall, there is no evidence of strong LD between the three blocks in the *AKR1C3* gene. During the course of this study, the Cancer Genetics Markers of Susceptibility study (CGEMS) of the National Cancer Institute performed a genome-wide association study for breast cancer using 1145 breast cancer cases and controls from the Nurses' Health Study. One of the SNPs having a $r^2 = 1$ with p.His5Gln was genotyped in the CGEMS study, which failed to discover any association with breast cancer risk. As illustrated in Fig. 1(Panels C and D), the current study shows a lack of effect of the p.His5Gln variation on enzymatic activity. Another non-synonymous variant, p.Glu77Gly, was previously reported to be associated with circulating testosterone levels [67]. As discussed previously, the assessment of the effect of this variant on the enzymatic activity of the 17β-HSD type 5 enzyme revealed a slightly significant difference in the conversion levels of androstenedione as compared to the wild-type enzyme (Fig. 1, Panels C and D).

Given that the association of a gene with disease may be haplotype-specific, the haplotype diversity of *AKR1C3* was estimated using the 35 common variants genotyped in this study (MAF ≥ 5%). 19 haplotypes were estimated, 6 of which represent 85% of all estimated haplotypes in our French-Canadian control sample set (Fig. 3, panel B). Thereafter, 11 tSNPs were identified in our population and determined to be necessary for the discrimination of all observed haplotypes (Fig. 3, panel C). However, in order to efficiently tag the 6 most common haplotypes only 5 tSNPs are necessary (rs10904415, rs2298305, rs481400, rs9329316, rs7917546). Considering the large number of variants found in this gene, narrowing down the genotyping to only five SNPs provides a valuable, cost-efficient tool for future association studies, at least in French Canadians.

3.2.2. HSD17B7

Three LD blocks were identified for *HSD17B7* in French Canadians as shown in Fig. 4, panel A by the r^2 plots, the majority of variants are not in strong LD, with the exception of four variants with a MAF around 30%. Using variants having a MAF ≥ 5% in healthy individuals, we estimated nine haplotypes (Fig. 4, panel B), four of which had frequencies ≥5% and which represented approximately 92% of all haplotypes estimated in controls. Thereafter, considering haplotypes having a frequency ≥5%, 4 tSNPs were identified in the 3 LD blocks from our population, namely SNP #3 (rs1704754) found in block 1, SNPs #9 (rs4095366) and #10 (rs10917597) in block 2, and SNP #11 (rs10917598) in block 3. All sequence variants with MAF ≥5% identified in our French Canadian cohort have a pairwise $r^2 \geq 0.8$ in association with at least one of the four tSNPs selected. None of the variants observed in the *HSD17B7* gene were genotyped in the CGEMS study.

3.2.3. HSD17B12

As illustrated in Fig. 5(panel A), the majority of variants were in LD and hence two blocks of LD are observed, covering the entire *HSD17B12* gene. Ten haplotypes were estimated (Fig. 5, panel B) and of these, five had frequencies ≥5% and represented approximately 92% of all haplotypes estimated in controls. Thereafter, 5 tSNPs were identified in the 2 LD blocks (Fig. 5, panel C), namely SNP #4 (rs12224705) and #6 in block 1, and SNP #12 (rs12801203), #17 (rs10838172) and #36 (rs1061810) in the second block.

Only two SNPs were genotyped in both our study and the CGEMS study, however, looking at LD between SNPs genotyped in CGEMS, those genotyped in HapMap and those identified in the current study, we are able to make an indirect correlation between the CGEMS SNPs and our variants. More precisely, using 9 of the SNPs genotyped in CGEMS, we were able to represent 84% of frequent haplotypes (≥5%) predicted in the current study. However, it should be noted that 5 of the 18 SNPs with MAFs ≥5%, which were used for haplotype prediction in the present study, could not be represented by any of the CGEMS SNPs, three of which were chosen as tagging SNPs by the Haploview software. Furthermore, although the CGEMS study did not find a significant association between their *HSD17B12* allelic variants and breast cancer risk, it should be noted that the "sporadic" breast cancer cases involved in their study are mostly post-menopausal, while our study only includes cases from high-risk families. Further analyses using larger cohorts, such as those used in recent studies [71,75] are warranted to determine the involvement of *HSD17B12* allelic variants in breast cancer risk. It would also be of importance to perform these large-scale studies in pre-menopausal breast cancer cases, especially taking into consideration the pivotal role of this enzyme in estradiol synthesis.

4. Conclusion

Our analysis does not suggest involvement of the *AKR1C3*, *HSD17B7*, *HSD17B12* genes in breast cancer susceptibility through high-risk alleles. Nonetheless, additional studies are needed to determine the functional contribution of the promoter sequence variants on *AKR1C3* gene expression. Further association studies involving larger cohorts of breast cancer cases are also warranted to establish whether *AKR1C3*, *HSD17B7* and *HSD17B12* could represent low or moderate penetrance genes involved in breast cancer risk.

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 cancérologie Charles Bruneau, Ste-Justine Hospital, for controls DNA samples. We would like to thank Dr Martine Dumont, Gilles Leblanc and Carolle Samson for sample management, mutation screening, and 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 genetic counselling 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. We thank Claudia Moreau at the Centre de Recherche de l'Hôpital Ste-Justine for help with control DNA samples. We also appreciate advice received from ethics committees. This work was supported by the Canadian Institutes of Health Research (CIHR) through the INHERIT BRCAs

research program – Grant #43822 and the Canadian Breast Cancer Research Alliance (CBCRA) – Grants #018361 and #019511. J.S. is Chairholder of the Canada Research Chair in Oncogenetics and F.D. is a recipient of a chercheur-boursier from the Fonds de la Recherche en Santé du Québec (FRSQ).

Appendix A. 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, Canada

Jocelyne Chiquette: Clinique des maladies du sein Deschênes-Fabia, Hôpital du Saint-Sacrement, Québec, Canada

Rachel Laframboise: Medical Genetics Division, Centre Hospitalier Universitaire de Québec, CHUL, Laval University, Québec, Canada

Jean Lépine: Haemato-Oncology Service, Centre Hospitalier Régional de Rimouski, Rimouski, Canada

Bernard Lespérance, Roxane Pichette: Department of Haemato-Oncology, Hôpital du Sacré-Coeur de Montréal, Montréal, Canada

Marie Plante: Gynecology Oncology Division, Hôtel-Dieu de Québec, Centre Hospitalier Universitaire de Québec, Laval University, Québec, Canada

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jsbmb.2009.05.005.

References

[1] D.F. Easton, How many more breast cancer predisposition genes are there? Breast Cancer Res. 1 (1) (1999) 14–17.
[2] B.A. Ponder, Cancer genetics, Nature 411 (6835) (2001) 336–341.
[3] D. Thompson, D. Easton, The genetic epidemiology of breast cancer genes, J. Mammary Gland Biol. Neoplasia 9 (3) (2004) 221–236.
[4] J. Simard, M. Dumont, A.M. Moisan, V. Gaborieau, H. Malouin, F. Durocher, J. Chiquette, M. Plante, D. Avard, P. Bessette, C. Brousseau, M. Dorval, B. Godard, L. Houde, Y. Joly, M.A. Lajoie, G. Leblanc, J. Lepine, B. Lesperance, H. Vezina, J. Parboosingh, R. Pichette, L. Provencher, J. Rheaume, D. Sinnett, C. Samson, J.C. Simard, M. Tranchant, P. Voyer, D. Easton, S.V. Tavtigian, B.M. Knoppers, R. Laframboise, P. Bridge, D. Goldgar, 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) (2007) 107–121.
[5] P.D. Pharoah, A. Antoniou, M. Bobrow, R.L. Zimmern, D.F. Easton, B.A. Ponder, Polygenic susceptibility to breast cancer and implications for prevention, Nat. Genet. 31 (1) (2002) 33–36.
[6] S.F. Doisneau-Sixou, C.M. Sergio, J.S. Carroll, R. Hui, E.A. Musgrove, R.L. Sutherland, Estrogen and antiestrogen regulation of cell cycle progression in breast cancer cells, Endocr. Relat. Cancer 10 (2) (2003) 179–186.
[7] F. Labrie, V. Luu-The, C. Labrie, A. Belanger, J. Simard, S.X. Lin, G. Pelletier, Endocrine and intracrine sources of androgens in women: inhibition of breast cancer and other roles of androgens and their precursor dehydroepiandrosterone, Endocr. Rev. 24 (2) (2003) 152–182.
[8] H. Seeger, A.O. Mueck, Are the progestins responsible for breast cancer risk during hormone therapy in the postmenopause? Experimental vs. clinical data, J. Steroid Biochem. Mol. Biol. 109 (1–2) (2008) 11–15.
[9] A. Fournier, A. Fabre, S. Mesrine, M.C. Boutron-Ruault, F. Berrino, F. Clavel-Chapelon, Use of different postmenopausal hormone therapies and risk of histology- and hormone receptor-defined invasive breast cancer, J. Clin. Oncol. 26 (8) (2008) 1260–1268.
[10] M.J. Lewis, J.P. Wiebe, J.G. Heathcote, Expression of progesterone metabolizing enzyme genes (AKR1C1, AKR1C2, AKR1C3, SRD5A1, SRD5A2) is altered in human breast carcinoma, BMC Cancer 4 (2004) 27.
[11] R. Shibuya, T. Suzuki, Y. Miki, K. Yoshida, T. Moriya, K. Ono, J. Akahira, T. Ishida, H. Hirakawa, D.B. Evans, H. Sasano, Intratumoral concentration of sex steroids and expression of sex steroid-producing enzymes in ductal carcinoma in situ of human breast, Endocr. Relat. Cancer 15 (1) (2008) 113–124.
[12] J.M. Jez, T.G. Flynn, T.M. Penning, A new nomenclature for the aldo-keto reductase superfamily, Biochem. Pharmacol. 54 (6) (1997) 639–647.
[13] T.M. Penning, M.E. Burczynski, J.M. Jez, C.F. Hung, H.K. Lin, H. Ma, M. Moore, N. Palackal, K. Ratnam, Human 3alpha-hydroxysteroid dehydrogenase isoforms (AKR1C1–AKR1C4) of the aldo-keto reductase superfamily: functional plasticity and tissue distribution reveals roles in the inactivation and formation of male and female sex hormones, Biochem. J. 351 (Pt 1) (2000) 67–77.
[14] I. Dufort, P. Rheault, X.F. Huang, P. Soucy, V. Luu-The, Characteristics of a highly labile human type 5 17beta-hydroxysteroid dehydrogenase, Endocrinology 140 (2) (1999) 568–574.
[15] T.M. Penning, M.E. Burczynski, J.M. Jez, H.K. Lin, H. Ma, M. Moore, K. Ratnam, N. Palackal, Structure-function aspects and inhibitor design of type 5 17beta-hydroxysteroid dehydrogenase (AKR1C3), Mol. Cell Endocrinol. 171 (1–2) (2001) 137–149.
[16] J.P. Wiebe, D. Muzia, J. Hu, D. Szwajcer, S.A. Hill, J.L. Seachrist, The 4-pregnene and 5alpha-pregnane progesterone metabolites formed in nontumorous and tumorous breast tissue have opposite effects on breast cell proliferation and adhesion, Cancer Res. 60 (4) (2000) 936–943.
[17] V. Luu-The, Analysis and characteristics of multiple types of human 17beta-hydroxysteroid dehydrogenase, J. Steroid Biochem. Mol. Biol. 76 (1–5) (2001) 143–151.
[18] G. Seth, R.S. McIvor, W.S. Hu, 17Beta-hydroxysteroid dehydrogenase type 7 (Hsd17b7) reverts cholesterol auxotrophy in NS0 cells, J. Biotechnol. 121 (2) (2006) 241–252.
[19] Z. Marijanovic, D. Laubner, G. Moller, C. Gege, B. Husen, J. Adamski, R. Breitling, Closing the gap: identification of human 3-ketosteroid reductase, the last unknown enzyme of mammalian cholesterol biosynthesis, Mol. Endocrinol. 17 (9) (2003) 1715–1725.
[20] A. Krazeisen, R. Breitling, K. Imai, S. Fritz, G. Moller, J. Adamski, Determination of cDNA, gene structure and chromosomal localization of the novel human 17beta-hydroxysteroid dehydrogenase type 7(1), FEBS Lett. 460 (2) (1999) 373–379.
[21] D. Song, G. Liu, V. Luu-The, D. Zhao, L. Wang, H. Zhang, G. Xueling, S. Li, L. Desy, F. Labrie, G. Pelletier, Expression of aromatase and 17beta-hydroxysteroid dehydrogenase types 1, 7 and 12 in breast cancer. An immunocytochemical study, J. Steroid. Biochem. Mol. Biol. 101 (2–3) (2006) 136–144.
[22] V. Luu-The, P. Tremblay, F. Labrie, Characterization of type 12 17beta-hydroxysteroid dehydrogenase, an isoform of type 3 17beta-hydroxysteroid dehydrogenase responsible for estradiol formation in women, Mol. Endocrinol. 20 (2) (2006) 437–443.
[23] Y.A. Moon, J.D. Horton, Identification of two mammalian reductases involved in the two-carbon fatty acyl elongation cascade, J. Biol. Chem. 278 (9) (2003) 7335–7343.
[24] N. Sakurai, Y. Miki, T. Suzuki, K. Watanabe, T. Narita, K. Ando, T.M. Yung, D. Aoki, H. Sasano, H. Handa, Systemic distribution and tissue localizations of human 17beta-hydroxysteroid dehydrogenase type 12, J. Steroid. Biochem. Mol. Biol. 99 (4–5) (2006) 174–181.
[25] O.M. Sinilnikova, J.D. McKay, S.V. Tavtigian, F. Canzian, D. DeSilva, C. Biessy, S. Monnier, L. Dossus, C. Boillot, L. Gioia, D.J. Hughes, M.K. Jensen, K. Overvad, A. Tjonneland, A. Olsen, F. Clavel-Chapelon, V. Chajes, V. Joulin, J. Linseisen, J. Chang-Claude, H. Boeing, S. Dahm, A. Trichopoulou, D. Trichopoulos, M. Koliva, K.T. Khaw, S. Bingham, N.E. Allen, T. Key, D. Palli, S. Panico, F. Berrino, R. Tumino, P. Vineis, H.B. Bueno-de-Mesquita, P.H. Peeters, C.H. van Gils, E. Lund, G. Pera, J.R. Quiros, M. Dorronsoro, C. Martinez Garcia, M.J. Tormo, E. Ardanaz, G. Hallmans, P. Lenner, G. Berglund, J. Manjer, E. Riboli, G.M. Lenoir, R. Kaaks, Haplotype-based analysis of common variation in the acetyl-coA carboxylase alpha gene and breast cancer risk: a case-control study nested within the European Prospective Investigation into Cancer and Nutrition, Cancer Epidemiol. Biomarkers Prev. 16 (3) (2007) 409–415.
[26] A.L. Dixon, L. Liang, M.F. Moffatt, W. Chen, S. Heath, K.C. Wong, J. Taylor, E. Burnett, I. Gut, M. Farrall, G.M. Lathrop, G.R. Abecasis, W.O. Cookson, A genome-wide association study of global gene expression, Nat. Genet. 39 (10) (2007) 1202–1207.
[27] M. Morley, C.M. Molony, T.M. Weber, J.L. Devlin, K.G. Ewens, R.S. Spielman, V.G. Cheung, Genetic analysis of genome-wide variation in human gene expression, Nature 430 (7001) (2004) 743–747.
[28] H.S. Feigelson, D.G. Cox, H.M. Cann, S. Wacholder, R. Kaaks, B.E. Henderson, D. Albanes, D. Altshuler, G. Berglund, F. Berrino, S. Bingham, J.E. Buring, N.P. Burtt, E.E. Calle, S.J. Chanock, F. Clavel-Chapelon, G. Colditz, W.R. Diver, M.L. Freedman, C.A. Haiman, S.E. Hankinson, R.B. Hayes, J.N. Hirschhorn, D. Hunter, L.N. Kolonel, P. Kraft, L. LeMarchand, J. Linseisen, W. Modi, C. Navarro, P.H. Peeters, M.C. Pike, E. Riboli, V.W. Setiawan, D.O. Stram, G. Thomas, M.J. Thun, A. Tjonneland, D. Trichopoulos, Haplotype analysis of the HSD17B1 gene and risk of breast cancer: a comprehensive approach to multicenter analyses of prospective cohort studies, Cancer Res. 66 (4) (2006) 2468–2475.
[29] H.S. Feigelson, R. McKean-Cowdin, G.A. Coetzee, D.O. Stram, L.N. Kolonel, B.E. Henderson, Building a multigenic model of breast cancer susceptibility: CYP17 and HSD17B1 are two important candidates, Cancer Res. 61 (2) (2001) 785–789.
[30] A. Mannermaa, H. Peltoketo, R. Winqvist, B.A. Ponder, H. Kiviniemi, D.F. Easton, M. Poutanen, V. Isomaa, R. Vihko, Human familial and sporadic breast cancer: analysis of the coding regions of the 17 beta-hydroxysteroid dehydrogenase 2 gene (EDH17B2) using a single-strand conformation polymorphism assay, Hum. Genet. 93 (3) (1994) 319–324.
[31] M. Plourde, C. Samson, F. Durocher, O. Sinilnokova, J. Simard, Characterization of HSD17B1 sequence variants in breast cancer cases from French Canadian families with high risk of breast and ovarian cancer, J. Steroid. Biochem. Mol. Biol. 109 (1–2) (2008) 115–128.
[32] V.W. Setiawan, S.E. Hankinson, G.A. Colditz, D.J. Hunter, I. De Vivo, HSD17B1 gene polymorphisms and risk of endometrial and breast cancer, Cancer Epidemiol. Biomarkers Prev. 13 (2) (2004) 213–219.
[33] A.H. Wu, A. Seow, K. Arakawa, D. Van Den Berg, H.P. Lee, M.C. Yu, HSD17B1 and CYP17 polymorphisms and breast cancer risk among Chinese women in Singapore, Int. J. Cancer 104 (4) (2003) 450–457.

[34] M.M. Gaudet, S. Chanock, A. Dunning, K. Driver, L.A. Brinton, J. Lissowska, B. Peplonska, P. Pharoah, M. Garcia-Closas, HSD17B1 genetic variants and hormone receptor-defined breast cancer, Cancer Epidemiol. Biomarkers Prev. 17 (10) (2008) 2766–2772.
[35] A. Jansson, J. Carlsson, A. Olsson, P. Storm, S. Margolin, C. Gunnarsson, M. Stenmark-Askmalm, A. Lindblom, B. Persson, O. Stal, A new polymorphism in the coding region of exon four in HSD17B2 in relation to risk of sporadic and hereditary breast cancer, Breast Cancer Research and Treatment, January 27 (2007) (epub ahead of print). doi:10.1007/s10549-10006-19477-10544.
[36] M. Plourde, C. Manhes, G. Leblanc, F. Durocher, M. Dumont, O. Sinilnikova, J. Simard, Mutation analysis and characterization of HSD17B2 sequence variants in breast cancer cases from French Canadian families with high risk of breast and ovarian cancer, J. Mol. Endocrinol. 40 (4) (2008) 161–172.
[37] A.C. Antoniou, F. Durocher, P. Smith, J. Simard, D.F. Easton, 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) (2006) R3.
[38] A.C. Antoniou, D.F. Easton, Polygenic inheritance of breast cancer: implications for design of association studies, Genet. Epidemiol. 25 (3) (2003) 190–202.
[39] F. Durocher, Y. Labrie, G. Ouellette, J. Simard, Genetic sequence variations and ADPRT haplotype analysis in French Canadian families with high risk of breast cancer, J. Hum. Genet. 52 (12) (2007) 963–977.
[40] F. Durocher, Y. Labrie, P. Soucy, O. Sinilnikova, D. Labuda, P. Bessette, J. Chiquette, R. Laframboise, J. Lepine, B. Lesperance, G. Ouellette, R. Pichette, M. Plante, S.V. Tavtigian, J. Simard, Mutation analysis and characterization of ATR sequence variants in breast cancer cases from high-risk French Canadian breast/ovarian cancer families, BMC Cancer 6 (2006) 230–250.
[41] A.M. Moisan, J. Fortin, M. Dumont, C. Samson, P. Bessette, J. Chiquette, R. Laframboise, J. Lepine, B. Lesperance, R. Pichette, M. Plante, L. Provencher, P. Voyer, D. Goldgar, P. Bridge, J. Simard, No evidence of BRCA1/2 genomic rearrangements in high-risk French-Canadian breast/ovarian cancer families, Genet. Test 10 (2) (2006) 104–115.
[42] J.C. Barrett, B. Fry, J. Maller, M.J. Daly, Haploview: analysis and visualization of LD and haplotype maps, Bioinformatics 21 (2) (2005) 263–265.
[43] M. Stephens, N.J. Smith, P. Donnelly, A new statistical method for haplotype reconstruction from population data, Am. J. Hum. Genet. 68 (4) (2001) 978–989.
[44] S.B. Gabriel, S.F. Schaffner, H. Nguyen, J.M. Moore, J. Roy, B. Blumenstiel, J. Higgins, M. DeFelice, A. Lochner, M. Faggart, S.N. Liu-Cordero, C. Rotimi, A. Adeyemo, R. Cooper, R. Ward, E.S. Lander, M.J. Daly, D. Altshuler, The structure of haplotype blocks in the human genome, Science 296 (5576) (2002) 2225–2229.
[45] K. Cartharius, K. Frech, K. Grote, B. Klocke, M. Haltmeier, A. Klingenhoff, M. Frisch, M. Bayerlein, T. Werner, MatInspector and beyond: promoter analysis based on transcription factor binding sites, Bioinformatics 21 (13) (2005) 2933–2942.
[46] M.B. Shapiro, P. Senapathy, RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression, Nucleic Acids Res. 15 (17) (1987) 7155–7174.
[47] L. Cartegni, J. Wang, Z. Zhu, M.Q. Zhang, A.R. Krainer, ESEfinder: A web resource to identify exonic splicing enhancers, Nucleic Acids Res. 31 (13) (2003) 3568–3571.
[48] P.C. Ng, S. Henikoff, Predicting deleterious amino acid substitutions, Genome Res. 11 (5) (2001) 863–874.
[49] V. Ramensky, P. Bork, S. Sunyaev, Human non-synonymous SNPs: server and survey, Nucleic Acids Res. 30 (17) (2002) 3894–3900.
[50] J.J. Tremblay, R.S. Viger, Novel roles for GATA transcription factors in the regulation of steroidogenesis, J. Steroid. Biochem. Mol. Biol. 85 (2–5) (2003) 291–298.
[51] Y.S. Piao, H. Peltoketo, P. Vihko, R. Vihko, The proximal promoter region of the gene encoding human 17beta-hydroxysteroid dehydrogenase type 1 contains GATA, AP-2, and Sp1 response elements: analysis of promoter function in choriocarcinoma cells, Endocrinology 138 (8) (1997) 3417–3425.
[52] L. Peng, Y. Huang, F. Jin, S.W. Jiang, A.H. Payne, Transcription enhancer factor-5 and a GATA-like protein determine placental-specific expression of the Type I human 3beta-hydroxysteroid dehydrogenase gene, HSD3B1, Mol. Endocrinol. 18 (8) (2004) 2049–2060.
[53] L.J. Martin, H. Taniguchi, N.M. Robert, J. Simard, J.J. Tremblay, R.S. Viger, GATA factors and the nuclear receptors, steroidogenic factor 1/liver receptor homolog 1, are key mutual partners in the regulation of the human 3beta-hydroxysteroid dehydrogenase type 2 promoter, Mol. Endocrinol. 19 (9) (2005) 2358–2370.
[54] C.E. Fluck, W.L. Miller, GATA-4 and GATA-6 modulate tissue-specific transcription of the human gene for P450c17 by direct interaction with Sp1, Mol. Endocrinol. 18 (5) (2004) 1144–1157.
[55] P. Jimenez, K. Saner, B. Mayhew, W.E. Rainey, GATA-6 is expressed in the human adrenal and regulates transcription of genes required for adrenal androgen biosynthesis, Endocrinology 144 (10) (2003) 4285–4288.
[56] J.J. Tremblay, R.S. Viger, GATA factors differentially activate multiple gonadal promoters through conserved GATA regulatory elements, Endocrinology 142 (3) (2001) 977–986.
[57] M.L. Asselin-Labat, K.D. Sutherland, H. Barker, R. Thomas, M. Shackleton, N.C. Forrest, L. Hartley, L. Robb, F.G. Grosveld, J. van der Wees, G.J. Lindeman, J.E. Visvader, Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation, Nat. Cell Biol. 9 (2) (2007) 201–209.
[58] H. Kouros-Mehr, S.K. Bechis, E.M. Slorach, L.E. Littlepage, M. Egeblad, A.J. Ewald, S.Y. Pai, I.C. Ho, Z. Werb, GATA-3 links tumor differentiation and dissemination in a luminal breast cancer model, Cancer Cell 13 (2) (2008) 141–152.
[59] H. Kouros-Mehr, J.W. Kim, S.K. Bechis, Z. Werb, GATA-3 and the regulation of the mammary luminal cell fate, Curr. Opin. Cell Biol. 20 (2) (2008) 164–170.
[60] K. Qin, D.A. Ehrmann, N. Cox, S. Refetoff, R.L. Rosenfield, Identification of a functional polymorphism of the human type 5 17beta-hydroxysteroid dehydrogenase gene associated with polycystic ovary syndrome, J. Clin. Endocrinol. Metab. 91 (1) (2006) 270–276.
[61] O.J. Lehmann, J.C. Sowden, P. Carlsson, T. Jordan, S.S. Bhattacharya, Fox's in development and disease, Trends Genet. 19 (6) (2003) 339–344.
[62] J.C. Kiefer, Back to basics: SOX genes, Dev. Dyn. 236 (8) (2007) 2356–2366.
[63] J.D. Graham, S.M. Hunt, N. Tran, C.L. Clarke, Regulation of the expression and activity by progestins of a member of the SOX gene family of transcriptional modulators, J. Mol. Endocrinol. 22 (3) (1999) 295–304.
[64] N. Srour, M.A. Reymond, R. Steinert, Lost in translation? A systematic database of gene expression in breast cancer, Pathobiology 75 (2) (2008) 112–118.
[65] P.C. Ng, S. Henikoff, Accounting for human polymorphisms predicted to affect protein function, Genome Res. 12 (3) (2002) 436–446.
[66] Q. Lan, J.L. Mumford, M. Shen, D.M. Demarini, M.R. Bonner, X. He, M. Yeager, R. Welch, S. Chanock, L. Tian, R.S. Chapman, T. Zheng, P. Keohavong, N. Caporaso, N. Rothman, Oxidative damage-related genes AKR1C3 and OGG1 modulate risks for lung cancer due to exposure to PAH-rich coal combustion emissions, Carcinogenesis 25 (11) (2004) 2177–2181.
[67] J. Jakobsson, E. Palonek, M. Lorentzon, C. Ohlsson, A. Rane, L. Ekstrom, A novel polymorphism in the 17beta-hydroxysteroid dehydrogenase type 5 (aldo-keto reductase 1C3) gene is associated with lower serum testosterone levels in caucasian men, Pharmacogenomics J. 7 (4) (2007) 282–289.
[68] U.C. Oppermann, C. Filling, H. Jornvall, Forms and functions of human SDR enzymes, Chem. Biol. Interact. 130–132 (1–3) (2001) 699–705.
[69] A.C. Antoniou, D.E. Goldgar, N. Andrieu, J. Chang-Claude, R. Brohet, M.A. Rookus, D.F. Easton, A weighted cohort approach for analysing factors modifying disease risks in carriers of high-risk susceptibility genes, Genet. Epidemiol. 29 (1) (2005) 1–11.
[70] J. Peto, T.M. Mack, High constant incidence in twins and other relatives of women with breast cancer, Nat. Genet. 26 (4) (2000) 411–414.
[71] A. Cox, A.M. Dunning, M. Garcia-Closas, S. Balasubramanian, M.W. Reed, K.A. Pooley, S. Scollen, C. Baynes, B.A. Ponder, S. Chanock, J. Lissowska, L. Brinton, B. Peplonska, M.C. Southey, J.L. Hopper, M.R. McCredie, G.G. Giles, O. Fletcher, N. Johnson, I. dos Santos Silva, L. Gibson, S.E. Bojesen, B.G. Nordestgaard, C.K. Axelsson, D. Torres, U. Hamann, C. Justenhoven, H. Brauch, J. Chang-Claude, S. Kropp, A. Risch, S. Wang-Gohrke, P. Schurmann, N. Bogdanova, T. Dork, R. Fagerholm, K. Aaltonen, C. Blomqvist, H. Nevanlinna, S. Seal, A. Renwick, M.R. Stratton, N. Rahman, S. Sangrajrang, D. Hughes, F. Odefrey, P. Brennan, A.B. Spurdle, G. Chenevix-Trench, J. Beesley, A. Mannermaa, J. Hartikainen, V. Kataja, V.M. Kosma, F.J. Couch, J.E. Olson, E.L. Goode, A. Broeks, M.K. Schmidt, F.B. Hogervorst, L.J. Van't Veer, D. Kang, K.Y. Yoo, D.Y. Noh, S.H. Ahn, S. Wedren, P. Hall, Y.L. Low, J. Liu, R.L. Milne, G. Ribas, A. Gonzalez-Neira, J. Benitez, A.J. Sigurdson, D.L. Stredrick, B.H. Alexander, J.P. Struewing, P.D. Pharoah, D.F. Easton, A common coding variant in CASP8 is associated with breast cancer risk, Nat. Genet. 39 (3) (2007) 352–358.
[72] D.F. Easton, K.A. Pooley, A.M. Dunning, P.D. Pharoah, D. Thompson, D.G. Ballinger, J.P. Struewing, J. Morrison, H. Field, R. Luben, N. Wareham, S. Ahmed, C.S. Healey, R. Bowman, K.B. Meyer, C.A. Haiman, L.K. Kolonel, B.E. Henderson, L. Le Marchand, P. Brennan, S. Sangrajrang, V. Gaborieau, F. Odefrey, C.Y. Shen, P.E. Wu, H.C. Wang, D. Eccles, D.G. Evans, J. Peto, O. Fletcher, N. Johnson, S. Seal, M.R. Stratton, N. Rahman, G. Chenevix-Trench, S.E. Bojesen, B.G. Nordestgaard, C.K. Axelsson, M. Garcia-Closas, L. Brinton, S. Chanock, J. Lissowska, B. Peplonska, H. Nevanlinna, R. Fagerholm, H. Eerola, D. Kang, K.Y. Yoo, D.Y. Noh, S.H. Ahn, D.J. Hunter, S.E. Hankinson, D.G. Cox, P. Hall, S. Wedren, J. Liu, Y.L. Low, N. Bogdanova, P. Schurmann, T. Dork, R.A. Tollenaar, C.E. Jacobi, P. Devilee, J.G. Klijn, A.J. Sigurdson, M.M. Doody, B.H. Alexander, J. Zhang, A. Cox, I.W. Brock, G. MacPherson, M.W. Reed, F.J. Couch, E.L. Goode, J.E. Olson, H. Meijers-Heijboer, A. van den Ouweland, A. Uitterlinden, F. Rivadeneira, R.L. Milne, G. Ribas, A. Gonzalez-Neira, J. Benitez, J.L. Hopper, M. McCredie, M. Southey, G.G. Giles, C. Schroen, C. Justenhoven, H. Brauch, U. Hamann, Y.D. Ko, A.B. Spurdle, J. Beesley, X. Chen, A. Mannermaa, V.M. Kosma, V. Kataja, J. Hartikainen, N.E. Day, D.R. Cox, B.A. Ponder, Genome-wide association study identifies novel breast cancer susceptibility loci, Nature 447 (7148) (2007) 1087–1093.
[73] M.J. Khoury, J. Little, M. Gwinn, J.P. Ioannidis, On the synthesis and interpretation of consistent but weak gene-disease associations in the era of genome-wide association studies, Int. J. Epidemiol. 36 (2) (2007) 439–445.
[74] A.C. Antoniou, D.F. Easton, Models of genetic susceptibility to breast cancer, Oncogene 25 (43) (2006) 5898–5905.
[75] B. Frank, M. Wiestler, S. Kropp, K. Hemminki, A.B. Spurdle, C. Sutter, B. Wappenschmidt, X. Chen, J. Beesley, J.L. Hopper, A. Meindl, M. Kiechle, T. Slanger, P. Bugert, R.K. Schmutzler, C.R. Bartram, D. Flesch-Janys, E. Mutschelknauss, K. Ashton, R. Salazar, E. Webb, U. Hamann, H. Brauch, C. Justenhoven, Y.D. Ko, T. Bruning, S. Silva Idos, N. Johnson, P.P. Pharoah, A.M. Dunning, K.A. Pooley, J. Chang-Claude, D.F. Easton, J. Peto, R. Houlston, G. Chenevix-Trench, O. Fletcher, B. Burwinkel, Association of a common AKAP9 variant with breast cancer risk: a collaborative analysis, J. Natl. Cancer Inst. 100 (6) (2008) 437–442.
[76] International HapMap Project, http://www.hapmap.org.
[77] The European Prospective Investigation into Cancer and Nutrition (EPIC), http://epic.iarc.fr/.
[78] Cancer Genetic Markers of Susceptibility Project, http://cgems.cancer.gov/.
[79] The Perlegen European Panel, http://www.ncbi.nlm.nih.gov.