Introduction

Breast cancer is one of the most common diseases among women in developed countries. Several breast cancer susceptibility genes have been identified, the most important being the high risk tumour suppressor genes BRCA1 (1) and BRCA2 (2), the latter also identified as FANCD1 (3,4). However, under the polygenic model of complex diseases, there are also low risk genes with variants that are common in the general population and that produce a mild susceptibility to the disease. Whether these variants interact together or with other environmental and lifestyle factors in an additive or multiplicative way, is not yet clear (5–8).

Abbreviations: FA, Fanconi anaemia; DSB, double-strand break; SNP, single nucleotide polymorphism. HR, homologous recombination; LD, linkage disequilibrium; MAF, minor allele frequency.

FANCD2 associated with sporadic breast cancer risk

The BRCA1 and BRCA2 (FANCD1) proteins are implicated in several cellular mechanisms, the most relevant being double-strand breaks (DSBs) by homologous recombination (HR) (9,10). Both BRCA1 and BRCA2 interact with the RAD51 protein, a key component of HR DNA damage repair, to form the nuclear repair foci after DNA damage (11–14). Furthermore, the FA protein family has been identified as a component of this DSB HR repair machinery (15–17). To date, 12 complementation groups have been described, for which 11 responsible FA genes have been identified: FANC-A, -B, -C, -D1, -D2, -E, -F, -G, -J, -L and -M. Eight of these (FANC-A, -B, -C, -E, -F, -G, -L and -M) form a multi-subunit nuclear FA complex (18–20). Some of the FA proteins are structural proteins in the FA complex. For example, FANCA acts as dimer, FANC-G plays an important role in the maintenance of the complex structure (21,20) and FANCF functions as a flexible adaptor protein which is required for the assembly of this complex (22). Other FA proteins have specific functions such as FANCL, a ubiquitin ligase (23,24), whose stability and function depends on FANC-B (25,26). Another example is FANCD2, a monoubiquitinated protein, which is activated by FANCL and passes from isoform FANCD2-S to FANCD2-L (27,16). FANCD2 is located downstream of the FA complex and linked with it through FANCE (20). FANCI, described previously as helicase BACH1/BRIP1 (28,29), and FANCM, a nuclease that may act as an engine that translocates the FA core complex along DNA, have been most recently identified (30).

The pathway activity that implicates FA proteins is initiated when DNA damage occurs, and ATM and ATR protein kinases function as DNA damage sensors (31,32). They phosphorylate FANCD2 which is required for the activation of an S-phase checkpoint (33,34) and for its appropriate monoubiquitination through the FA complex (35). They also phosphorylate several proteins in the FA complex (36) which are necessary for the activation of FANCL ubiquitin ligase activity. After monoubiquitination of FANCD2-S, FANCD2-L translocates to the nuclear repair complex foci, helped by BRCA1 in its formation and accumulation (37), where BRCA2 and RAD51 are also located (27). Thus, the monoubiquitination of FANCD2 is important because it allows its chromatin-associated foci assembly (38,39), either in response to DNA damage (40) during the S-phase of the normal cell cycle (41,42), or in replicative stress (43). It has therefore been proposed that FANCD2 could mediate the BRCA2–RAD51 interaction and also influence the speed and efficiency of DNA repair (17,40,44) although FANC-G may also be implicated in this process (45). However, an alternative model proposes that FANCD2 functions in a DNA repair sub-pathway that is independent of DNA crossover and sister chromatid exchange, where RAD51 focus formation depends on BRCA2, but not FANCD2 (46). It has been demonstrated that FA proteins are not essential for HR repair machinery. Furthermore, FA cells have a mild defect in the single-strand annealing pathway, which is also
involved in DSB repair. Thus, FA proteins could work in other DNA damage repair pathways of HR (47).

We focused our study on the hypothesis that genes involved in the mechanism of DNA damage repair, acting similarly to tumour suppressors, would act as low susceptibility genes in the development of sporadic breast cancer, as proposed previously (48). In this study, we analysed seven genetic variants in three FA genes: two variants in FANCA, three in FANCL, and two in FANC D2. The first two genes encode proteins in the FA complex and the latter encodes a protein downstream of the FA complex.

Materials and methods

Subjects

This case–control study included a total of 897 consecutive and non-related sporadic breast cancer cases and 1033 healthy controls. The case samples had a mean age of diagnosis of 50 years (range 23–90 years) (Table I). They were collected from seven Spanish public hospitals: 258 (29%) from Hospital Monte Naranco, in Oviedo; 164 (18%) from Fundación Jiménez Díaz, 132 (15%) from Hospital La Paz and 100 (11%) from Hospital Clínico San Carlos, all in Madrid; 115 (13%) from Hospital San Pau, in Barcelona; 100 (11%) from Hospital de Salamanca, in Salamanca; 28 (3%) from Hospital de Valencia, in Valencia. Control samples were collected from seven Spanish public hospitals: 258 (29%) from Hospital Monte Naranco, in Oviedo; 164 (18%) from Fundación Jiménez Díaz, 132 (15%) from Hospital La Paz and 100 (11%) from Hospital Clínico San Carlos, all in Madrid; 115 (13%) from Hospital San Pau, in Barcelona; 100 (11%) from Hospital de Salamanca, in Salamanca; 28 (3%) from Hospital de Valencia, in Valencia. Controls (between 22 and 85 years of age) were unaffected Spanish women collected from six centres: 462 (45%) from the Menopause Research Centre at the Instituto Palafoxios; 234 (28%) from the National Blood Transfusion Centre; 231 (21%) from Fundación Jiménez Díaz; 63 (6%) from Hospital Clínico San Carlos, all in Madrid.

Extraction of genomic DNA

Genomic DNA from cases and controls was isolated from peripheral blood lymphocytes using the MagNA Pure LC Instrument (Roche Molecular Biochemicals, Germany) following manufacturer’s instructions (49), and a stock solution of DNA at 50 ng/μl was obtained, measured using the PicoGreen dsDNA Quantiﬁcation Kit (Molecular Probes, Eugene, OR).

SNP selection

Three public databases available via internet were used to collect information about single nucleotide polymorphisms (SNPs) in the three genes (FANCA, FANCL and FANC D2): NCBI (http://www.ncbi.nlm.nih.gov), Ensembl (http://www.ensembl.org) and HapMap (http://www.hapmap.org). SNPs were considered if they had a minor allele frequency (MAF) greater than or equal to 5%. Information on linkage disequilibrium (LD) (blocks within chromosomes that are more probably linked and segregate together than expected at random), was used to select representative SNPs in high LD blocks. Haploview v3.11 (50) was used to select tag-SNPs (a subset of SNPs that represent variation in a gene or chromosomal region) where possible.

Genotyping assays

Genotyping was carried out using the TaqMan platform. SNPs were designed using Applied Biosystems Assay-by-design and Assay-on-Demand probes (Applied Biosystems, Foster City, CA, USA) (primer and probe information is provided in Supplementary Table I). PCR amplifications were carried out following the manufacturer’s standardized protocol. The genotype of each sample was automatically assigned by measuring allele-specific ﬂuorescence in the ABI Prism 7900HT Detection System, using the SDVs.2.1 (Sequence Detection Software) for allele discrimination, by Applied Biosystems. Three non-DNA and nine duplicate samples were included in each plate of 384 samples used, as an internal control.

Statistical analysis

Deviations from Hardy–Weinberg equilibrium were tested using the genhwi command in STATAv8.0 (51), applying Pearson’s Chi-squared test or Fisher’s Exact Test where appropriate. Differences in the MAFs between cases and controls were tested using the Chi-squared test. In order to assess associations between genotypes, haplotypes (combination of variants along a chromosome) and diplotypes (haplotype combinations in individuals), and breast cancer risk, several analyses were performed. PHASE,2.0 software (52,53) was used to impute haplotypes and compare their frequency distributions in cases and controls. Unconditional logistic regression was applied using STATAv8.0 for the analysis of genotypes, haplotypes and diplotypes, the most likely imputed haplotype being assumed for the latter two. Odds ratios (OR), their 95% conﬁdence intervals (CI) and Wald statistic P-values were estimated using the most frequent genotype, haplotype and diplotype, respectively, among controls as reference. In the analysis of genotypes, dominant, recessive and multiplicative (single parameter) codominant models were assessed against the two parameter codominant model via the likelihood ratio test, and where the latter was not the best fit (nominal P-value greater than 0.05), the best fitting model was determined by parsimony. Age was included as a categorical variable (<40, 40–44, 45–49, 50–54, 55–59, 60–64, 65+) in the analysis of haplotypes. Analysis of haplotypes was repeated using the haplo.stats library implemented in R which compares haplotype frequencies in cases and controls in an unbiased way by including haplotype uncertainty in the estimation of ORs. For all statistical tests, results were considered signiﬁcant for two-sided P-values less than 0.05.

Table I. Age categories of control and case samples

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Controls</th>
<th>Cases</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>N (%)</td>
<td>N (%)</td>
<td>N (%)</td>
<td></td>
</tr>
<tr>
<td>&lt;40</td>
<td>261 (25)</td>
<td>209 (23)</td>
<td>470 (24)</td>
</tr>
<tr>
<td>40–44</td>
<td>61 (6)</td>
<td>30 (3)</td>
<td>91 (5)</td>
</tr>
<tr>
<td>45–49</td>
<td>90 (9)</td>
<td>45 (5)</td>
<td>135 (7)</td>
</tr>
<tr>
<td>50–54</td>
<td>156 (15)</td>
<td>72 (8)</td>
<td>228 (12)</td>
</tr>
<tr>
<td>55–59</td>
<td>147 (14)</td>
<td>74 (8)</td>
<td>221 (12)</td>
</tr>
<tr>
<td>60–64</td>
<td>91 (9)</td>
<td>72 (8)</td>
<td>163 (8)</td>
</tr>
<tr>
<td>65+</td>
<td>22 (2)</td>
<td>112 (13)</td>
<td>134 (7)</td>
</tr>
<tr>
<td>Unknown</td>
<td>205 (20)</td>
<td>283 (32)</td>
<td>488 (25)</td>
</tr>
<tr>
<td>Total</td>
<td>1033 (100)</td>
<td>897 (100)</td>
<td>1930 (100)</td>
</tr>
</tbody>
</table>

Screening for functional variants in the FANC D2 gene

Haploview v3.11 was used to graphically analyse the extent of LD in the FANC D2 region (Figure 1B). Selected sequences in FANC D2 (promoter region and nine exons) were further analysed by PCR, dHPLC and sequencing in samples carrying haplotype H2 in the FANC D2 gene. Primers in flanking exons, intron-exon boundaries and the promoter region were designed using Primer3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/ primer3 www.cgi) (54), based on genomic information collected from NCBI (http://www.ncbi.nlm.nih.gov). PCR products, obtained under standardized conditions, were analysed by dHPLC using the HAPLOMAKER software from Transgenomics (PCR primers and conditions, and dHPLC elution conditions are showed in Supplementary Table II). Those DNAs with heterozygous peaks detected by dHPLC analysis were purified using the High Pure PCR Product Purification Kit (Roche) and sequenced with the ABI Prism Big DyeTM System and ABI 3700 sequencer (Applied Biosystems, Foster City, CA, USA) according to the manufacturer’s protocol. Additional validation for the identified variants ss49839682, rs17032268, rs9811771 and rs7647987 in FANC D2 was carried out by dHPLC analysis (conditions shown in Supplementary Table II) in 95 case and 95 control non-H2 carriers. A sample of 551 case and 559 control non-H2 carriers were further genotyped using TaqMan probes (Supplementary Table I) in order to validate the results obtained for ss49839682 and rs9811771, FANC D2 variants were assessed for functional importance using PupaSNP (http://pupasnp.bioinfo.ohio.fib.es), which synthesizes information from: Ensembl v25.34e, dbSNP 122, HapMap release #12, Match v2.2, Transfac Professional v8.3, Haploview v3.0 and PMUT (55). Transfac Professional v8.3 was also applied to screen for transcription factor consensus binding sites using Rvista v2.0 (http://rvista.dcode.org) (56). Phylogenetic conservation analysis was performed using ECR Browser (http://ecrbrowser.dcode.org) (57). Phylogenetic sequence comparisons and construction of phylogenetic trees among mammal organisms were carried out using Tcoffee (http://fips-server.cns.mrs.fr/ Tcoffee/tcoffe_cgi/index.cgi) (58) (see Figures 2A and B in Supplementary Figure 1).
Results

Genotyping

Genotyping call rates were over 97%, and concordance rates for non-DNA and duplicate samples were 100%, for all SNPs genotyped (data are shown in Table II). All seven SNPs complied with Hardy–Weinberg equilibrium among controls.

Analysis of genotypes

A higher MAF was observed among cases than controls for rs2272125 in FANCD2 (0.13 versus 0.10, respectively, \( P = 0.005 \)). The other six polymorphisms tested did not show any evidence of differences in the allelic frequencies (data not shown). Comparison of genotype frequencies suggested that the risk of developing sporadic breast cancer associated with rs2272125 followed a codominant model, with the estimated OR for rare CC homozygotes relative to common AA homozygotes (OR = 2.77; 95% CI = 1.08–7.13; \( P = 0.03 \)) more than double that for heterozygotes (OR = 1.28; 95% CI = 1.02–1.62; \( P = 0.04 \)). The best fitting model was a single parameter codominant one, with an associated multiplicative per-allele OR of 1.35 (95% CI = 1.09–1.67, \( P = 0.005 \)). This association was consistently observed after adjusting for age (Table I) (OR = 1.32, \( P = 0.03 \)). We did not find any other associations among the remaining SNPs in this study. Results for all genotyped SNPs are summarized in Table II.

The program Structure v.2.0 was applied to a set of 28 unlinked biallelic markers genotyped in a random subset of 163 cases and 180 controls, finding no evidence of population substructure, with an estimated posterior probability of
approximately one for $K$ (number of clusters) = 1, and minimum and maximum ancestry coefficients of 0.45 and 0.55, respectively for $K = 2$, consistent across independent runs (R.L. Milne unpublished data).

**Analysis of haplotypes and diplotypes**

We compared the imputed haplotypes and diplotypes formed by the SNPs in $FANCD2$ between cases and controls and the results are presented in Table III. The most common haplotype (H1) found in both cases and controls carried the more common alleles ‘T’ in rs722509 and ‘A’ in rs2272125. Comparison of haplotype frequencies between cases and controls suggested that one of the four haplotypes was associated with the risk of breast cancer. This haplotype (H2) carried the combination ‘TC’ and the estimated OR relative to H1 was 4.46 (95% CI = 1.51–13.21; $P = 0.007$) assuming the most
confirmed that they were at increased risk (OR $P = 0.007$). Haplotype $H_2$ represented people carrying at least one haplotype $H_2$ ($D_1-2$ and $D_2-4$).

Counts and frequencies in cases and controls with ORs and $P$-values for tests of association with sporadic breast cancer. OR, odds ratio; CI, confidence interval. D2- represent people carrying at least one haplotype H2 ($D_1-2$ and $D_2-4$).

## Table III. Statistical analysis of haplotypes (a) and diplotypes (b) in the FANC2 gene. Bold typeface indicate statistically significant haplotype and diplotype analyses.

<table>
<thead>
<tr>
<th></th>
<th>Controls</th>
<th>Cases</th>
<th>OR (95% CI)</th>
<th>$P$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N$ (%)</td>
<td>$N$ (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a) Haplotype $^a$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$H_1$</td>
<td>1328 (89)</td>
<td>1340 (87)</td>
<td>1.00 (Reference)</td>
<td></td>
</tr>
<tr>
<td>$H_2$</td>
<td>4 (0)</td>
<td>18 (1)</td>
<td>$4.46 (1.51-13.21)$</td>
<td>$0.007$</td>
</tr>
<tr>
<td>$H_3$</td>
<td>4 (0)</td>
<td>3 (0)</td>
<td>$0.74 (0.7-3.33)$</td>
<td>0.70</td>
</tr>
<tr>
<td>$H_4$</td>
<td>150 (10)</td>
<td>185 (12)</td>
<td>$1.22 (0.97-1.54)$</td>
<td>0.09</td>
</tr>
<tr>
<td>(b) diplotype $^b$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$D_1-1$</td>
<td>592 (80)</td>
<td>581 (75)</td>
<td>1.00 (Reference)</td>
<td></td>
</tr>
<tr>
<td>$D_1-2$</td>
<td>4 (1)</td>
<td>14 (2)</td>
<td>$3.57 (1.17-10.90)$</td>
<td>0.03</td>
</tr>
<tr>
<td>$D_1-3$</td>
<td>2 (0)</td>
<td>3 (0)</td>
<td>$1.53 (0.25-9.18)$</td>
<td>0.64</td>
</tr>
<tr>
<td>$D_1-4$</td>
<td>138 (19)</td>
<td>161 (21)</td>
<td>$1.19 (0.92-1.53)$</td>
<td>0.18</td>
</tr>
<tr>
<td>$D_2-4$</td>
<td>0 (0)</td>
<td>4 (1)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$D_3-3$</td>
<td>1 (0)</td>
<td>0 (0)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$D_3-4$</td>
<td>6 (1)</td>
<td>10 (1)</td>
<td>$1.70 (0.61-4.70)$</td>
<td>0.31</td>
</tr>
<tr>
<td>$D_2-4$</td>
<td>4 (1)</td>
<td>18 (2)</td>
<td>$4.59 (1.54-13.63)$</td>
<td>0.006</td>
</tr>
</tbody>
</table>

likely haplotypes were actually observed, and 4.81 (95% CI 1.63–14.18; $P = 0.004$) allowing for haplotype uncertainty [Table III (a)].

Analysis of diplotypes found that carrying H1 and H2 (diploptype $D_1-2$) was associated with increased risk of breast cancer relative to $D_1-1$ (OR 3.57; 95% CI 1.17–10.90; $P = 0.03$). Moreover, only four women carried H2 in a diplotype other than $D_1-2$, all in combination with $H_4$ ($D_2-4$) and all four were cases. Pooling all carriers of H2 ($D_1-2$ and $D_2-4$) confirmed that they were at increased risk (OR $= 4.59; 95\%$ CI 1.54–13.63; $P = 0.006$) [Table III (b)].

Screening for putative causal mutations in FANC2 risk haplotype carriers

rs2272125 is unlikely to directly cause the observed increased breast cancer risk because it is a synonymous coding SNP. Furthermore, we evaluated its potential implication in exon–intron splicing and found no exonic splicing enhancer (ESE) motifs in that region of FANC2. The observed association might be due to a different causal variant in the same LD block as rs2272125, which extends over the whole region including FANC2 (over 100kb, see Figure 1B). Because it was unclear where the causative variation could be located, we searched for putative causal variants in the region among the carriers of the risk haplotype $H_2$ ($D_1-2$ and $D_2-4$). We studied the described functional regions in the FANC2 protein by DNA sequencing: the phosphorylation motif, located in exon 9, and the monoubiquitination motif, located in exon 19 (Figure 1A). We did not detect any variation from the reference sequence used (NC_000004).

We also sought to characterize the region surrounding rs2272125, (from exon 38 to exon 43), as well as the promoter (1500 bp upstream of ATG) and 3′-UTR regions, the latter two as putative functional domains. Three variants were found in the putative promoter region: rs781871 (MAF not described), rs17032268 (NCBI MAF = 0.22) and a novel SNP ss49839682 (−544 bp). A transitional change A > G in intron 38, 13 bp from exon 38 (IV38 +13), was identified in the public databases as rs9811771, with unknown MAF. Finally, four SNPs in the 3′-UTR region were detected, one (rs7647987) located before the polyA sequence (NCBI MAF = 0.12), and three others located after the polyA sequence: rs3826, rs11716842, and a novel polymorphism ss49839681 (+421 bp from the stop codon), all with unknown MAF (Table IV).

All these variants detected in the H2 carriers were in high LD and, according to PHASEv2.0, the variant-haplotype VH2 (CTGGAG) was observed (see areas highlighted in grey in Table IV). Four individuals carried sections of VH2 and a fifth carried the entire haplotype, together representing 30% of the H2 carriers.

We genotyped four of these eight variants in 190 non-carriers of H2 (95 cases and 95 controls) in order to investigate whether any appeared accumulated in H2 carriers. These variants were: ss49839682, rs17032268, rs9811771 and rs7647987, chosen because our hypothesis was focused on detecting variants with functional implications that also shared the main part of VH2 (Table IV). Only two out of the four variants selected (ss49839682 and rs9811771) appeared to be more frequent in H2 carriers than non-carriers although neither was statistically significant (OR $= 2.04$; 95% CI $= 0.73–5.66$; $P = 0.2$; and OR $= 1.68$; 95% CI $= 0.61–4.62$; $P = 0.3$, respectively). We subsequently genotyped a sample of 1100 DNAs for these two SNPs, but did not observe statistical significance, although the tendency of the minor allele accumulating in H2 carriers was also observed (Table V).

Analysis of sequence conservation was performed for these four additional FANC2 variants in several mammal organisms. Only ss49839682 and rs17032268 (and their surrounding regions) have been phylogenetically maintained in mammals. Furthermore, these two variants located in the promoter region contain several consensus motifs for transcriptional factors (Figure 2A and B, respectively). The variant rs9811771, located near the exon–intron splicing site, was only weakly conserved in Pan troglodytes and Canis familiaris and rs7647987, located at the 3′polyA site, was not conserved at all (see Supplementary Figures 1A and B, respectively).

Discussion

Previous studies have demonstrated that several genes involved in DNA repair, other than BRCA1 or BRCA2, are related to breast cancer susceptibility (59–63). BRCA2 has been identified as FANC1 (3), and several interactions between both BRCA1 and BRCA2 and FA proteins have been described (64,45,40,65). It is therefore plausible that alterations in other FA genes might be implicated in breast cancer risk (66). It is known that alterations in FA genes are linked to several types of cancer, included breast cancer, in the general (non-FA) population (67,68). For example, the absence of FANCF by aberrant promoter methylation (69,70), and BACH1/FANCJ mutations (71) have both been found in breast cancer cases. Furthermore, one study found that FANC2 is more often expressed in high proliferation rate tissues, such as mammary glands (72), and mouse model studies have shown that a germline disruption of the murine Fancd2 gene results in ovarian and epithelial breast cancers (73,74). However, despite the evidence in the literature about
In close proximity or in LD with this SNP. The region of high
is therefore that there is a causative functional change located
in any ESE motif. A likely explanation for this association
rs2272125 is of a synonymous nature and it is not located
(estimated greater than 4-fold risk of breast cancer
ation, with carriers of at least one copy of the H2 haplotype at
Both haplotype and diplotype analyses confirmed this associ-
FANCD2

Variants found by dHPLC and sequencing in samples carrying haplotype H2.
Table IV.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Promoter</th>
<th>Intron 38</th>
<th>Exon 43</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rs781871</td>
<td>rs9839682</td>
<td>rs17032268</td>
</tr>
<tr>
<td>Major allele</td>
<td>T</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>Case-1</td>
<td>T/T</td>
<td>T/C</td>
<td>C/T</td>
</tr>
<tr>
<td>Case-2</td>
<td>T/A</td>
<td>T/T</td>
<td>C/C</td>
</tr>
<tr>
<td>Case-3</td>
<td>T/T</td>
<td>T/C</td>
<td>C/T</td>
</tr>
<tr>
<td>Case-4</td>
<td>T/A</td>
<td>T/T</td>
<td>C/C</td>
</tr>
<tr>
<td>Case-5</td>
<td>T/T</td>
<td>T/T</td>
<td>C/C</td>
</tr>
<tr>
<td>Case-6</td>
<td>T/T</td>
<td>T/T</td>
<td>C/C</td>
</tr>
<tr>
<td>Case-7</td>
<td>T/T</td>
<td>T/T</td>
<td>C/C</td>
</tr>
<tr>
<td>Case-8</td>
<td>T/T</td>
<td>T/T</td>
<td>C/C</td>
</tr>
<tr>
<td>Case-9</td>
<td>– – A/A – A/A</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>Case-10</td>
<td>– – A/A – A/A</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>Case-11</td>
<td>– – A/A – A/A</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>Case-12</td>
<td>T/A</td>
<td>T/T</td>
<td>C/C</td>
</tr>
<tr>
<td>Case-13</td>
<td>T/T</td>
<td>T/C</td>
<td>C/T</td>
</tr>
<tr>
<td>Case-14</td>
<td>T/T</td>
<td>T/T</td>
<td>C/C</td>
</tr>
<tr>
<td>Control-1</td>
<td>T/T</td>
<td>T/C</td>
<td>C/C</td>
</tr>
<tr>
<td>Control-2</td>
<td>T/T</td>
<td>T/C</td>
<td>C/C</td>
</tr>
<tr>
<td>Control-3</td>
<td>T/T</td>
<td>T/C</td>
<td>C/C</td>
</tr>
</tbody>
</table>

−, unknown genotype, not amplified.

Samples sharing sections of the same haplotype (VH2) highlighted in grey.

Variants between the dashed lines were selected for further stud

Table V. Observed allele frequencies for four putative functional SNPs, by H2 carrier status

<table>
<thead>
<tr>
<th>Variant</th>
<th>H2 carriers</th>
<th>190 H2 non-carriers</th>
<th>1100 H2 non-carriers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N_m/Na</td>
<td>MAF</td>
<td>N_m/Na</td>
</tr>
<tr>
<td>ss49839682</td>
<td>5/30 0.17*</td>
<td>34/380 0.09*</td>
<td>248/2166 0.11*</td>
</tr>
<tr>
<td>rs17032268</td>
<td>4/26 0.13</td>
<td>35/346 0.10</td>
<td>–</td>
</tr>
<tr>
<td>rs9811771</td>
<td>5/34 0.15**</td>
<td>35/376 0.09**</td>
<td>265/2154 0.12**</td>
</tr>
<tr>
<td>rs77647987</td>
<td>6/28 0.18</td>
<td>45/298 0.18</td>
<td>–</td>
</tr>
</tbody>
</table>

N_m, observed number of minor alleles; Na, total number of alleles;
MAF, minor allele frequency.

*P (comparison of MAFs) = 0.2 and 0.4, using 190 and 1083 non-H2
carrier samples, respectively.

**P (comparison of MAFs) = 0.3 and 0.7, using 188 or 1077 non-H2
carrier samples, respectively.

the potential susceptibility to breast and other cancers associ-
ated with the FA family (75), there have been few studies
investigating these associations (76).

In this study, we have investigated associations between
genetic variants in three genes belonging to the FA family
and sporadic breast cancer in the Spanish population. We
genotyped seven SNPs in FANCA, FANCL and FANCD2.
Whereas no associations with breast cancer risk were found
for SNPs in FANCA and FANCL, we were able to identify an
increased risk associated with the minor ‘C’ allele of
rs2272125 in FANCD2 (OR = 1.35 per allele) (Table II).
Both haplotype and diplotype analyses confirmed this associ-
ation, with carriers of at least one copy of the H2 haplotype at
an estimated greater than 4-fold risk of breast cancer
(P = 0.007 and P = 0.006, respectively) (Table III).

Despite being located in a coding region, the variant
rs2272125 is of a synonymous nature and it is not located
in any ESE motif. A likely explanation for this association
is therefore that there is a causative functional change located
in close proximity or in LD with this SNP. The region of high
LD that includes rs2272125 is over 100 Kb and spans the entire
FANCD2 gene. Applying two methods of LD inference, the
Confidence Intervals Method (77) and the 4 Gametes Rule
(78), it was found that this LD block contains FANCD2 and
three other genes of unknown function. VHL, a gene close to
FANCD2 which has been implicated in some cancers (79,80),
was excluded because it was not included in the LD block
according to the latter method (Figure 1B). We therefore
focused on FANCD2, our hypothesis being that this gene
could carry the putative causal variant.

In-depth analysis of several regions of the FANCD2 gene
among H2 carriers allowed us to detect another eight variants,
some of them described previously in the SNP databases, but
others identified for the first time (ss49839682 and ss49839681)
(Table IV). Four of these variants could have functional
implications because of their location (promoter, exon–intron
splicing site, and 3’-UTR regions). In addition, they seemed to
form a variant-haplotype (VH2) within H2 carriers (Table IV).
We therefore investigated how frequent these variants were in
people not carrying the H2 haplotype. One of the two variants
located in the likely promoter region fell in a strongly phylo-
genetically conserved motif with consensus binding sites for
several transcriptional factors (Figure 2A), which made us
speculate that it may have functional implications; however,
we have been unable to validate statistically this hypothesis
(Table V). It is also possible that other variants along
FANCD2 or the immediate surrounding region could be
responsible for the increased risk associated with this gene.

Another, less likely possibility, is that the observed associ-
ation could be due to confounding by unmeasured factors, such
as race. Although population stratification was tested for and
not detected, this we done using only a subset of the sample
and using only 28 markers. Population stratification cannot
therefore be entirely discarded, although it appears unlikely
to have influenced the results of this study. Alternatively, the
observed association could be due to chance. Replication of
these findings using samples from other populations is required
to address these two concerns.
In conclusion, we found that a synonymous variant (rs2272125) and haplotype (H2) in the FANCD2 gene are associated with sporadic breast cancer. This is the first study to identify this association. We have also identified a putative functional variant (ss49839682) in the same high LD block, but were not able to conclude that this is causal. Further studies will be required to clarify the full implication of the FANCD2 protein in sporadic breast cancer risk.

Supplementary material
Supplementary material is available online at http://www.carcin.oupjournals.org.

Acknowledgments
This study was supported by the grant BF2003-03852. EB and LPF were funded by the Comunidad Autonoma de Madrid and by the Spanish Ministry of Science and Technology (MCT), respectively. We would like to thank Alvaro Ruibal (University Hospital, Santiago de Compostela), and Santiago Palacios (Instituto Palacios, Madrid) for the use of samples of cases and controls. We would also like to thank Raquel Rodriguez-Lopez for her contribution to the beginning of this work, and Fatima Mercadillo, Alicia Barroso, Emilia Gonzalez, Jesus Lopez, Victoria Fernandez and Rocio Leton for their expert technical skills.

Conflict of Interest Statement
None declared.

References
5. Risch,N. (2001) The genetic epidemiology of cancer: interpreting family studies will be required to clarify the full implication of the putative functional variant (ss49839682) in the same high LD study to identify this association. We have also identified a associated with sporadic breast cancer. This is the first study to identify this association. We have also identified a putative functional variant (ss49839682) in the same high LD block, but were not able to conclude that this is causal. Further studies will be required to clarify the full implication of the FANCD2 protein in sporadic breast cancer risk.