Materials and methods

Study subjects, data collection and DNA extraction

The study included a total of 547 consecutive and non-related sporadic BC cases and 552 control women. Cases were recruited from 1st January 2002 to 31st December 2006 from three Spanish public hospitals: 256 (47%) from Monte Naranco Hospital, in Oviedo; 164 (30%) from the Fundación Jiménez Díaz, and 127 (23%) from La Paz University Hospital, both in Madrid. Controls were unaffected Spanish women, recruited at two centres in Madrid: 460 (83%) from the Menopause Research Centre at the Instituto Palacios, and 92 (17%) from the Fundación Jiménez Diaz. All cases and controls were women and controls were selected so that their age range was comparable to that of cases.

Information about personal characteristics of cases and controls (age at diagnosis for cases or age at blood sample collection for controls, age at menarche, parity, menopausal status, and survival), and clinical and tumour characteristics for cases (nodal metastasis, and ER and PR immunohistochemical markers), were either collected by the treating physician or extracted by review of medical records [1].

Gene and SNP selection

A total of ten FANC genes and five genes related to Fanconi anemia pathway were selected.

Three public databases were used to collect information about SNPs in FANC genes and related genes: 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 10%, the exception being putative coding SNPs with MAF ≥ 5%. SNPs with a putative function in the gene expression regulation due to their location into gene have been considered:
SNPs in coding regions, promoter, 3'UTR, exon-intron boundaries, ISEs and ESEs. The following bioinformatics tools were used in order to get SNP functional information: PupaSuite [2], the Functional Element SNP Database (http://combio.kribb.re.kr/FESD/), rVista v2.0 (http://rvista.dcode.org), Pfam (http://www.sanger.ac.uk/Software/Pfam/) and EMBL-EBI (http://www.ebi.ac.uk/). Information on linkage disequilibrium (LD), or blocks within chromosomes that are more probably linked and segregate together than expected at random, was used to select representative SNPs or tag-SNPs (a subset of SNPs that represent variation in a gene or chromosomal region) using Haploview v3.11 when possible. Those SNPs conserved in several mammalian species were selected according ECR Browser (http://ecrbrowser.dcode.org). A total of 75 SNPs were selected following the described criteria.

**Genotyping assays**

Genotyping was carried out using the TaqMan, iPLEX, Amplifluor and Illumina GoldenGate platforms following the manufacturer’s instructions. Amplifluor genotyping platform was used only in case of failed TaqMan and iPLEX assay. TaqMan genotyping assays were designed using Applied Biosystems Assay-by-Design and Assay-on-Demand probes (Applied Biosystems, Foster City, CA, USA). Amplifluor oligonucleotides were designed using Amplifluor AssayArchitect Software. The genotype of each sample using TaqMan and Amplifluor was automatically determined by measuring final allele-specific fluorescence in the ABI Prism 7900HT Detection System, using the SDS 2.1 software for allele discrimination (Applied Biosystems, Foster city, USA). iPLEX oligonucleotides were designed following the manufacture’s instructions and were analysed using the MassARRAY Workstation software v3.3. SNPs genotyped by Illumina GoldenGate platform were included within an OPA (custom oligonucleotide pool) of 768 SNPs in cancer genes. Genotyping designs for 33 SNPs genotyped using TaqMan, 13 SNPs by iPLEX, 2 SNPs by Amplifluor and 28
SNPs by Illumina Golden Gate (one of them also genotyped using TaqMan), can be given upon request.

As a quality control measure, we included at least 2 sample duplicates and 1 non-template sample per 96-well plate. Genotypes were scored by two different personnel in the laboratory. We obtained a concordance rate of 100% for all 72 successful SNP genotyping assays studied. Three genotyping assays failed.

**Statistical analyses**

For all polymorphisms studied, Fisher’s exact test was used both to test for deviations from Hardy-Weinberg equilibrium (HWE) among controls and to compare differences in the MAF distributions between cases and controls. We found evidence of departure from HWE for eleven of the 72 successfully genotyped SNPs, removed from the following statistical analyses. A total of 61 SNPs were included in the statistical analysis.

In order to assess associations between genotypes, haplotypes and cancer risk, several analyses were performed. Genotype-related odds ratios (ORs), their corresponding 95% confidence intervals (CIs) and associated p-values were estimated via unconditional logistic regression. This was done for each of heterozygotes and minor-allele homozygotes relative to common-allele homozygotes (data do not shown), as well as under an additive model, in the latter case estimating an effect per copy of the minor allele carried. Known or suspected risk factors for BC (age, number of live births, age at menarche, and menopause status) were evaluated for potential confounding effects by including them in multivariate analyses (data not shown). Survival association according genotype for each SNP was analysed using Kaplan-Meier and Cox regression assays.

Associations between polymorphisms genotyped and various clinical and tumour characteristics were assessed via logistic regression in order to determine their potential modifying effects on BC risk. This was done for cases only: nodal metastasis
(yes versus no), estrogen receptor status (positive versus negative) and progesterone receptor status (positive versus negative), were considered as the outcome variables. In the case of association, assessment of real phenotypic association was performed by stratification of case samples and comparison to complete set of controls, via logistic regression of subset of samples.

SPSS v11.0 was used to carry out these analyses. All p-values were two-sided.
