Familial breast cancer and DNA repair genes: Insights into known and novel susceptibility genes from the GENESIS study, and implications for multigene panel testing


To cite this version:


HAL Id: inserm-02438452
https://www.hal.inserm.fr/inserm-02438452
Submitted on 14 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Familial breast cancer and DNA repair genes: Insights into known and novel susceptibility genes from the GENESIS study, and implications for multigene panel testing

Elodie Girard1,2,3,4, Séverine Eon-Marchais1,2,3,4, Robert Olaso5, Anne-Laure Renault1,2,3,4, Francesca Damiola6, Marie-Gabrielle Dondon1,2,3,4, Laure Barjhoux5, Didier Goidin7, Vincent Meyer9, Dorothée Le Gal1,2,3,4, Juana Beauvallet1,2,3,4, Noura Mebirouk1,2,3,4, Christine Lonjou1,2,3,4, Juliette Coignard1,2,3,4,8, Morgane Marcou1,2,3,4, Eve Cavaciuti1,2,3,4, Céline Baulard9, Marie-Thérèse Bihoreau5, Odile Cohen-Haguenauer9, Dominique Leroux10, Clotilde Penet11, Sandra Fert-Ferrer12, Chryselle Colas13,14, Thierry Frebourg15, François Eisinger16, Claude Adenis17, Anne Fajac18, Laurence Gladieff19, Julie Tinat11, Anne Floquet20, Jean Chiesa21, Sophie Giraud22, Isabelle Mortemousque23, Florent Soubrier24, Séverine Audebert-Bellanger25, Jean-Marc Limacher26, Christine Lasset27, Sophie Lejeune-Dumoulin28, Hélène Dreyfus29, Yves-Jean Bignon30, Michel Longy20, Pascal Pujol31, Laurence Venat-Bouvet32, Valérie Bonadonna27, Pascaleine Berthe33, Elisabeth Luporsi34, Christine M. Maugard35, Catherine Noguès16, Capucine Delnatte36, Jean-Pierre Fricker37, Paul Gesta38, Laurence Faivre39, Alain Lortholary40, Bruno Buecher14, Olivier Caron41, Marion Gauthier-Villars14, Isabelle Coupier31, Nicolas Servant1,2,3,4, Anne Boland5, Sylvie Mazoyer42, Jean-François Deleuze5, Dominique Stoppa-Lyonnet14,43,44, Nadine Andrieu1,2,3,4 and Fabienne Lesueur5,1,2,3,4

1Inserm, Paris, France
2Institut Curie, Paris, France
3Mines ParisTech, Fontainebleau, France
4PSL Research University, Paris, France
5Centre National de Recherche en Génomique Humaine, Institut de Biologie François Jacob, CEA, Evry, France
6Département de Biopathologie, Centre Léon Bérard, Lyon, France
7Life Sciences and Diagnostics Group, Agilent Technologies France, Les Ulis, France
8Université Paris Sud, Paris, France
9Hôpital Saint-Louis, Paris, France
10Département de Génétique, CHU de Grenoble, Hôpital Couple-Enfant, Grenoble, France
11Consultation d’Oncogénétique, Institut Jean-Godinot & ICC Courlancy, Reims, France
12Service de Génétique, Centre Hospitalier de Chambéry, Chambéry, France
13Département de Génétique Groupe Hospitalier Pitié-Salpêtrière, APHP, Paris, France
14Service de Génétique, Institut Curie, Paris, France
15Département de Génétique, Hôpital Universitaire de Rouen, Rouen, France
16Institut Paoli Calmettes, Département d’Anticipation et de Suivi des Cancers, Oncogénétique Clinique, Institut Paoli-Calmettes & Aix Marseille Université, Inserm, IRD, SESSTIM, Marseille, France

Key words: breast cancer, exome sequencing, multigene panel testing, variant, case–control study

Abbreviations: 95%CI: 95% confidence interval; BC: breast cancer; ER: estrogen receptor; HBOC: hereditary breast and ovarian cancer; IGV: integrative genomics viewer; LoF: loss-of-function variant; MAF: minor allele frequency; MV: missense variant; OR: odds ratio; P: probability value; PR: progesterone receptor; SNV: single nucleotide variant; VUS: variant of uncertain clinical significance; WES: whole-exome sequencing.

Additional Supporting Information may be found in the online version of this article.

Conflict of interest: All authors, except for DG, FE and OC declare no conflict of interest. DG is employed at Agilent Technologies Inc. FE is a consultant for ROCHE S.A., Genome Quebec and the International Biobank of Luxembourg. OC received honoraria for AstraZeneca and Ipsen.

Grant sponsor: comprehensive cancer center SiRIC; Grant numbers: INCa-DGOS-4654; Grant sponsor: Fondation ARC pour la recherche sur le cancer; Grant numbers: PJA 20151203365; Grant sponsor: France Génomique National infrastructure; Grant numbers: ANR-10-INBS-09; Grant sponsor: Institut National du Cancer (INCa); Grant numbers: b2008-029/LL-LC; Grant sponsor: Ligue Comité de Paris ; Grant numbers: RS15/75-78; Grant sponsor: Ligue Nationale Contre le Cancer; Grant numbers: PREDOS/DGL PRE07/DGL PRE11/NA

DOI: 10.1002/ijc.31921

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

History: Received 22 Apr 2018; Accepted 25 Sep 2018; Online 10 Oct 2018.
Correspondence to: Fabienne Lesueur, Inserm, U900, Institut Curie, Paris, France, E-mail: fabienne.lesueur@curie.fr
Pathogenic variants in **BRCA1** and **BRCA2** only explain the underlying genetic cause of about 10% of hereditary breast and ovarian cancer families. Because of cost-effectiveness, multigene panel testing is often performed even if the clinical utility of testing most of the genes remains questionable. The purpose of our study was to assess the contribution of rare, deleterious-predicted variants in DNA repair genes and specific breast cancer relative risks associated with rare deleterious-predicted variants in **BRCA1** and **BRCA2** mutation and having a sister with BC (OR\textsubscript{het} = 1.6; \textit{p}_{\text{het}} = 0.002). Hence, our approach allowed us to specify BC relative risks associated with deleterious-predicted variants in **PALB2**, **ATM**, and **CHEK2** and to add **FANCI**, **MAML1**, **POLH**, and **RTF1** to the list of DNA repair genes possibly involved in BC susceptibility. We also highlight that different types of variants within the same gene can lead to different risk estimates.

**What’s new?**

Pathogenic variants in **BRCA1** and **BRCA2** only explain the genetic cause of about 10% of hereditary breast and ovarian cancer families, and the clinical usefulness of testing other genes following the recent introduction of cost-effective multigene panel sequencing in diagnostics laboratories remains questionable. This large case-control study describes genetic variation in 113 DNA repair genes and specifies breast cancer relative risks associated with rare deleterious-predicted variants in **PALB2**, **ATM**, and **CHEK2**. Importantly, different types of variants within the same gene can lead to different risk estimates. The results may help improve risk prediction models and define gene-specific consensus management guidelines.
Introduction
Women from families with multiple cases of breast or ovarian cancer have a higher risk of developing breast cancer (BC) during their lifetime than women in the general population. Since the identification of BRCA1 and BRCA2 in the late 1990s, genetic testing has become a routine clinical assessment for individuals with clinical features suggestive of a hereditary breast and ovarian cancer (HBOC) predisposition. However, a pathogenic variant in one of these two genes is found in only 10% of patients attending the cancer genetics clinics. HBOC families with no pathogenic variants in the known predisposition genes present challenges for molecular genetic diagnostics. In such families, predictive testing for healthy relatives, genetic counseling and preventive medical management are hampered.

Paradoxically in recent years, the introduction of cost-effective multigene panel sequencing in diagnostics laboratories has led to an explosion of genetic data in cancer-prone families. Very often, 20–25 genes or more are being screened in parallel to BRCA1 and BRCA2 in HBOC families, but data attributing a causative role for potentially deleterious variants identified in many of these genes are usually lacking, and precise risk estimates per variant category (e.g. loss-of-function variants (LoF) or likely deleterious missense variants (MV)) have not been determined yet.

To date, the screened genes include known cancer susceptibility genes associated with a moderate (2- to 5-fold) to high relative risk of BC (>5-fold) because of their direct or indirect functional link with BRCA1 and BRCA2, such as PALB2, ATM and CHEK2. Other DNA repair genes, such as RAD51 paralogs, RINT1 and genes of the MRE11A-RAD50-NBN complex are also screened although replication studies conducted in independent at-risk populations are lacking or failed to replicate initial findings. The interpretation of the genetic data is also complicated. Most of the known pathogenic variants associated with BC are LoF (i.e. frameshift, stop-gain and canonical splice variants predicted to result in a truncated protein) but a significant fraction of BC risk may be attributable to rare MV in some genes such as ATM, CHEK2 and TP53, and this fraction may be even bigger than that attributable to LoF.

In the light of these issues, whether identifying a LoF or a MV in genes usually screened in parallel to BRCA1 and BRCA2 is clinically useful remains questionable. We therefore conducted a study to estimate BC relative risks associated with rare potentially pathogenic variants in DNA repair genes in a large well-characterized and homogeneous population.

Participants, Material and Methods
Study participants
The study population consisted of women participating in GENESIS (GENE SISTers), a French resource for familial BC research. In summary, 1,721 women affected with infiltrating Table 1. Distribution of cases and controls by age, by population ancestry/ethnicity and by family history of cancer

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Controls (N = 1,199)</th>
<th>All cases (N = 1,207)</th>
<th>Pseudo-incident cases (N = 663)</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>9</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>31–40</td>
<td>61</td>
<td>20</td>
<td>13</td>
</tr>
<tr>
<td>41–50</td>
<td>265</td>
<td>269</td>
<td>171</td>
</tr>
<tr>
<td>51–60</td>
<td>429</td>
<td>465</td>
<td>250</td>
</tr>
<tr>
<td>61</td>
<td>435</td>
<td>453</td>
<td>229</td>
</tr>
<tr>
<td>Mean age (SE)</td>
<td>56.1 (0.3)</td>
<td>57.5 (0.3)</td>
<td>56.6 (0.4)</td>
</tr>
<tr>
<td>Range</td>
<td>25–83</td>
<td>31–90</td>
<td>31–85</td>
</tr>
</tbody>
</table>

Population ancestry/ethnicity

| Caucasian | 1,176 | 1,152 | 631 |
| Ashkenazi | 0     | 4     | 4   |
| African/Afro-Caribbean | 1 | 3 | 1 |
| Arab/Berber/Asian/Arab +Berber | 4 | 13 | 6 |
| Mixed origin | 9 | 18 | 7 |
| Unknown | 9 | 17 | 14 |

Number of cancers in family

Breast cancer (C50 + D05.1)

| 0 | 851 | 0 | 0 |
| 1 | 261 | 0 | 0 |
| 2 | 68  | 294 | 167 |
| 3 | 14  | 467 | 252 |
| 4 | 3   | 280 | 161 |
| 5 | 1   | 104 | 46 |
| 6 | 0   | 46  | 28 |
| 7 | 1   | 11  | 6  |
| 8 | 0   | 5   | 3  |

Ovarian cancer (C56)

| 0 | 1,164 | 1,139 | 623 |
| 1 | 33    | 61    | 34  |
| 2 | 1     | 6     | 6   |
| 3 | 0     | 1     | 0   |
| 5 | 1     | 0     | 0   |

Cancers from Lynch syndrome spectrum

| 0 | 781    | 678 | 396 |
| 1 | 306    | 335 | 158 |
| 2 | 81     | 130 | 74  |
| 3 | 20     | 44  | 23  |
| 4 | 5      | 14  | 8   |
| 5 | 5      | 5    | 4   |
| 6 | 1      | 0    | 0   |
| 7 | 0      | 1    | 0   |

1Cases diagnosed with BC less than 5 years before enrollment in GENESIS.
2Age at diagnosis for cases and age at inclusion for controls.
3Number after excluding GENESIS index case and one affected sister.
4Colon (C18), small bowel (C17), rectum (C19, C20), endometrium (C53, C54, C55), ovary (C56), stomach (C16), bile duct cancers (C24).
mammary or ductal adenocarcinoma, not carrying a pathogenic variant in *BRCA1* and *BRCA2*, and having a sister with BC were enrolled in GENESIS between 2007 and 2013 through the national network of cancer genetics clinics (http://www.unicancer.fr/en/unicancer-group). Affected sisters (*N* = 826) and 1,419 unrelated cancer-free friends or colleagues of index cases (controls) were also included. These latter were aged-matched (+/-3 years) to cases at interview. Blood samples, clinical and epidemiological data were collected for each participant. Written informed consent for our study was obtained from all participants. GENESIS study protocol and genetic subsequent studies were approved by the appropriate ethics committee (CCP Ile-de-France III) and by the French data protection authority (CNIL).

**Selection of cases for the gene discovery stage.** Women included in the WES stage consisted of 100 unrelated cases presenting with a strong personal or family history of BC. Among those, 53 cases belonged to a family with at least one case having developed bilateral BC, 43 cases had at least three sisters affected with BC, and 11 cases had a male first-degree relative affected with BC. Seven of the selected cases fulfilled two of these criteria. The mean age at diagnosis of first BC was 48.4 (range: 28–71 years). Ninety-eight of the cases were of European origin.

**Selection of subjects for BC risk assessment.** After quality control procedures, we analyzed DNA from 1,207 unrelated

---

**Figure 1. Strategy for genes prioritization.** The DNA repair genes list was compiled from ACSN maps (https://acsn.curie.fr/), KEGG (http://www.genome.jp/kegg/pathway.html) and after review of literature.

**CASE-CONTROL VARIANT SCREENING**

Targeted sequencing in 1,207 cases and 1,199 unrelated controls
cases (none of whom had been included or were related to an individual who was included in the discovery stage), and 1,199 unrelated controls. The screened sample represented 77% of the GENESIS population. Distribution of cases and controls by age, ethnicity and number of cancers in the family is shown in Table 1.

**DNA preparation**
Peripheral blood samples (10 mL) were collected in the presence of anticoagulant (sodium citrate), and genomic DNA (gDNA) extracted from peripheral blood leukocytes with a standard inorganic method using the DNA extractor Autopure-LS (Qiagen). DNA handling (normalization and aliquoting) was done using a TECAN EVO instrument.

**Sequencing, data processing and variant discovery**
Library preparation, whole exome sequencing (WES) and gene panel sequencing were performed at the Centre National de Recherche en Génomique Humaine (Evry, France). Exomes were captured using the SureSelectXT Human All Exon V5 kit, and coding exons of candidate genes were captured using a custom SureSelect target Enrichment system (Agilent Technologies, Santa Clara, CA, USA). Final libraries were sequenced on an Illumina HiSeq 2500 instrument with 100-bp

---

**Figure 2.** Result of the association tests per variant type for the 113 genes. Legend: *, no LoF identified in cases and/or in controls; **, no LoF and no likely deleterious MV identified in cases and/or in controls.
paired end reads. Reads were mapped to the reference genome hg19 (1,000 Genomes GRCh37 build) using BWA aln v0.6.24 with the after parameters: two mismatches allowed in the 32 bp-seed alignment and four mismatches allowed in the whole read, one alignment reported per read. For the exome-sequenced samples, Picard Tools v.1.103 (http://picard.sourceforge.net) was used to remove duplicate reads in order to mitigate biases introduced by data generation steps such as PCR amplification. SNV and insertions/deletions (indels) were then called using three strategies: i. Freebayes (v0.9.9.9.2),\textsuperscript{25} ii. GATK’ Unified Genotyper (v2.2)\textsuperscript{26} on original alignment files and iii. on preprocessed alignment files using GATK modules for local realignment around indels and base quality score recalibration using databases of known polymorphisms, after best practice recommendations. All variant calls were applied using a minimum base quality of 17 and a minimum mapping quality of 10. A dedicated filtering process was applied per sample to prioritize variants of interest: variants covered by less than 5× and with phred quality of less than 30 were discarded, while only heterozygous variants (defined by an allelic ratio between 20 and 80%) and variants detected by at least two calling strategies were considered. For some complex events (indels or SNVs falling in repeated sequences) we also performed visual inspection of aligned reads using the Integrative Genomics Viewer (IGV)\textsuperscript{27,28} to reduce the risk of false positives and help characterizing them.
The joint variant calling file was finally annotated with refGene regions, SNV effects, as well as 1,000 Genomes (1000G) minor allele frequency (MAF; August 2015 version) using Annovar (October 2015 version, http://annovar.openbioinformatics.org). Only exonic or splicing variants presenting a MAF below 0.5% in all represented ethnic groups of the 1000G database were kept in further analyses. In the gene discovery stage, variants carried by more than five cases and absent from 1000G were excluded as they were likely to represent a sequencing artifact or an unreported common polymorphism in our population. We then only considered LoF, inframe indels and conserved MV (PhyloP score > 0) predicted as deleterious by SIFT and/or PolyPhen2 in the analyses. In the case–control study, CADD (for all genes) and Align-GVGD (for ATM, CHEK2 and PALB2) were also used to annotate and filter MV. In this latter stage, we applied a supplementary filter on MAF in order to exclude any variant with MAF > 0.5% in controls, as we observed that the MAF of some variants was underestimated in the public databases.

Statistical analyses

All LoF and MV with a phred CADD score equal or above 20 were considered in the analyses. Their frequencies among cases and controls were compared by unconditional logistic regressions, adjusting for ethnicity and age at inclusion in GENESIS. To address issues of multiple testing, by examining 113 genes and applying a Bonferroni correction, statistical significance was defined as \( p < 0.0004 \).

We used Likelihood ratio tests to test for heterogeneity according to variant type (LoF vs. MV), and polytomous logistic regressions to test for heterogeneity according to hormone status and according to time between inclusion and diagnosis.

All statistical tests were two-sided and data was analyzed using STATA, version 11.2 (StataCorp, College Station, TX).

Results

Gene discovery stage

WES was performed on 100 unrelated high-risk women to identify candidate genes. The average read depth achieved for target regions was 110X (range: 55X–207X). At least 80% (average: 96.45%; range: 79.77%–99.35%) of the capture target regions were covered by 20 or more sequenced reads for all samples (Supporting Information Fig. S1). Whole exome sequencing performance and variants count for each sequenced DNA sample is shown in Supporting Information Table S1. After initial single nucleotide variants (SNV) filtering steps, an average of 18.2 (11–29) LoF, 6.8 (1–14) inframe indels and 133.7 (105–170) MV were identified per individual. In total, 2,584 genes were altered by such variants in at least two unrelated cases. Among those, 123 genes were known to be involved in DNA repair pathways. After having reviewed the potential involvement of each of them in cancer etiology or progression, we selected 77 genes for further investigation in the case–control study. This gene panel was completed with...
Table 3. Analyses by hormone receptor status and restricted to pseudo-incident cases, for the 7 genes associated with BC (P ≤ 0.05)

<table>
<thead>
<tr>
<th>Gene</th>
<th>Analysis^</th>
<th>Any variant</th>
<th>Loe</th>
<th>Likely deleterious MV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Case</td>
<td>OR^2 (95%CI)</td>
<td>p-Value</td>
<td>Case</td>
</tr>
<tr>
<td>ATM</td>
<td>ER positive</td>
<td>47</td>
<td>2.1 (1.4, 3.3)</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>ER negative</td>
<td>8</td>
<td>2.1 (0.9, 4.5)</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>PR positive</td>
<td>38</td>
<td>2.1 (1.3, 3.4)</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>PR negative</td>
<td>12</td>
<td>1.7 (0.9, 3.4)</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>HER2 positive</td>
<td>3</td>
<td>1.3 (0.4, 4.2)</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>HER2 negative</td>
<td>24</td>
<td>1.8 (1.1, 3.1)</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>Pseudo-incident cases</td>
<td>41</td>
<td>1.9 (1.2, 2.9)</td>
<td>0.006</td>
</tr>
<tr>
<td>CHEK2</td>
<td>ER positive</td>
<td>39</td>
<td>3.5 (2.1, 6.0)</td>
<td>0.000004</td>
</tr>
<tr>
<td></td>
<td>ER negative</td>
<td>3</td>
<td>1.3 (0.4, 4.5)</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td>PR positive</td>
<td>35</td>
<td>3.9 (2.2, 6.7)</td>
<td>0.000001</td>
</tr>
<tr>
<td></td>
<td>PR negative</td>
<td>6</td>
<td>1.6 (0.7, 4.1)</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>HER2 positive</td>
<td>5</td>
<td>3.8 (1.4, 10.4)</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>HER2 negative</td>
<td>27</td>
<td>4.1 (2.3, 7.3)</td>
<td>0.000002</td>
</tr>
<tr>
<td></td>
<td>Pseudo-incident cases</td>
<td>37</td>
<td>3.3 (1.9, 5.6)</td>
<td>0.00001</td>
</tr>
<tr>
<td>FANCI</td>
<td>ER positive</td>
<td>9</td>
<td>0.7 (0.3, 1.6)</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>ER negative</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>PR positive</td>
<td>8</td>
<td>0.9 (0.4, 1.9)</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>PR negative</td>
<td>1</td>
<td>0.2 (0.0, 0.8)</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>HER2 positive</td>
<td>1</td>
<td>0.7 (0.1, 5.4)</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>HER2 negative</td>
<td>3</td>
<td>0.4 (0.1, 1.4)</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>Pseudo-incident cases</td>
<td>6</td>
<td>0.5 (0.2, 1.2)</td>
<td>0.11</td>
</tr>
<tr>
<td>MAST1</td>
<td>ER positive</td>
<td>8</td>
<td>1.8 (0.7, 5.0)</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>ER negative</td>
<td>3</td>
<td>3.8 (1.0, 14.4)</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>PR positive</td>
<td>5</td>
<td>1.2 (0.4, 3.9)</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>PR negative</td>
<td>4</td>
<td>3.0 (0.9, 10.0)</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>HER2 positive</td>
<td>2</td>
<td>4.2 (0.9, 20.0)</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>HER2 negative</td>
<td>2</td>
<td>0.8 (0.2, 3.7)</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>Pseudo-incident cases</td>
<td>8</td>
<td>1.9 (0.7, 5.0)</td>
<td>0.20</td>
</tr>
<tr>
<td>PALB2</td>
<td>ER positive</td>
<td>21</td>
<td>4.6 (2.1, 10.1)</td>
<td>0.0002</td>
</tr>
<tr>
<td></td>
<td>ER negative</td>
<td>1</td>
<td>1.1 (0.1, 8.8)</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>PR positive</td>
<td>14</td>
<td>3.7 (1.6, 8.7)</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>PR negative</td>
<td>7</td>
<td>4.9 (1.8, 13.4)</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>HER2 positive</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>HER2 negative</td>
<td>13</td>
<td>4.7 (2.0, 11.1)</td>
<td>0.0004</td>
</tr>
<tr>
<td></td>
<td>Pseudo-incident cases</td>
<td>15</td>
<td>3.2 (1.4, 7.3)</td>
<td>0.007</td>
</tr>
<tr>
<td>POLH</td>
<td>ER positive</td>
<td>4</td>
<td>0.6 (0.2, 1.8)</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>ER negative</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>PR positive</td>
<td>4</td>
<td>0.7 (0.2, 2.2)</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>PR negative</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>HER2 positive</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

(Continues)
36 genes currently included in commercial BC multigene panels. The workflow for gene prioritization and the final list of analyzed genes are shown on Figure 1.

Breast cancer risk assessment

Targeted sequencing of the 113 genes identified 3,930 distinct variants with MAF lower than 0.5% in GENESIS controls. These included 264 LoF variants (155 in cases and 132 in controls) and 1,994 MV bioinformatically predicted to alter protein function or stability (1,213 in cases and 1,103 in controls). The distribution of the number of altered genes per subject, according to their case/control status is shown on Supporting Information Figure S2. There was no overall difference in the distribution of number of altered genes between cases and controls (Chi² (6df ) = 10.0, p = 0.12) and only a modest increased risk of BC for subjects with at least two altered genes as compared to subjects with 0 or 1 altered gene (OR = 1.2; 95%CI, 1.0–1.4; p = 0.02). Nevertheless, we performed sensitivity analyses adjusted on the number of altered genes in the subsequent analyses. For each gene, we aggregated all rare LoF and MV, and also examined separately the two types of variants. The distribution of variants in cases and controls, per gene, and associated risk estimates are provided in Supporting Information Table S2, and results for all genes are summarized in Figure 2. Among those, PALB2, ATM, CHEK2, FANCI, MAST1, POLH and RTEL1 showed significant association with BC although only CHEK2 remained associated with BC after correcting for multiple testing.

Both LoF and MV in PALB2, ATM and CHEK2 were associated with an increased BC risk, but carriers of an ATM LoF had a significantly higher risk of developing BC than carriers of an ATM MV (OR_{LoF} = 17.4 vs. OR_{MV} = 1.6; p_{Het} = 0.002) (Table 2). In contrast for CHEK2 and PALB2 there was no significant heterogeneity in risk according to variant type (CHEK2: OR_{LoF} + MV = 3.0; 95% CI 1.9–5.0; PALB2: OR_{LoF} + MV = 3.5; 95% CI 1.7–7.5).

The increased risk associated with variants in MAST1 was essentially driven by carrying a MV (OR_{MV} = 2.5; 95%CI 1.0–6.0). Likewise, the decreased risk associated with variants in FANCI, POLH and RTEL1 was driven by carrying a MV (FANCI: OR_{MV} = 0.4; 95%CI 0.2–1.0; POLH: OR_{MV} = 0.3; 95%CI 0.1–1.0; RTEL1: OR_{MV} = 0.4; 95%CI 0.2–0.9). Intriguingly, four cases and one control carried a FANCI LoF, leading to a nonsignificant increase in BC risk (OR_{LoF} = 3.7 vs. OR_{MV} = 0.4; p_{Het} = 0.04) (Table 2).

Furthermore, we found that carriers of a LoF in CHEK2 were diagnosed with BC at a younger age than noncarriers of a CHEK2 variant (46.7 years vs. 52.1 years, p_{T test} = 0.007). Such a difference was not observed for carriers of a MV (50.1 years vs. 52.1 years, p_{T test} = 0.16). For all other genes, no difference in age at diagnosis was observed (data not shown).

Breast cancer risk by hormone receptor status

Estrogen receptor (ER), progesterone receptor (PR) and HER2 gene amplification status of the breast tumor were available for 67%, 63% and 40% of the sequenced cases, respectively. For FANCI, POLH and RTEL1, there were too few observations in each stratum to draw any conclusion on whether carrying a likely deleterious variant was associated with a particular status. Information on HER2 status among carriers of a variant in any of the genes was too scarce to conclude. In the analyses comparing BC risk according to ER or PR status, none of the heterogeneity tests was significant. However, CHEK2 variants were associated with a higher BC risk point estimate for ER-positive (OR = 3.5) compared to ER-negative BC (OR = 1.3); a similar pattern was observed for PR status (Table 3). PALB2 variants were also associated with a higher
BC risk point estimate for ER-positive (OR = 4.6) compared to ER-negative BC (OR = 1.1). MAST1 MV were associated with a higher point estimate for ER-negative (OR = 4.3) compared to ER-positive BC (OR = 2.1).

**Sensitivity analyses**

Because GENESIS cases are prevalent cases, i.e. diagnosed with BC sometimes several years before being invited to participate in the study, cases carrying a pathogenic variant in a gene associated with a poor prognosis would have had less chance to be included and would be therefore under-represented. To examine a survival bias, we excluded cases with a delay between diagnosis of BC and inclusion in GENESIS of greater than 5 years and performed analyses on pseudo-incident cases (N = 663). There was no significant heterogeneity in BC risk according to time between diagnosis and inclusion for ATM, CHEK2, PALB2, FANC1, MAST1 and POLH (Table 3). The observed decreased BC risk associated with carrying a MV in RTEL1 seemed different, although not statistically significant, between cases diagnosed more than 5 years before inclusion and cases diagnosed less than 5 years before inclusion (OR≤5years = 1.1 vs. OR>5years = 0.7, p = 0.51). Therefore, we cannot conclude here that RTEL1 variant carriers have a poorer survival, but more studies are warranted to clarify this.

There was no change in BC risk estimates when adjusting for the number of self-reported cancer cases in the family (BC, ovarian cancer or cancers from the Lynch syndrome spectrum; data not shown) or for the number of altered genes (Table 2). Finally, results by hormone receptor status were unchanged when restricting the analyses to pseudo-incident cases (data not shown).

**Discussion**

By analyzing 113 DNA repair genes selected from either a WES or a candidate gene approach in familial BC cases and a well-identified control group, we confirmed associations between variants predicted to impair protein function in PALB2, ATM and CHEK2 and occurrence of BC. We found that the risk associated with ATM variants can greatly vary according to variant type (OR_{ATM-LoF} \approx 10 x OR_{ATM-MV}). The risk conferred by LoF would still significantly differ from the risk conferred by MV in this gene even if the number of ATM LoF carriers was four times higher in controls. There were no ATM LoF that were observed more than once in cases, and thus none could contribute disproportionately to the risk estimate (Supporting Information Table S3).

We also identified four genes (FANCI, MAST1, POLH and RTEL1) that had not been previously associated with BC susceptibility. Should the association with BC be confirmed in larger studies, the decreased BC risk associated with MV in FANCI, POLH and RTEL1 is puzzling. Because of a difference in BC risk between cases included more than 5 years after BC diagnosis and those included less than 5 years after BC diagnosis, we concluded that RTEL1 alterations might be associated with a poor prognosis. The same hypothesis was tested for carriers of a variant in FANCI or POLH and results were unchanged. Another possible explanation could be that FANCI and POLH are associated with another syndrome than the HBOC syndrome. Therefore due to recruitment criteria in GENESIS, carriers of a variant in one of these genes would be under-represented among the cases of our study.

Apart from FANCI, which we included in our panel because it belongs to the same Fanconi Anemia gene family as BRCA1/ FANCS, BRCA2/FANCD1 and PALB2/FANCN, we investigated MAST1, POLH and RTEL1 because they had been highlighted in the WES exploratory stage, as were the known BC genes ATM and CHEK2. This demonstrates the efficiency of our approach to prioritize new candidate genes. Another strength of our study is that the full coding sequence of the genes was screened in all subjects using the same sequencing techniques and platforms, thus avoiding biases that may arise when only the variants identified in cases are tested in controls. Moreover, unlike many other studies, controls were specifically recruited for the GENESIS study. They were cancer-free friends or colleagues belonging to the same birth cohort as the cases, and were therefore naturally matched on many unmeasured potential confounders.

The literature on MAST1, POLH and RTEL1 has indicated relationships between these genes and carcinogenesis. Putative germline pathogenic variants in MAST1, encoding a microtubule-associated serine–threonine kinase, were described in familial lung cancer cases. Transcriptome analysis of BC cell lines and tissue identified somatic recurrent rearrangements involving MAST1 and over-expression of the resulting gene fusion. Inactivating variants of POLH, which encodes the translesion DNA polymerase η are responsible for Xeroderma Pigmentosum, a rare recessive syndrome associated with hypersensitivity to sunlight and a high frequency of skin cancers at an early age. Of note, a germline POLH MV (c.2074A>G) was identified by WES in a Lebanese BC-prone family but this variant, predicted to be
benign by various tools, was not found in our population. \textit{RTEL1} encodes the regulator of telomere elongation helicase 1, which is essential for telomere maintenance and regulation of homologous recombination. Germline biallelic \textit{RTEL1} variants are responsible for dyskeratosis congenita and its severe variant Hoyerad-Hreidarsson syndrome,\textsuperscript{39} while heterozygous carriers with pulmonary fibrosis,\textsuperscript{40-42} myelodysplasia and liver disease have been described.\textsuperscript{43} Thus, our findings on \textit{FANCI}, \textit{MAST1}, \textit{POLH} and \textit{RTEL1} would warrant further investigation in other large-scale studies on familial BC and the setup of functional studies to assess the impact of the variants.

There was some discrepancy between our findings on ATM in GENESIS and previous findings from a pooled analysis of seven case–control studies that included the Breast Cancer Family Registry (BCFR) study\textsuperscript{9} in terms of risk estimates. In these latter analyses, we had found that carriers of a likely deleterious MV may be at higher risk of BC than LoF carriers.\textsuperscript{9} One explanation could be the difference in the tool predicting the impact of the MV. We therefore reanalyzed GENESIS data for MV in ATM, \textit{CHEK2} and \textit{PALB2} using Align-GVGD\textsuperscript{33} which was used in the BCFR\textsuperscript{9,44} and other studies.\textsuperscript{45,46} Results obtained with Align-GVGD were consistent with those obtained with CADD (Table 4). Another explanation could be the difference in the ages at diagnosis between the studies. In GENESIS, carriers of an ATM LoF or MV were diagnosed on average at age 50.0 (SD = 2.6 years) and 50.9 (SD = 1.2 years), respectively, while most of the cases in the pooled analysis were diagnosed before age 46. Even if in GENESIS carrying an ATM MV was associated with a higher BC risk point estimate in early-onset cases (OR$_{\geq 45\text{years}}$ = 2.2 vs. OR$_{\leq 45\text{years}}$ = 1.2, \textit{p}_{\text{Het}} = 0.06), the risk estimate associated with MV remained much lower than the risk estimate associated with LoF. No difference in BC risk was observed for LoF when stratifying by age (OR$_{\geq 45\text{years}}$ = 18.6 vs. OR$_{\leq 45\text{years}}$ = 17.9, \textit{p}_{\text{Het}} = 0.95). Of note, the MV c.7271T>G (p.Val2424Gly), which had been noted in the BCFR\textsuperscript{9,44} and other studies,\textsuperscript{45,46} and in particular in the Australian population, was not found in GENESIS cases and was found only once in controls. Moreover, there was no evidence that the observed association was driven by a specific variant (either a LoF or a MV) in our study set (Supporting Information Table S3). Conversely, the \textit{CHEK2} LoF c.1100delC was quite prevalent in the French population (Supporting Information Table S3). However, when excluding carriers of this variant in the analysis we found that other \textit{CHEK2} variants (both LoF and MV) contribute significantly to BC risk (OR = 2.5; 95%CI, 1.5–4.4; \textit{p} = 0.0004).

There was also a noteworthy difference in \textit{PALB2} BC relative risk estimates when compared to those estimated by Antoniou \textit{et al.}\textsuperscript{8} for carriers of a LoF. This may be explained by the difference in populations under study. Indeed, Antoniou \textit{et al.} estimated BC risk from highly selected \textit{PALB2} families with uncertainty in modes of ascertainment. Although analyses were performed using modeling to reduce bias, their \textit{PALB2} families may have aggregated many other unmeasured genetic and nongenetic BC risk factors. Another notable finding in our study is the relatively high OR associated with MV in \textit{PALB2}, which have not been systematically assessed in previous reports. Interestingly, the \textit{PALB2} MV c.2816T>G (p.Leu939Trp) was quite frequent in our study population: half of the \textit{PALB2} MV carriers carried this variant (12 out of 20 carriers in cases, and 3 out of 6 in controls; OR$_{c.2816T>G}$ = 4.3; 95%CI, 1.2–15.4; \textit{p} = 0.02). To understand the origin of its high prevalence, which could explain our ability to detect its effect, we performed a haplotype analysis of the 16p12.2 locus containing \textit{PALB2} on GENESIS carriers of the c.2816T>G variant using MERLIN.\textsuperscript{50} In a companion study, genotypes of 38 SNPs localized +/-500 kb around the variant were obtained with the iCOGS array in the GENESIS population (authors’ unpublished data). This latter analysis suggests that the c.2816T>G variant could be a founder allele in the French BC cases (data not shown). A first study showed that this variant falls in the \textit{PALB2} WD40 domain and results in altered \textit{PALB2}-BRCA2 binding.\textsuperscript{51} However, it was then demonstrated that it does not disrupt the HR-mediated DNA repair activity of \textit{PALB2}.\textsuperscript{52} Two case–control studies conducted in the German population\textsuperscript{53} and in the British population\textsuperscript{6} on over 800 and 923 familial BC cases, respectively, showed that the c.2816T>G variant was not associated with BC. More recently, a multicenter case–control study conducted by the Breast Cancer Association Consortium (BCAC) on 42,671 invasive BC and 42,164 controls did not show any association with this variant (OR$_{c.2816T>G}$ = 1.05; 95%CI 0.83–1.32),\textsuperscript{49} leading to the conclusion that the \textit{PALB2} c.2816T>G variant should be classified as a likely neutral variant. However, even if we cannot rule out a false-positive result in our study, one asset of the GENESIS population is to be homogeneous, in contrast with the countries heterogeneity in BCAC. Thus, further investigation on this variant may be worthwhile. Nonetheless, the risk estimates associated with \textit{PALB2} MVs was not driven by this peculiar variant (after excluding carriers of c.2816T>G, OR$_{MV}$ = 3.1; 95%CI, 1.2–7.9; \textit{p} = 0.02).

With respect with the confirmed BC genes, deleterious variants in \textit{ATM}, \textit{CHEK2} and \textit{PALB2} are present in 14% of GENESIS BC cases, of whom 1% carry a LoF in the actionable gene \textit{PALB2} and another 1% carry a \textit{PALB2} MV that would, for the time being, probably be considered as a variant of uncertain clinical significance in a cancer clinics setting. To further evaluate the potential clinical significance of variants considered in our case–control analysis, we determined how many of the 264 LoF and 1,994 likely deleterious MV identified in the 113 genes were present in ClinVar.\textsuperscript{54,55} Only 78 (3.5%) were reported as "pathogenic" or "likely pathogenic" variants (Supporting Information Fig. S3). They were located in 30 genes including \textit{PALB2}, \textit{TP53}, \textit{CDH1}, \textit{RAD51C}, \textit{RAD51D}, \textit{MLH1}, \textit{MSH2} and \textit{PMS2} for which specific management guidelines have been elaborated in France in some specific cancer predisposition contexts (http://www.unicancer.fr/recherche/les-groupes-recherche/groupe-genetique-et-cancer-ggc). This illustrates that although ClinVar is a useful tool for disease gene and variant classification, it should not be the sole tool for the clinical management of hereditary cancer patients, pending the implementation of large-scale studies to determine a precise clinical impact in the future.
tool for variant interpretations, the majority of variants that will be identified by multigene panel testing will be unreported, and efforts to facilitate sharing of data, in particular observation in unaffected populations, together with methods for variant assessment and expertise in specific disease areas are needed.

We chose to investigate each gene independently in order to be able to compare our relative risk estimates with those of other published data. However, we also performed the analyses adjusting for the number of altered genes, and no significant change in relative risk estimates was observed (Table 2). Nevertheless, one cannot exclude that there is a residual quantitative effect shift of number of altered genes in cases as compared to controls (Supporting Information Fig. S2). A potential polygenic effect might be further investigated in larger studies where a larger number of genes can be tested.

In conclusion, this case–control study conducted in a well-characterized population allowed us to describe genetic variation in 113 DNA repair genes and to specify BC relative risks associated with rare deleterious-predicted variants in PALB2, ATM and CHEK2. More generally, this work highlights the importance of considering all types of variants in the analyses, as their contribution might differ from one gene to another. Further family-based studies and gene-specific prospective cohorts are needed to assess absolute risk for carriers of a deleterious variant by taking into account other factors such as family history of cancer, polygenic models or even environmental/lifestyle factors. They will be useful to improve risk prediction models and will help defining gene-specific consensus management guidelines.

Authors’ contribution
Conceived and designed the experiments: RO, FD, AB, JFD, FL. Performed the experiments: LB, NM, ALR. Analyzed the data: EG, ALR, NA, FL. Contributed reagents/materials/analyses tools: MGD, DG, VM, CL, JC, NS. Coordinated the GENESIS study: SM, DSL, NA. Included GENESIS participants, collected data: SEM, DLG, JB, NM, MM, EC. Invited GENESIS index cases: OCH, DL, CP, SFF, CC, TF, FE, AC, CA, AF, LG, JT, AF, JC, SG, IM, FS, SAB, JML, CL, SL, HD, YJB, ML, PP, LVB, VB, PB, EL, CM, CN, CD, JF, PG, LF, AL, BB, OC, MGV, IC, DSL. Wrote the paper: EG, NA, FL. All authors read and approved the final version of the study.

Acknowledgements
We wish to thank all the families who so willingly participated in the GENESIS study.

This article is dedicated to the memory of Olga M. Sinilnikkova who died prematurely on June 30, 2014. Olga participated decisively in structuring research around hereditary predisposition to BC and in leading the GENESIS study with Nadine Andrieu and Dominique Stoppa-Lyonnet. She also contributed to the design of the GENESIS-NGS project.


References
10. Renwick A, Thompson D, Seal S, et al. ATM mutations that cause ataxia-telangiectasia are
18. Li H, Durbin R. Fast and accurate long-read align-
19. Tavtigian SV, Chenevix-Trench G. Growing rec-
20. Ognjanovic S, Olivier M, Bergemann TL, et al. Sarcomas in TP53 germline mutation car-
21. Tavtigian SV, Chenevix-Trench G. Growing rec-
43. Li H, Durbin R. Fast and accurate long-read align-
46. Cogan JD, Kropski JA, Zhao M, et al. Rare vari-
48. Golgde DE, Healey S, Dowty JG, et al. Rare vari-