Supporting Information

Kiiski et al. 10.1073/pnas.1407909111

SI Results and Discussion

In addition to the *FANCM* c.5101C>T variant, five other variants were included in the final analysis. Combining the phase I, II, and III genotyping, the *SLX4* c.2484G>C missense variant and the *MPG* c.40–1G>T splicing variant had similar frequencies between cases and controls (Table S2). The *FANCA* c.4228T>G missense variant was detected in two familial breast cancer cases (0.2%) and the *RUVBL1*c.950G>A missense variant in three breast cancer cases [two familial (0.2%) and one unselected (0.1%)] and one ovarian cancer case (0.2%), but not in any of the controls. The frequency of the *NEIL1* c.314dupC was similar between all breast cancer cases and controls (0.6% and 0.5% in the Helsinki dataset, respectively, and 0.4% for cases and controls in the Tampere dataset], with a somewhat higher frequency among the Tampere sporadic cases [1.7%; OR = 4.57, 95% CI = 1.17–17.75, $P = 0.016$ vs. controls]. One of the unselected breast cancer patients in the Helsinki dataset was homozygous for the c.314dupC mutation. Results for the other subgroups are nevertheless consistent between both datasets, and no statistically significant difference between cases and controls is observed (Table S2). In addition, frequencies for the *BARD1* c.2282G>A and *PAXIP1* c.803C>T missense variants after phase II genotyping were similar to phase I frequencies or too rare for statistical assessment in our dataset, and thus these variants were excluded from further analysis.

Biallelic mutations in the *SLX4* gene have recently been found to cause FA-P, a new subtype of FA (1). Landwehr et al. (2) further studied *SLX4* association with hereditary breast cancer. They identified four previously undescribed variants, but none of them clearly associated with breast cancer. However, a minor contribution for rare variants cannot be ruled out, and, again, larger datasets are needed for more accurate prediction. The *FANCA* c.4228T>G variant has not been previously reported in the FA Mutation Database (www.rockefeller.edu/fanconi/mutate), even though *FANCA* is the most mutated gene in FA, with more than 350 unique mutations reported (3). The role of *FANCA* c.4228T>G as a putative breast cancer-predisposing allele warrants further studies. RUVBL1 is overexpressed in several types of cancer, associates with c-Myc transformation, is a component of chromatin remodeling complexes, and participates in DNA repair (4). Further studies are needed to evaluate the contribution of *RUVBL1* to breast cancer predisposition. NEIL1 is an S-phase regulated protein that excises base lesions from ssDNA, and hypothetically interacts with replication-associated repair and is possibly connected with FA repair pathway. Levels of NEIL1 protein are greatly diminished from cells depleted for FA proteins, but the exact mechanism or correlation of *NEIL1* and FA pathway remains to be further investigated (5).

Truncating and missense mutations in *BARD1* have been identified in breast and ovarian cancer families (6, 7), but the contribution of the gene to breast cancer predisposition remains unclear. The missense variant c.2282G>A identified in our study was in silico predicted to be pathogenic, but it was detected in similar frequency among cases and controls. For *PAXIP1* c.803C>T missense variant, the size of our dataset was not sufficient to disclose the difference in frequencies between cases and controls, and thus the association with breast cancer cannot be excluded.

SI Materials and Methods

Subjects. ***Helsinki breast cancer series.*** The unselected breast cancer patient series from Helsinki was collected at Helsinki University Central Hospital Department of Oncology in 1997–1998 and 2000 (884 patients, including 79% of all consecutive newly diagnosed breast cancer cases during the collection periods) (8, 9) and Department of Surgery in 2001–2004 (986 patients, including 87% of all consecutive newly diagnosed breast cancer cases) (10). Altogether, 1,730 female patients with invasive breast cancer were included in the analysis. The additional familial breast cancer patients were collected at Helsinki University Central Hospital Departments of Clinical Genetics and Oncology (10, 11). When combining the additional familial cases and the unselected series, 524 patients had strong family background with at least three breast or ovarian cancers among first- or second-degree relatives, including the proband, and 568 patients had one first-degree relative affected with breast or ovarian cancer. The samples were genomic DNA isolated from the peripheral blood of the patients. All patients with strong family background were tested negative for *BRCA1/2* mutations, and the patients with one affected relative were tested negative for the Finnish *BRCA1/2* founder mutations as previously described (12–14). The genealogies were confirmed through population registries and cancer diagnoses through the Finnish Cancer Registry and hospital records. Information on ER and progesterone receptor status was collected from pathology reports (15). HER2 status was based on immunohistochemistry and gene amplification with chromogenic in situ hybridization on breast cancer tissue microarrays (16).

Tampere breast cancer series. The unselected breast cancer series from Tampere was collected at Tampere University Hospital as previously described (8, 10), and an additional 336 incident cases were collected in 1996–2004 at Tampere University Hospital. Only invasive cases were included in the analysis. The samples were genomic DNA isolated from peripheral blood.

Helsinki ovarian cancer series. An unselected ovarian cancer series ($n = 233$) was collected at Helsinki University Central Hospital Department of Obstetrics and Gynecology in 1998 as previously described (17). The patients had been treated for invasive epithelial ovarian carcinoma at Helsinki University Central Hospital between 1989 and 1998, and the blood samples were collected in routine follow-up visits to the clinic during 1998 from those patients who were still alive. Additional DNA samples from blood and tumor tissue samples were prospectively collected from patients treated for ovarian cancer at Helsinki University Central Hospital Department of Obstetrics and Gynecology between 1998 and 2006, totaling 569 samples (345 genomic DNA and 204 tumor DNA) in the analysis.

Iceland breast cancer series. The unselected breast cancer series from Iceland consisted of 965 breast cancer cases diagnosed in the period 1987–2004 at Landspitali University Hospital. Among these patients, 92 had a family history of breast cancer (first- and/or second-degree relative diagnosed with breast cancer) and were tested negative for the Icelandic *BRCA1/2* founder mutations (18). The genealogies were conducted by the Genetic Committee of the University of Iceland, and cancer diagnoses were conducted through the Department of Pathology of Landspitali University Hospital and the Icelandic Cancer Registry. Genomic DNA samples were isolated from blood of the patients. The study was approved by the Icelandic Data Protection Authority (reference no. 20010505239 and later amendments) and the National Bioethic Committee (reference no. 99/051-B1/B2 and later amendments).

Population controls. For controls, we genotyped DNA samples from healthy female blood donors from the same geographical regions: 1,274 from Helsinki and 816 from the Tampere region.

Exome Data Analysis. FASTQ-files were quality checked with FASTX-toolkit (hannonlab.cshl.edu/fastx_toolkit/). Sequence reads were aligned to the hg19 human genome build using Burrows–Wheeler Aligner software (version 0.5.9) (19). Sequence realignment and duplicate detection was performed with Picard and Samtools (20). Samtools was used for SNV and indel calling. Integrative Genomics Viewer (21) was used for metrics computing. SnpSift (22) and SnpEff were used for dbSNP and variant effect annotation.

Sanger Sequencing. DNA extracted from blood was amplified by PCR (Table S3) with Biotools DNA polymerase (Biotools). Primers for PCR were designed with Primer3 (frodo.wi.mit.edu/primer3) using NCBI37/Hg19 as a reference sequence. The PCR products were visualized on 2% (wt/vol) agarose gel and purified with ExoSAP-IT (Affymetrix). ABI BigDyeTerminator 3.1 Cycle Sequencing kit (Applied Biosystems) was used for sequencing reactions, and the sequencing was performed at FIMM (University of Helsinki) by using 3730xl DNA Analyzer (Applied Biosystems). Sequence chromatograms were visualized with Variant Reporter (Applied Biosystems) and FinchTV software (Geospiza).

Sequenom Genotyping. The SNV genotyping with Sequenom MassARRAY system was performed by using iPLEX Gold assays (Sequenom) at FIMM (University of Helsinki).

TaqMan Genotyping. TaqMan real-time PCR was performed in a 7500 Fast RealTime PCR System or a 9800 Fast Thermal Cycler by using TaqMan SNP Genotyping assays and TaqMan Genotyping MasterMix (Applied Biosystems). Genotype calling was performed with 7500 Fast RealTime PCR System and ABI Prism 7500 SDS software (version 1.4; Applied Biosystems). The *FANCM* c.5101C>T mutation was genotyped in the Icelandic breast cancer patients by using TaqMan SNP Genotyping assays and TaqMan Genotyping MasterMix. The PCR was run on a StepOne Real-Time PCR system (Life Technologies), and data were collected and analyzed with StepOne software (version 2.0; Life Technologies).

Nonsense-Mediated mRNA Decay Analysis of *FANCM* c.5101C>T. Quantitative allele-specific quantitative RT-PCR was performed for WT and c.5101C>T mutant *FANCM* alleles on RNA extracted from lymphoblastoid cells from a heterozygous carrier of the c.5101C>T variant, a noncarrier from the Finnish population, and five non-Finnish white noncarriers on a 7900HT system (applied Biosystems). PCR primers were as follows: common *FANCM* reverse, AAATTCAGCGATGTCTGTTTGC; WT *FANCM* forward, CAAGCACTGTTAAGAAGAACAAACG**A**C; and c.5101C>T mutant *FANCM* forward, GCACTGTTAAGAAGAACAAAC**GAT**. Both forward primers included a modification of the third nucleotide from the 3′ end to promote allele-specific PCR. Experiments were repeated by using RNA from cells treated with cycloheximide (100 μg/mL) for 4 h to inhibit nonsense-mediated RNA decay. Levels of *FANCM* alleles were normalized relative to GAPDH levels.

1. Kim Y, et al. (2011) Mutations of the SLX4 gene in Fanconi anemia. *Nat Genet* 43(2): 142–146.
2. Landwehr R, et al. (2011) Mutation analysis of the SLX4/FANCP gene in hereditary breast cancer. *Breast Cancer Res Treat* 130(3):1021–1028.
3. Litim N, et al.; INHERIT BRCAs (2013) Polymorphic variations in the FANCA gene in high-risk non-BRCA1/2 breast cancer individuals from the French Canadian population. *Mol Oncol* 7(1):85–100.
4. Gorynia S, et al. (2011) Structural and functional insights into a dodecameric molecular machine - the RuvBL1/RuvBL2 complex. *J Struct Biol* 176(3):279–291.
5. Macé-Aimé G, Couvé S, Khassenov B, Rosselli F, Saparbaev MK (2010) The Fanconi anemia pathway promotes DNA glycosylase-dependent excision of interstrand DNA crosslinks. *Environ Mol Mutagen* 51(6):508–519.
6. De Brakeleer S, et al. (2010) Cancer predisposing missense and protein truncating BARD1 mutations in non-BRCA1 or BRCA2 breast cancer families. *Hum Mutat* 31(3): E1175–E1185.
7. Ratajska M, et al. (2012) Cancer predisposing BARD1 mutations in breast-ovarian cancer families. *Breast Cancer Res Treat* 131(1):89–97.
8. Syrjäkoski K, et al. (2000) Population-based study of BRCA1 and BRCA2 mutations in 1035 unselected Finnish breast cancer patients. *J Natl Cancer Inst* 92(18):1529–1531.
9. Kilpivaara O, et al. (2005) Correlation of CHEK2 protein expression and c.1100delC mutation status with tumor characteristics among unselected breast cancer patients. *Int J Cancer* 113(4):575–580.
10. Fagerholm R, et al. (2008) NAD(P)H:quinone oxidoreductase 1 NQO1*2 genotype (P187S) is a strong prognostic and predictive factor in breast cancer. *Nat Genet* 40(7): 844–853.
11. Eerola H, Blomqvist C, Pukkala E, Pyrhönen S, Nevanlinna H (2000) Familial breast cancer in southern Finland: How prevalent are breast cancer families and can we trust the family history reported by patients? *Eur J Cancer* 36(9):1143–1148.
12. Vehmanen P, et al. (1997) Low proportion of BRCA1 and BRCA2 mutations in Finnish breast cancer families: Evidence for additional susceptibility genes. *Hum Mol Genet* 6(13):2309–2315.
13. Vahteristo P, et al. (2002) A CHEK2 genetic variant contributing to a substantial fraction of familial breast cancer. *Am J Hum Genet* 71(2):432–438.
14. Vahteristo P, Eerola H, Tamminen A, Blomqvist C, Nevanlinna H (2001) A probability model for predicting BRCA1 and BRCA2 mutations in breast and breast-ovarian cancer families. *Br J Cancer* 84(5):704–708.
15. Eerola H, et al. (2005) Histopathological features of breast tumours in BRCA1, BRCA2 and mutation-negative breast cancer families. *Breast Cancer Res* 7(1):R93–R100.
16. Tommiska J, et al. (2008) The DNA damage signalling kinase ATM is aberrantly reduced or lost in BRCA1/BRCA2-deficient and ER/PR/ERBB2-triple-negative breast cancer. *Oncogene* 27(17):2501–2506.
17. Sarantaus L, et al. (2001) BRCA1 and BRCA2 mutations among 233 unselected Finnish ovarian carcinoma patients. *Eur J Hum Genet* 9(6):424–430.
18. Arason A, et al. (1998) A population study of mutations and LOH at breast cancer gene loci in tumours from sister pairs: Two recurrent mutations seem to account for all BRCA1/BRCA2 linked breast cancer in Iceland. *J Med Genet* 35(6):446–449.
19. Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26(5):589–595.
20. Li H, et al.; 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
21. Robinson JT, et al. (2011) Integrative genomics viewer. *Nat Biotechnol* 29(1):24–26.
22. Cingolani P, et al. (2012) Using drosophila melanogaster as a model for genotoxic chemical mutational studies with a new program, SnpSift. *Front Genet* 3:35.

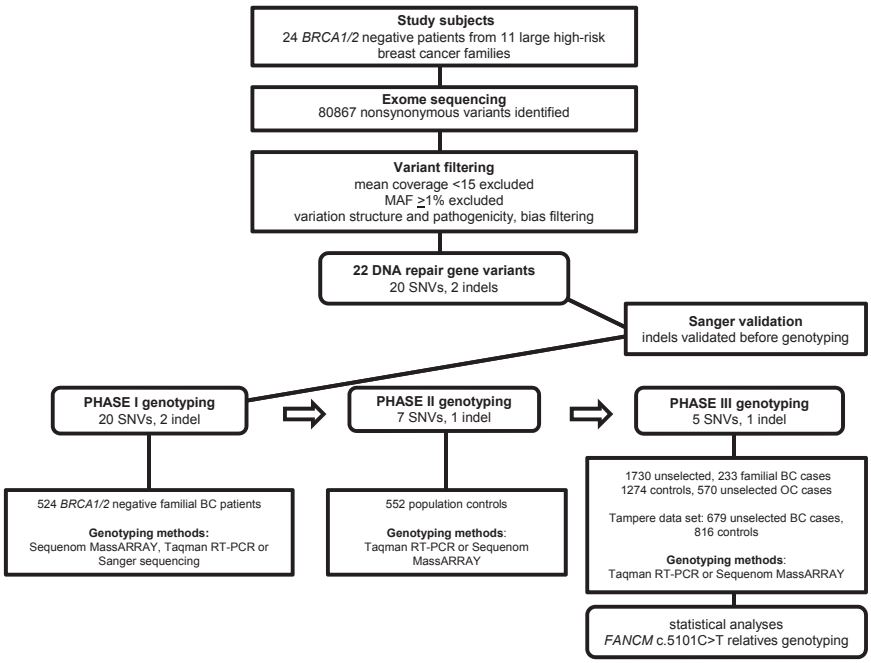

Fig. S1. Schematic representation of the study. We performed exome sequencing for 24 *BRCA1/2*-negative high-risk familial breast cancer patients. The exome data was filtered to prioritize rare variants in DNA repair genes. The selected variants were further genotyped in three consecutive phases, allowing us to concurrently exclude from further analysis plausibly too rare (unique) variants or those found prevalent in the population controls. After phase III genotyping of the full dataset of breast and ovarian cancer cases and population controls, breast cancer association of *FANCM* c.5101C>T nonsense mutation was confirmed with statistical analyses. Consecutively relatives of *FANCM* c.5101C>T mutation carriers were genotyped. BC, breast cancer. OC, ovarian cancer.

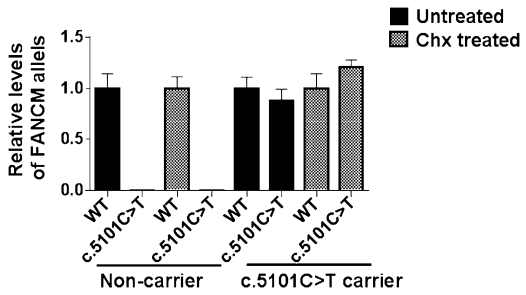

Fig. S2. Results of the nonsense-mediated mRNA decay analysis. Quantitative allele-specific *FANCM* RT-PCR of RNA extracted from lymphoblastoid cells from a noncarrier and a heterozygous carrier of the c.5101C>T mutant. Alleles were quantified in untreated and cycloheximide-treated (to inhibit nonsense-mediated decay) cells. Levels of *FANCM* alleles are normalized relative to GAPDH levels.

Table S1. Details of the variants selected for genotyping phases I–III

Position	Gene	Rs number	HGVS name	Protein level	Transcript	Exonic function	Genotyping phases and methods
11:108202202	ATM	—	c.7547T>G	p.Phe2516Cys	ENST00000278616	Missense	iPLEX (phase I)
9:130549857	CDK9	—	c.586G>A	p.Val196Ile	ENST00000373265	Missense	iPLEX (phase I)
10:115612530	DCLRE1A	rs41292634	c.412C>T	p.Arg138*	ENST00000361384	Nonsense	iPLEX (phase I)
11:108559787	DDX10	—	c.973G>T	p.Glu325*	ENST00000526794	Nonsense	iPLEX (phase I)
16:89846310	FANCA	—	c.1682C>T	p.Thr561Met	ENST00000389301	Missense	iPLEX (phase I)
7:99693716	MCM7	—	c.748G>C	p.Glu250Gln	ENST00000354230	Missense	iPLEX (phase I)
4:178272607	NEIL3	rs141868607	c.943C>T	p.Arg315Trp	ENST00000264596	Missense	iPLEX (phase I)
3:51980263	PARP3	—	c.1180A>G	p.Ile394Val	ENST00000417220	Missense	iPLEX (phase I)
17:62489111	POLG2	rs139282177	c.590T>C	p.Leu197Pro	ENST00000539111	Missense	iPLEX (phase I)
8:95416345	RAD54B	—	c.904G>C	p.Glu302Gln	ENST00000336148	Missense	iPLEX (phase I)
18:20581567	RBBP8	—	c.2162A>G	p.Asn721Ser	ENST00000327155	Missense	iPLEX (phase I)
22:45750901	SMC1B	—	c.3056C>A	p.Pro1019Gln	ENST00000357450	Missense	iPLEX (phase I)
3:48508028	TREX1	—	c.139G>A	p.Gly47Ser	ENST00000422277	Missense	iPLEX (phase I)
6:146243948	SHPRH	—	c.3577_3580delCTTA	p.Leu1193Glnfs*7	ENST00000367505	Frame-shift deletion	Sanger sequencing (phase I)
7:154767677	PAXIP1	rs61752015	c.803C>T	p.Pro268Leu	ENST00000404141	Missense	TaqMan (phases I/II)
2:215593452	BARD1	rs142155101	c.2282G>A	p.Ser761Asn	ENST00000260947	Missense	iPLEX (phase I), TaqMan (phase II)
3:127816209	RUVBL1	—	c.950G>A	p.Arg316His	ENST00000322623	Missense	iPLEX (phases I-III)
16:3641155	SLX4	rs199656607	c.2484G>C	p.Glu828Asp	ENST00000294008	Missense	iPLEX (phases I-III)
16:89805322	FANCA	—	c.4228T>G	p.Cys1410Gly	ENST00000389301	Missense	iPLEX (phases I-III)
14:45658326	FANCM	rs147021911	c.5101C>T	p.Gln1701*	ENST00000267430	Nonsense	iPLEX (phases I-III)
16:129423	MPG	rs146600185	c.40–1G>T	p.?	ENST00000219431	Splice site	iPLEX (phases I-III)
15:75641560	NEIL1	—	c.314dupC	p.Pro106Alafs*50	ENST00000355059	Frame-shift insertion	Taqman (phases I-III)

Genomic positions are indicated according to genome build hg19.

Table S2. Genotype frequencies for variants that were included in phase I-II or I-III

Variation	Study cohort	No. of individuals	WT	%	mut	%	OR	95% CI	*P* value
			GG		GA				
RUVBL1 c.950G>A	Helsinki region								
p.Arg316His	Population control	1,273	1,273	100	0	0	—	—	—
	Familial BC	1,076	1,074	99.8	2	0.2	—	—	—
	Unselected BC	1,716	1,715	99.9	1	0.1	—	—	—
	All BC	2,410	2,407	99.9	3	0.1	—	—	—
	Unselected OC	551	550	99.8	1	0.2	—	—	—
	Tampere region								
	Population control	810	810	100	0	0	—	—	—
	Familial BC	254	254	100	0	0	—	—	—
	All BC (unselected)	675	675	100	0	0	—	—	—
			TT		TG				
FANCA c.4228T>G	Helsinki region								
p.Cys1410Gly	Population control	1,271	1,271	100	0	0	—	—	—
	Familial BC	1,075	1,073	99.8	2	0.2	—	—	—
	Unselected BC	1,714	1,714	100	0	0	—	—	—
	All BC	2,408	2,406	99.9	2	0.1	—	—	—
	Unselected OC	547	547	100	0	0	—	—	—
	Tampere region								
	Population control	810	810	100	0	0	—	—	—
	Familial BC	254	254	100	0	0	—	—	—
	All BC (unselected)	674	674	100	0	0	—	—	—
			GG		GC				
SLX4 c.2484G>C	Helsinki region								
p.Glu828Asp	Population control	1,270	1,264	99.5	6	0.5	—	—	—
	Familial BC	1,064	1,059	99.5	5	0.5	0.99	0.30–3.27	1
	Unselected BC	1,713	1,708	99.7	5	0.3	0.62	0.19–2.03	0.5442
	All BC	2,396	2,388	99.7	8	0.3	0.71	0.24–2.04	0.5175
	Unselected OC	547	546	99.8	1	0.2	0.29	0.04–2.33	0.6817
	Tampere region								
	Population control	810	808	99.8	2	0.2	—	—	—
	Familial BC	254	253	99.6	1	0.4	1.60	0.14–17.68	0.5591
	All BC (unselected)	675	672	99.6	3	0.4	1.80	0.30–10.83	0.6641
			C		CC*				
NEIL1 c.314dupC	Helsinki region								
p.Pro106Alafs*50	Population control	1,274	1,266	99.4	8	0.6	—	—	—
	Familial BC	1,085	1,080	99.5	5	0.5	0.73	0.11–10.18	0.7817
	Unselected BC	1,730	1,719	99.4	11	0.6	1.01	0.41–2.52	0.9785
	All BC	2,815	2,799	99.4	16	0.6	0.90	0.39–2.12	0.8174
	Tampere region								
	Population control	813	810	99.6	3	0.4	—	—	—
	Familial BC	257	256	99.6	1	0.4	1.05	0.11–10.18	1
	All BC (unselected)	678	670	98.8	8	1.2	3.22	0.85–12.20	0.0684
			GG		GT				
MPG c.40–1G>T	Helsinki region								
p.?	Population control	1,272	1,261	99.1	11	0.9	—	—	—
	Familial BC	1,075	1,062	98.8	13	1.2	1.40	0.63–3.15	0.4085
	Unselected BC	1,713	1,687	98.5	26	1.5	1.77	0.87–3.59	0.1108
	All BC	2,425	2,394	98.7	31	1.3	1.48	0.74–2.96	0.2597
	Unselected OC	548	540	98.5	8	1.5	1.70	0.68–4.26	0.2490
	Tampere region								
	Population control	809	797	98.5	12	1.5	—	—	—
	Familial BC	254	251	98.8	3	1.2	0.79	0.22–2.84	1
	All BC (unselected)	684	669	97.8	15	2.2	1.49	0.69–3.20	0.3052
			GG		GA				
BARD1 c.2282G>A	Helsinki region								
p.Ser761Asn	Population control	549	539	98.2	10	1.8	—	—	—
	Familial BC	505	494	97.8	11	2.2	1.20	0.51–2.85	0.6789

Table S2. Cont.

Variation	Study cohort	No. of individuals	WT	%	mut	%	OR	95% CI	*P* value
			GG		GA				
PAXIP1 c.803C>T	Helsinki region								
p.Pro268Leu	Population control	547	544	99.5	3	0.5	—	—	—
	Familial BC	489	483	98.8	6	1.2	2.25	0.56–9.06	0.3204

BC, breast cancer; mut, mutant; OC, ovarian cancer.
*Includes one homozygous *NEIL1* c.314dupC carrier.

Table S3. Primers used in Sanger validation and genotyping

Gene	Mutation	Forward primer	Reverse primer	T_m (°C)
SHPRH	c.3577_3580delCTTA	TCTGAATTTGAGGAAATTGACC	TGGTTTTTCATACCTTGTCCTTG	59
NEIL1	c.314dupC	GACCCTCCGAGTTCTCCTCT	GTTGGCCAAGAAGGCACTAA	58
FANCM	c.5101C>T	GAGTCTTGCAAAGGCCAATC	TCAGCGATGTCTGTTTGCTC	58

T_m, annealing temperature.