Published in final edited form as:

Cancer Discov. 2014 July ; 4(7): 804–815. doi:10.1158/2159-8290.CD-14-0212.

Rare mutations in *RINT1* predispose carriers to breast and Lynch Syndrome-spectrum cancers

Daniel J. Park^{1,*}, Kayoko Tao^{2,*}, Florence Le Calvez-Kelm^{3,*}, Tu Nguyen-Dumont¹, Nivonirina Robinot³, Fleur Hammet¹, Fabrice Odefrey¹, Helen Tsimiklis¹, Zhi L. Teo¹, Louise B. Thingholm¹, Erin L. Young², Catherine Voegelé³, Andrew Lonie⁴, Bernard J. Pope^{4,5}, Terrell C. Roane⁶, Russell Bell², Hao Hu⁷, Shankaracharya⁷, Chad D. Huff⁷, Jonathan Ellis⁸, Jun Li⁸, Igor V. Makunin⁸, Esther M. John^{9,10} for the Breast Cancer Family Registry, Irene L. Andrulis¹¹, Mary B. Terry¹², Mary Daly¹³, Sandra S. Buys¹⁴, Carrie Snyder¹⁵, Henry T. Lynch¹⁵, Peter Devilee¹⁶, Graham G. Giles^{17,18}, John L. Hopper^{18,19}, Bing J. Feng^{14,20}, Fabienne Lesueur^{3,21,#}, Sean V. Tavtigian^{2,#}, Melissa C. Southey^{1,#,^} for the Kathleen Cuninghame Foundation Consortium for Research into Familial Breast Cancer, and David E. Goldgar^{14,20,#,^}

¹Genetic Epidemiology Laboratory, The University of Melbourne, Victoria 3010, Australia

²Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, UT 84112, USA ³Genetic Cancer Susceptibility Group, International Agency for Research on Cancer, 69372 Lyon, France ⁴Victorian Life Sciences Computation Initiative, Carlton, Victoria 3010, Australia ⁵Department of Computing and Information Systems, The University of Melbourne, Victoria 3010, Australia ⁶University of Texas at Austin, Austin, TX 78712, USA ⁷Department of Epidemiology, The University of Texas M.D. Anderson Cancer Center, Houston, TX 77030, USA ⁸The QIMR Berghofer Medical Research Institute, 300 Herston Road, Herston, QLD 4006, Australia ⁹Cancer Prevention Institute of California, Fremont, CA 94538, USA ¹⁰Department of Health Research and Policy, Stanford Cancer Institute, Stanford, CA 94305, USA ¹¹Department of Molecular Genetics, Samuel Lunenfeld Research Institute,

*Authors for correspondence: Professor Melissa C. Southey, Department of Pathology, The University of Melbourne, Victoria 3010, Australia. msouthey@unimelb.edu.au, Phone +61 3 8344 4895, Fax +61 3 8344 4004; Professor David E. Goldgar, Department of Dermatology, Huntsman Cancer Institute, Salt Lake City 84112, USA. david.goldgar@hsc.utah.edu, Phone +1 801 585 0337, Fax +1 801 585 0990.

*These authors contributed equally to the work.

#These authors contributed equally to the work.

Conflict of Interest: The authors declare that they have no conflicts of interest related to the work presented in this manuscript
Supplemental Data: See supplemental data.doc

Web Resources: Align-GVGD website: <http://agvgd.iarc.fr/alignments>

GATK v.1.0.4418: <http://gatk.sourceforge.net/>

Genome Viewer (IGV v.1.5.48): <http://www.broadinstitute.org/software/igv/>

Mendelian Inheritance in Man: <http://www.omim.org>

MGI Mammalian Phenotype Ontology: http://www.informatics.jax.org/searches/MP_form.shtml

Picard v.1.29: <http://sourceforge.net/projects/picard/>

PolyPhen2: <http://genetics.bwh.harvard.edu/pph2/>

kConFab resource: www.kconfab.org

SOLiD Baylor protocol 2.1.: http://www.hgsc.bcm.tmc.edu/documents/Preparation_of_SOLiD_Capture_Libraries.pdf

UCSC Genome Browser: <http://genome.ucsc.edu/cgi-bin/hgGateway>

TCOFFEE: <http://www.tcoffee.org/>

The Exome Variant Server: <http://evs.gs.washington.edu/EVS/>

1000 Genomes: <http://browser.1000genomes.org/index.html>

Mount Sinai Hospital, Toronto, ON M5G 1X5, Canada ¹²Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY 10032, USA ¹³Fox Chase Cancer Center, Philadelphia, PA 19111, USA ¹⁴Huntsman Cancer Institute, University of Utah Health Sciences Center, Salt Lake City, UT 84112, USA ¹⁵Department of Preventive Medicine, Creighton University, Omaha, NE 68178 ¹⁶Department of Human Genetics, Leiden University Medical Center, Leiden, Netherlands ¹⁷Centre for Cancer Epidemiology, The Cancer Council Victoria, Carlton, Victoria 3004, Australia ¹⁸Centre for, Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Victoria 3010, Australia ¹⁹School of Public Health, Seoul National University, Seoul, Korea ²⁰Department of Dermatology, University of Utah School of Medicine, Salt Lake City, UT 84132 USA ²¹Genetic Epidemiology of Cancer Team, Inserm, U900, Institut Curie, Mines ParisTech

Abstract

Approximately half of the familial aggregation of breast cancer remains unexplained. A multiple-case breast cancer family exome sequencing study identified three likely pathogenic mutations in *RINT1* (NM_021930.4) not present in public sequencing databases: *RINT1* c.343C>T (p.Q115X), c.1132_1134del (p.M378del) and c.1207G>T (p.D403Y). Based on this finding, a population-based case-control mutation-screening study was conducted and identified 29 carriers of rare (MAF < 0.5%), likely pathogenic variants: 23 in 1,313 early-onset breast cancer cases and 6 in 1,123 frequency-matched controls (OR=3.24, 95% CI 1.29-8.17; p=0.013). *RINT1* mutation screening of probands from 798 multiple-case breast cancer families identified 4 additional carriers of rare genetic variants. Analysis of the incidence of first primary cancers in families of women in *RINT1*-mutation carrying families estimated that carriers were at increased risks of Lynch syndrome-spectrum cancers (SIR 3.35, 95% CI 1.7-6.0; P=0.005), particularly for relatives diagnosed with cancer under age 60 years (SIR 10.9, 95% CI 4.7-21; P=0.0003).

Keywords

Breast Cancer; *RINT1*; Lynch Syndrome; Genetic Susceptibility

Introduction

Large population-based studies have established that women with a family history of breast cancer are at ~2–3-fold increased risk of the disease (1-3). Although there has been substantial progress in the last 20 years in identifying genetic causes of breast cancer, a considerable proportion of the familial risk remains unexplained. The current estimate of the proportion of familial risk explained by rare mutations in the high-risk breast cancer susceptibility genes *BRCA1* (MIM 113705), *BRCA2* (MIM 600185) and *PALB2* (MIM 610355) is 20%, and another 28% is explained by associations with common genetic variants (4), although these proportions are highly dependent on age at diagnosis. A further ~5% is likely to be explained by mutations in genes in the homologous recombination repair (HRR) detection and signalling pathways, including *MRE11A* (MIM600814), *RAD50* (MIM604040), *NBN* (MIM 602667), *ATM* (MIM 607585), and *CHEK2* (MIM 604373) and

the RAD51 paralogs *RAD51C* (MIM 602774), *RAD51D* (MIM 602954), and *XRCC2* (MIM 600375) that are all currently under investigation. The potential for intra-family exome-sequencing approaches to identify additional breast cancer susceptibility genes has been demonstrated recently (5-7). While larger scale validation of findings from these studies has provided further insight into the possible roles and clinical significance of these susceptibility genes in cancer predisposition, these studies have also illustrated the challenges posed by conducting these studies in the context of breast cancer, a disease that has variable phenotypes and likely to involve a great number of intermediate- to high-risk rare variants in multiple susceptibility genes (5,6,8-11).

In many, if not most, cases, mutations in specific cancer susceptibility genes are associated with high risks of cancer at one or two primary sites, but are also associated with more modest increases in risk for a number of other cancer types. For instance, *BRCA2* mutations confer high risks of breast and ovarian cancer but are also associated with increased risks of prostate cancer with an estimated relative risk of 2.5 (12) and pancreatic cancer with increased risks of six-fold in two independent large studies (12,13). Perhaps more noteworthy examples are the genes in which defects lead to deficiencies in mismatch repair that cause Lynch syndrome (MIM 120435); their defining associations are with colorectal and endometrial cancers, but they are also associated with a spectrum of other tumour types including ovarian cancer, stomach cancer, bladder cancer, kidney cancer, pancreatic cancer, and brain tumours (14-16). Thus it is plausible that mutations in other breast cancer susceptibility genes might also predispose to other cancers and should be considered in susceptibility gene discovery efforts.

Here, we conducted whole exome sequencing of women affected with breast cancer at an early age from highly selected multiple-case breast cancer families followed by gene-specific mutation screening using large international breast cancer research resources to identify *RINT1* as a putative breast cancer predisposition gene that appears also to be associated with risk for a spectrum of other cancers similar to those that have previously been described for Lynch syndrome.

Results

Whole exome sequencing

Bioinformatic analysis of the exome sequences of 89 women with early-onset breast cancer from 47 families identified three mutations in the gene coding for the RAD50-interacting protein 1 (*RINT1*; NM_021930.4): c.343C>T (p.Q115X), c.1132_1134del (p.378del) and c.1207G>T (p.D403Y), each occurring in a separate family and not found in the public sequencing databases. The nonsense substitution was carried by both women included in the WES from the family in which it was observed. The in-frame deletion (c.1132_1134del) and missense substitution (c.1207G>T; p.D403Y) were observed in one WES subject each, but both fall at protein positions that are highly evolutionarily conserved and are thus likely to be damaging to protein function. (Figures 1 and 2).

Additional genotyping in WES pedigrees

In the family in which *RINT1* c.343C>T (p.Q115X) was identified, the mutation was carried by both women for whom WES was conducted (II-2 and I-4, Figure 1a), and subsequent testing revealed that the two other female relatives diagnosed with breast cancer in the family (II-1 and II-4) and a male relative (I-2) affected by bladder and lung cancers were also carriers, Figure 1a. The *RINT1* c.1207G>T (p.D403Y) variant was identified in one of the two affected women selected for WES (III-8 but not III-1, Figure 1b). This family had an extensive family history of cancer including nine diagnoses of breast cancer in seven female members (two women were found to be carriers (III-8, II-10), two found to be non-carriers (III-1 and II-12) and three could not be genotyped (II-1, II-7 and I-4) for c.1207G>T. The family also had four individuals (II-4, II-9, II-10, III-10) who had malignant melanoma at early ages (39, 39, 36 and 33, respectively). Of these cases, three were carriers of c.1207G>T (II-4, II-9, II-10) and one could not be genotyped. The *RINT1* c.1132_1134del (p.378del) variant was carried by one of the two women selected for WES (II-1 but not III-1) while the carrier's mother (I-4), affected by breast cancer at age 42, could not be genotyped. Of the eight additional relatives diagnosed with other cancers, seven could not be genotyped and one, affected by brain cancer at the age of 32, did not carry this variant (Figure 1c). Figure 1d shows the evolutionary conservation of the relevant amino acid positions for the missense (c.1207G>T; D403Y) and in-frame deletion (c.1132_1134del) variants described above.

Based on the findings from exome sequencing and pedigree follow-up, and the gene ontology-based plausibility of *RINT1* as a breast cancer susceptibility gene, we conducted two further studies in parallel: (i) Case-control mutation screening of *RINT1* and mutation screening of *RINT1* in a series of index cases from multiple-case breast cancer families and (ii) *RINT1* mutation screening in probands of multiple-case breast cancer families, as described in Methods.

Case-control mutation screening

We identified 31 distinct rare variants (MAF<0.5%) as shown in Table 1. Table 2 shows the results of comparisons of the frequency of the various classes of rare sequence *RINT1* variants in cases and controls. There was a statistically significant difference in the frequency of rare sequence variants (independent of *in silico* assessment) in cases compared to controls (OR=2.2; p=0.019). When we consider only those variants that are scored as potentially damaging by PolyPhen2 or Align-GVGD, six of 1,123 controls carried such a variant compared to 23 of 1,313 cases (OR=3.24, 95%CI 1.29-8.17; p=0.013). These variants were distributed across ethnicity as follows: Caucasians 15 pathogenic variants in cases vs. 4 in controls; East Asians 0 cases vs. 1 control; Latinas 5 cases vs. 1 control; Recent African Ancestry 3 cases vs. 0 controls.

RINT1 screening in probands from multiple-case breast cancer families

Screening in an additional 798 multiple-case breast cancer families identified five rare sequence variants in six families (minor allele frequency (MAF) <0.5% in all public sequencing database populations): the missense variants *RINT1* c.413C>T, p.A138V (2 families); c.1256C>G, p.P419R (1 family); and c.1385C>T, p.S462L (1 family); these latter

two families were also included in the case-control mutation screening study through independent ascertainment. All of these variants are predicted to be ‘probably pathogenic’ by PolyPhen2 or have a grade greater than C0 by Align-GVGD *in silico* analysis. In this set of families we also identified the variants c.1333+1G>A (1) and c.129 C>A, p.V43V (1), predicted to affect splicing of the *RINT1* transcript and further described below.

Assessment of variants predicted to cause aberrant splicing

The case-control mutation screening revealed a complex variant, c.{1334-5delA; 1334-1_1335delGTT}, falling within the intron 9 splice acceptor sequence. Application of a minigene splicing assay demonstrated that the variant results in an in-frame deletion of the first three nucleotides of *RINT1* exon 10 (Supplemental Figure 1). Screening in multiple case families revealed two other sequence variants that were predicted to interfere with splicing: c.1333+1G>A and c.129C>A, p.V43V. RT-PCR assays performed from LCL cDNA derived from a carrier of 1333+1G>A revealed that this donor variant causes skipping of exon 9; since exon 9 is 226 nucleotides long, the result is a frameshift. RT-PCR assays performed from LCL cDNA derived from a carrier of 129C>A, p.V43V revealed that this variant creates a *de novo* splice donor within exon 3; the resulting deletion of 147 nucleotides from the end of exon 3 produces an in-frame deletion of 49 amino acids including a number of evolutionarily invariant positions in the protein.

Cancer incidence in *RINT1* mutation-positive families

Preliminary comparisons of cancer incidence between the 23 families with rare likely pathogenic *RINT1* variants and the 1260 *RINT1* negative families for cancer anatomical sites/site groups showed nominally statistically significant ($p<0.05$) excesses of all non-breast cancers ($p=0.03$), digestive tract cancers ($p=0.02$), and female genital cancers ($p=0.0009$). In addition, there were significantly fewer respiratory cancers in *RINT1* mutation-carrying families ($p=0.026$) than in the families in which the affected proband was found not to carry a *RINT1* variant.

Given this observed constellation of cancers showing evidence of increased incidence in *RINT1* mutation-positive families, we created additional groupings as follows: 1) Classic Lynch syndrome consisting of cancers of the colon, stomach, intestine, rectum and endometrium (17); 2) Lynch Spectrum cancers consisting of all digestive cancers except for esophageal, female genital cancers, eye/brain/CNS cancers, and urinary tract cancers, as these have all been shown to be in excess in Lynch syndrome families (14-16); 3) All non-breast cancers except Lynch syndrome as noted in 1) above and 4) All non-breast cancers except those included in Lynch spectrum as noted in 2) above. These groupings were then analyzed using two more detailed approaches as described in Methods. The results of these analyses are presented in Table 3. In general, we see increased risks of Lynch spectrum cancers in both the comparison of relatives in *RINT1* mutation-carrying families to those in the families of cases without rare variants in *RINT1*, and in the comparison of risk to *RINT1* mutation carriers with expected values based on population incidence rates. There was no evidence for increased risks associated with cancers that fall outside the Lynch syndrome/extended Lynch spectrum. Analyses with follow-up time restricted to before age 60 and/or

incorporating genotyping data resulted in higher estimates of risk for the sites examined, which further strengthens the role of *RINT1* in susceptibility to these cancers.

Discussion

Here, we combined whole exome sequencing of women affected with breast cancer from highly selected multiple-case breast cancer families and additional mutation screening in large international breast cancer research resources to identify a new breast cancer predisposition gene, *RINT1*, that appears also to be associated with risk for a spectrum of other cancers similar to those that have previously been described for Lynch syndrome.

Ideally, the transition from WES to detailed studies of individual candidate genes in a multi-step susceptibility gene identification study such as this one would be driven by exome-wide statistical analysis of the WES data. After most of the analyses described in this work were completed, a new tool for pedigree-based studies, pVAAST (Hu et al., submitted), was available within the sequence analysis software package VAAST (18,19). We performed a genome-wide analysis of the breast cancer exomes using pVAAST, accounting for the genealogic relationships between the subjects sequenced and comparing against 88 CEU (Utah Residents (CEPH) with Northern and Western European ancestry) and GBR (British in England and Scotland) control exomes from the 1000 Genomes Project. Post hoc, there was a significant excess of *RINT1* variants ($p=0.029$ based on a permutation test of cases and controls).

RINT1 was originally identified as a RAD50-interacting protein, and overexpression of truncated *RINT1* was found to result in a defective radiation-induced G2/M checkpoint (20). Subsequent to its discovery as a RAD50 interactor, *RINT1* was also shown to be involved in several aspects of Golgi apparatus dynamics. The protein appears to play a role in pericentriolar positioning of the Golgi, Golgi assembly, and partitioning of the Golgi apparatus between daughter cells during mitosis (21-23). *RINT1* is also a component of the NAG, *RINT1*, ZW10 (NRZ) complex, which likely acts as a tethering complex that harnesses dynein for retrograde trafficking of coat protein I (COPI) coated vesicles from the Golgi to the endoplasmic reticulum (24,25). Nonetheless, neither the distribution of protein truncating mutations nor the distribution of missense substitutions that we observed at conserved *RINT1* residues provide immediate insight into which *RINT1* function(s) are important to its role as a cancer susceptibility gene.

The analyses of non-breast cancers in the *RINT1* mutation positive families further extends our understanding of breast cancer susceptibility in the broader context of cancer susceptibility syndromes that include increased risk to breast cancer. The pattern of cancer susceptibility observed in the families included in this report is very similar to that in Lynch syndrome, which is associated with DNA mismatch repair deficiency and characterized by high increased risks of colorectal and endometrial cancers and more moderate increased risk of a number of other cancers including ovarian, small intestine and skin cancers (17).

Although not part of any formal analysis, Figure 1 provides additional, anecdotal support of the hypothesis that *RINT1* mutations are associated with the spectrum of Lynch syndrome-

like cancers. These families were recruited and included in our exome sequencing project due to having multiple-cases of breast cancer, but these families also have a number of other cancers including bladder (BIC; dx 68,40 and 86 years) and uterine (UC; dx 30 years) cancers consistent with a Lynch syndrome-like (mismatch repair deficiency) phenotype. These families also have a number of other cancers that are not considered part of Lynch syndrome, including malignant melanoma (MM; dx 39, 40, 36, 33 years) and lung cancer (LC; dx 81, 61, 43, 66 years) for which our broader family-based analyses did not demonstrate significant enrichment in *RINT1* mutation carrier families. In the five families from the targeted resequencing study of candidate genes in probands from the kConFab resource, we also identified a number of cancers in relatives of the probands including colon (dx 47 years), small intestine (dx 63 years), bladder (dx 73 and 72 years), cervix (dx 26 years) and pancreas (dx 79 years).

In our sensitivity analyses, we observed some variation with regard to the choice of the year of start of follow-up. In general, results derived from beginning follow-up before 1920 showed smaller/less significant effects than those yielded from follow-up after 1920 until 1940 when the smaller sample sizes led to more unstable estimates. When we included control families in the analysis, the HR estimates were in general slightly lower, but in many cases were associated with lower p-values. Lastly, upon relaxing our criteria for inclusion of variants, we found that when comparing all rare *RINT1* variants, the HR estimates were broadly similar to prior analyses (in some cases higher), indicating that among the variants excluded from the main analysis, there are likely to be some that are associated with increased risk. Functional analyses may be required to identify the relevant pathogenic variants from the entire set; this refinement would be expected to yield increased estimates of risk for associated cancers.

Because many of the cancers (particularly those at sites other than breast or ovary) are taken from reports of relatives (most often the family proband) we also repeated some analyses using only cancers verified by pathology report, death certificate, medical or records or self-report. Although this reduced the number of events by 50 – 70% depending on site, the effects we observed remained statistically significant, albeit with smaller effect sizes. For example, the main Lynch syndrome HR went from 3.24 to 2.64 and the associated p-value increased from 4×10^{-5} to 0.004. However, we note that a similar effect was seen in the analysis of cancers outside the Lynch spectrum so that the relative magnitude of the estimates remained quite similar. Further, the accuracy of reporting of relative's cancers is unlikely to be related to *RINT1* sequence variant status, as such genetic status was not known to the proband at the time of the reporting. Finally, the pattern of increased risk in Lynch-spectrum cancers compared to all other cancer sites was remarkably similar when the analyses were restricted to verified or self-report cancers only. However, validation of *RINT1* as a susceptibility gene for other sites require the sequencing of the gene in large series of cases and controls, as well as families that fit the Lynch spectrum criteria but without mutations in the known mismatch repair genes.

Our report further illustrates the capacity of massively parallel sequencing in the discovery of additional breast cancer susceptibility genes when used with an appropriate study design. Our report also gives weight to a concept that information from mouse gene-phenotype

relationships and other similar resources could complement a gene ontology-driven analysis and potentially be very helpful in identifying genes of interest from sequencing data sets acquired during studies of cancer susceptibility. Based on the data presented here, we would recommend that *RINT1* be added to genes currently being tested by targeted gene panels that are now being increasingly used in commercial/diagnostic genetic testing of cancer patients and their families.

Material and Methods

Subjects

Whole Exome Sequencing—DNA samples from 89 women with breast cancer from 47 families were selected for exome capture followed by massively parallel sequencing (whole exome sequencing, WES). These samples were contributed by a number of breast cancer research resources (26-28). Selection criteria for inclusion were three cases of breast cancer in the family diagnosed under age 50 with two cases who were more distant than first-degree relatives of each other having a DNA sample available for WES analysis. All participants provided written informed consent for participation in genetic studies. WES studies were approved by the University of Utah Institutional Review Board (IRB) and The University of Melbourne, Human Research Ethics Committee (HREC).

Case-control mutation screening—Eligible participants included women ascertained by population-based sampling by the Australia, Northern California, and Ontario sites of the BCFR between 1995 and 2005 (26). For the Ontario and Northern California sites, cases diagnosed at age 35 years or older or 36 years and older, respectively, that met high-risk criteria (bilaterality, Ashkenazi origin, family history) were preferentially included; and cases from the Australian site oversampled cases with early age at diagnosis. For the present study, selection criteria for cases (n=1,313) were diagnosis of breast cancer at/or before 45 years and self-reported race/ethnicity plus grandparents' country of origin information consistent with Caucasian, East Asian, Hispanic/Latino, or African American racial/ethnic heritage. The controls (n=1,123) were frequency matched to the cases within each center by racial/ethnic group, with age at selection not more than 10 years older or younger than the age range at diagnosis of the cases ascertained at the same center. Because of the shortage of available controls in some ethnic and age groups, the frequency matching was not one-to-one in all subgroups. The design of this study has been described in detail previously (6,29-31). Recruitment and genetic studies were approved by the Ethics Committee of the International Agency for Research on Cancer (IARC), the University of Utah Institutional Review Board (IRB), and the local IRBs of the BCFR centers from which we received samples. Written informed consent was obtained from each participant. Distribution of cases and controls by center and ethnicity is shown in Supplemental Table 1.

Multiple-case breast cancer families—Index cases from 798 multiple-case breast or breast/ovarian cancer families were selected for mutation screening of *RINT1*. These consisted of 114 youngest affected members of families participating in the Australian Breast Cancer Family Study (ABCFS) (26) and 684 youngest affected members of families participating in the kConFab familial breast cancer resource (27), selected on the basis of

strength of family history and availability of DNA samples. Informed consent was as described for the women whose samples were selected for exome sequencing described above.

Laboratory and Bioinformatic Methods

Exome Sequencing—Capture was performed using SeqCap EZ Exome v2.0 (Roche Nimblegen) (n=86) and TruSeq Exome Enrichment Kit (Illumina) (n=3). Sequencing was performed using SOLiD™ 3.5 (n=8), SOLiD™ 4 (n=28), 5500xl SOLiD™ (n=12) (Life Technologies), GAIIX (n=2) and HiSeq (n=39) instruments (Illumina). Mapping to the human reference genome build hg19 was performed using BioScope v1.2 (SOLiD™ 3.5 data), BioScope v1.3 (SOLiD™ 4 data), LifeScope 2_5.1 (5500xl SOLiD™ data), bwa 0.6.2 (32) (GAIIX data), Novoalign (HiSeq data, n=40) and bwa 0.7.5a (HiSeq data, n=3). Local realignment was performed using the GATK 1.5-25 RealignerTargetCreator and IndelRealigner modules, duplicates were removed using Picard MarkDuplicates 1.29, and sorting and indexing were performed using SAMtools 0.1.8. Variant calling was conducted using the GATK 1.5-25 UnifiedGenotyper module. Variant calls were filtered to the capture target regions using BEDtools 2.16.2 (33) and bed files provided by the manufacturers. Further filtering was performed using FAVR (34). Annotation was performed using ANNOVAR (35).

Case-control mutation screening—For mutation screening of the coding exons and proximal splice junction regions of *RINT1* (NM021930), we used 30 ng of mixed whole-genome amplified (WGA) DNA obtained from two independent WGA reactions. The laboratory process was as described in detail for our recent studies of *ATM*, *CHEK2*, *XRCC2*, and *RAD51* (6,29-31). Our semi-automated approach, handled by a Laboratory Information Management System (LIMS) (36,37), relies on mutation scanning by High-Resolution Melt (HRM) curve analysis followed by direct Sanger sequencing of the individual samples for confirmation of the sequence variant. In a previous work, we showed that the HRM technique performed with high sensitivity and specificity (1.0, and 0.8, respectively, for amplicons of <400 bp) for mutation screening by comparing the results with those obtained with Sanger sequencing (38).

All exonic sequence variants, plus intronic sequence variants that fell within 20 bp of a splice acceptor or eight bp of a splice donor, and were either unreported or had an allele frequency of <0.5% in the large scale reference groups “Caucasian Americans”, “African Americans” and “East Asians” based on exome variant server (EVS) and 1,000 genomes project (1000G) data, were re-amplified from genomic DNA or from the two independent WGA reaction products to confirm the presence of the variant.

All samples that failed either at the primary PCR, secondary PCR, or sequencing reaction stage were re-amplified from WGA DNAs or genomic DNAs. Primer and probe sequences are available in Supplemental Table 2.

Sequence analysis of multiple-case breast cancer families—The 114 probands from the ABCFS study were screened by HRM followed by Sanger sequencing confirmation of aberrant melt curves as described above. The 684 cases from kConFab

resource were screened for mutations in *RINT1* via targeted capture followed by massively parallel sequencing. Briefly, as part of a custom target enrichment (Agilent Technologies) totalling 1.49 Mb we included the capture of the 15 coding and flanking exonic regions of *RINT1* (chr07: 105,172,532-105,208,124; hg19) followed by 100bp paired-end sequencing on HiSeq 2000 (Illumina) at Axseq Technologies, Inc (Rockville MD)/Macrogen.

Alignments and scoring of missense substitutions—We used M-Coffee, which is part of the T-Coffee (Tree-based Consistency Objective Function for alignment Evaluation) software suite of alignment tools to prepare a protein multiple sequence alignment for *RINT1* in order to predict the effect of missense substitutions. The alignment included sequences from 17 species: *Homo sapiens*, *Macaca mulatta*, *Callithrix jacchus*, *Mus musculus*, *Bos taurus*, *Loxodonta africana*, *Dasyus novemcinctus*, *Monodelphis domestica*, *Ornithorhynchus anatinus*, *Gallus gallus*, *Xenopus tropicalis*, *Latimeria chalumnae*, *Danio rerio*, *Branchiostoma lanceolatum*, *Strongylocentrotus purpuratus*, *Nematostella vectensis*, and *Trichoplax adhaerens*. The alignment was characterized using the Protpars routine of Phylogeny Inference Package version 3.2 software (PHYLIP) (39) to make a maximum parsimony estimate of the number of substitutions that occurred along each clade of the underlying phylogeny. The sequence alignment, or updated versions thereof, is available at the Align-GVGD website. Missense substitutions observed during our mutation screening of *RINT1* were scored using Align-GVGD with our curated alignment, and with PolyPhen2 using its precompiled alignments.

In silico prediction on splicing—All exonic sequence variants, plus intronic variants detected in the vicinity of the splice junction sequences, with allele frequencies <0.5%, were scored for their potential impact on splicing using MaxEntScan (MES), which computes the maximum entropy score of a given sequence using splice site models trained on human data (40). We calibrated MES by calculating the average and standard deviation of MES scores for the wild-type splice junctions in *BRCA1*, *BRCA2*, and *ATM*, allowing us to convert raw MES scores into z-scores. Based on *BRCA1* and *BRCA2* mutation screening data used previously to calibrate Align-GVGD (41,42), we found that rare variants that fall within the acceptor or donor region and reduce the MES score for the splice signal in which they fall show an approximately 95% probability to damage splice junction function when they result in a calibrated MES score of $z < -2$, or approximately 40% probability when they result in a calibrated MES score of $-2 < z < -1$. Additionally, exonic rare variants that increase the MES donor score of their sequence context and result in a calibrated MES donor score of $z > 0$ have an approximately 64% probability to create a *de novo* donor, while if they result in a calibrated MES score of $-2 \leq z \leq 0$ the probability to create a *de novo* donor is approximately 30% (Vallee *et al.* manuscript in preparation). These MES-based rules were used to identify rare sequence variants that are likely to alter *RINT1* gene mRNA splicing.

Assessment of c.129C>A (p.V43V) and c.1333+1G>A *RINT1* transcripts—Epstein-Barr Virus-transformed lymphoblastoid cell lines (LCL) bearing either reference sequence *RINT1*, *RINT1* c.129C>A or *RINT1* c.1333+1G>A were cultured without or with cycloheximide (to stabilize transcripts sensitive to decay) and prepared for RNA extraction as described previously (43,44). To guard against amplification of genomic DNA, the

reverse transcription primer and all of the PCR primers except one (due to restrictions in length) were designed to span *RINT1* exon-exon boundaries. cDNA was synthesized using the Thermoscript RT-PCR system kit (Life Technologies), and *RINT1* products were PCR amplified using Amplitaq Gold DNA Polymerase (Life Technologies). Non-reverse transcribed templates (extracted total RNA added in place of cDNA during PCR) were used to control for genomic DNA contamination effects. PCR products were separated by agarose gel electrophoresis. Bands were excised from the gel and purified using the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany) prior to Sanger sequencing.

Minigene assay characterization of *RINT1* c.{1334-5delA; 1334-1_1335delGTT}

—We used a minigene assay to characterize the effect of c.{1334-5delA; 1334-1_1335delGTT} on *RINT1* RNA splicing. A region spanning some or all of *RINT1* intron 9, exon 10, and 924bp of intron 10 was amplified from the human *RINT1*-bearing BAC clone RP11-96B13 (BACPAC Resources Center, Children's Hospital, Oakland USA) and inserted into the exon trapping vector pSPL3 (a generous gift from Dr. Bernd Wissinger and Dr. Nicole Weisschuh). The mutation was introduced into an intron 9 PCR product by incorporating c.{1334-5delA; 1334-1_1335delGTT} into an intron 9 reverse primer; the mutation was then inserted into the minigene construct by *in-vitro* homologous recombination using the Cold Fusion Cloning Kit (System Biosciences).

For expression assays, the minigene constructs were transfected into both COS-7 and HEK293 cells using TurboFect (Thermo Scientific) for COS-7, or FuGene 6 (Promega) for HEK293, following the manufacturers' recommended protocols. Transfected cells were cultured for 44 hours and then cultured for an additional 4 hours without or with cycloheximide (to stabilize transcripts bearing frameshift mutations). Total RNA was isolated using RNeasy kits (Qiagen), treated with Turbo DNase (Life Technologies) to remove contaminating genomic DNA, and then reverse transcribed using SuperScript VILO (Life Technologies). Minigene transcripts were PCR amplified using NEBNext High-Fidelity 2× PCR Master Mix (New England Biolabs) and then separated by agarose gel electrophoresis. PCR products from the mutation-bearing construct were cloned into the pCR-BluntII-TOPO vector (Life Technologies), and clone sequences evaluated by Sanger sequencing. Primer sequences for all transcript assessments are shown in Supplemental Table 3.

Statistical Analyses

Case-Control Analysis—Missense variants in *RINT1* identified by HRM/Sanger sequencing were assessed using the *in-silico* analysis tools Align-GVGD and PolyPhen2 (see above) for missense variants. Sequence variants were classified as pathogenic by Align-GVGD if they were any grade higher than C0 (equivalent to being conserved in mammals), and by PolyPhen2 if assessed as 'probably damaging'. In addition, sequence variants predicted to result in a truncated protein were considered pathogenic (although all potential splice variants were validated experimentally as described above). The frequency of different groupings of variants among cases and controls based on these definitions were compared by unconditional logistic regression, adjusting for ethnicity and study center.

Analysis of cancers in RINT1 mutation-carrier families—We compared the incidence of cancers at all sites in the family members of the 23 case probands with likely pathogenic *RINT1* rare variants to that in family members of 1,290 case probands without any such variants. From these 1,313 families we excluded: three families who did not have pedigree data; 19 families that had been subsequently found to include carriers of pathogenic mutations in *BRCA1* (6 families), *BRCA2* (6), and the specific variants *PALB2* cPlease add “.” between c and 31133113G>A (p.W1038*) (4) and *ATM* c.7271T>G (p.V2424G)(3), known to be associated with high-risk of breast cancer (45,46); and eight families in which the proband was found to carry a rare *RINT1* missense variant that was not scored as pathogenic by either *in silico* prediction tool.

Probands and their relatives were censored at the earliest of: their age at last contact/vital status information, age at death, or age at first invasive cancer (at any site). We excluded from this latter comparison set individuals with missing year of birth or who were born before 1920 as diagnostic and censoring information was likely to be less reliable in this cohort. With these restrictions, the data set consisted of 373 members of 23 *RINT1* mutation-carrying families with a total of 16,037 person-years and 20,918 members of 1,260 families without *RINT1* mutations comprising 914,178 person-years. In these analyses, cancers at sites other than breast diagnosed in probands were included if they had occurred prior to the breast cancer diagnosis. For these analyses and those described below, we used the definition of likely pathogenic as any rare variant that was scored by either of the two *in silico* methods used for variant assessment (corresponding to line 4 of Table 2).

In a first exploratory analysis we used a simple incidence rate ratio approach to assess differences in incidences between members of families in which the proband carried a likely pathogenic *RINT1* variant and members of families not harbouring rare variants in *RINT1* to examine the overall incidences of cancers other than breast cancers, according to the following topological site groups: oral cavity and pharynx (C00-14), digestive organs (C15-26), respiratory organs (C30-39), bone (C40-41), skin (C43-44), soft tissue (C45 – 49), female genital organs (C51-58), male genital organs (C60-63), urinary tract (C64-68), eye/brain/CNS (C69-72), thyroid/endocrine (C73-75), and lymphoid tissues (C81-96).

In more detailed analyses, we examined the incidence of first primary cancers using two different approaches. Firstly, we used Cox proportional hazards model in a retrospective cohort analysis to estimate the hazard ratio (HR) for occurrence of cancer by site group among members of families of probands found to carry a *RINT1* likely pathogenic sequence variant compared to members of families not carrying a rare *RINT1* variant. These analyses included adjustment for study center and used a robust variance estimate to account for clustering of cancers within families. We obtained DNA samples of 36 family members of probands carrying a *RINT1* rare variant from the relevant study centers to enable genotyping of these individuals for the variant identified in the index case of the family. These genotypes were used to inform the estimation of carrier probabilities for our second set of analyses as described below.

In this analysis we estimated carrier probabilities for each individual using the program MENDEL (47) based on the observed phenotypes, the family genotypes, and assuming the

HR for breast cancer as estimated from the case-control analysis. We then calculated a weighted standardized incidence ratio (SIR) by comparing to age- sex- site- and registry-based incidence rates averaged over the period 1988–2002 using data from Cancer Incidence in Five Continents versions 7,8, and 9 (IARC). Cancer Registries were matched to the location/ethnicity of the cases as follows: Canada: Ontario; Australia: Victoria, New South Wales; California: Hispanic, East Asian, Black with weights equal to the probability of being a *RINT1* mutation carrier. $SIR = \sum w_i O_i / \sum w_i E_i$, where O_i is a binary variable indicating the presence/absence of the given cancer type/group, E_i is the cumulative incidence for that individual based on their censoring age, gender, and residence, and w_i is the estimated carrier probability for the i th individual.

Exact p-values were calculated assuming the (estimated) observed number of cases in carriers, $\sum w_i O_i$, follows a Poisson distribution with mean $\sum w_i E_i$. Ninety-five percent confidence intervals were calculated using the accurate Poisson-based approximation as detailed in Rothman and Boice (48).

We also performed several sensitivity analyses by: varying the inclusion criterion of birth year after 1920 using a range from 1900 to 1940; employing a broader definition of pathogenicity for the observed sequence variants than that in the main analyses; and incorporating the families of controls with and without *RINT1* variants. All statistical analyses were performed using STATA v 12.0 (StataCorp, College Station, TX).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the Cancer Council Victoria (Grant ID 628774 to MCS); the United States National Institute of Health, R01CA155767 to DEG, SVT and MCS, R01CA121245 to SVT, and P30CA042014; The Australian National Health and Medical Research Council (NHMRC; Grant IDs 466668, 509038, and APP1025145); The Victoria Breast Cancer Research Consortium; The University of Melbourne (infrastructure award to JLH); a Victorian Life Sciences Computation Initiative (VLSCI) grant number VR0053 on its Peak Computing Facility at the University of Melbourne, an initiative of the Victorian State Government, and by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research, and the Ministère de l'enseignement supérieur, de la recherche, de la science, et de la technologie du Québec through Génome Québec.

The Australian Breast Cancer Family Registry (ABCFR; 1992-1995) was supported by the Australian NHMRC, the New South Wales Cancer Council, the Victorian Health Promotion Foundation (Australia). We wish to thank Margaret McCredie for key role in the establishment and leadership of the ABCFR in Sydney, Australia, and the families who donated their time, information and biospecimens. The Genetic Epidemiology Laboratory at the University of Melbourne has also received generous support from Mr B. Hovey and Dr and Mrs R.W. Brown to whom we are most grateful. We also thank Mr G. Keough for assistance with manuscript preparation.

The work of the Breast Cancer Family Registry (BCFR) centers (BCFR-AU (ABCFS), BCFR-NC, BCFR-NY, BCFR-ON, BCFR-PA (FCCC) and BCFR-UT) was supported by grant UM1 CA164920 from the National Cancer Institute. The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Breast Cancer Family Registry, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government or the Breast Cancer Family Registry.

BCFR-ON work was additionally supported by the Canadian Institutes of Health Research "CIHR Team in Familial Risks of Breast Cancer" program.

We wish to thank Heather Thorne, Eveline Niedermayr, all the Kathleen Cuninghame Foundation Consortium for research into Familial Breast Cancer (kConFab) research nurses and staff, the heads and staff of the Family Cancer Clinics, and the Clinical Follow Up Study (funded by NHMRC grants 145684, 288704 and 454508) for their contributions to this resource, and the many families who contribute to kConFab. kConFab has been supported by grants from the National Breast Cancer Foundation (Australia), the NHMRC, the Queensland Cancer Fund, the Cancer Councils of New South Wales, Victoria, Tasmania and South Australia, the Cancer Foundation of Western Australia and NIH via (U01 CA69638).

MCS is a NHMRC Senior Research Fellow, JLH is a NHMRC Senior Principal Research Fellow. TN-D is a Susan G. Komen for the Cure Postdoctoral Fellow.

References

1. Claus EB, Schildkraut JM, Thompson WD, Risch NJ. The genetic attributable risk of breast and ovarian cancer. *Cancer*. 1996; 77:2318–2324. [PubMed: 8635102]
2. Goldgar DE, Easton DF, Cannon-Albright LA, Skolnick MH. Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. *J Natl Cancer Inst*. 1994; 86:1600–1608. [PubMed: 7932824]
3. Hemminki K, Granstrom C, Czene K. Attributable risks for familial breast cancer by proband status and morphology: a nationwide epidemiologic study from Sweden. *Int J Cancer*. 2002; 100:214–219. [PubMed: 12115572]
4. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet*. 2013; 45:353–61. 361e1. [PubMed: 23535729]
5. Thompson ER, Doyle MA, Ryland GL, Rowley SM, Choong DY, Tothill RW, et al. Exome Sequencing Identifies Rare Deleterious Mutations in DNA Repair Genes FANCC and BLM as Potential Breast Cancer Susceptibility Alleles. *PLoS Genet*. 2012; 8:e1002894. [PubMed: 23028338]
6. Park DJ, Lesueur F, Nguyen-Dumont T, Pertesi M, Odefrey F, Hammet F, et al. Rare mutations in XRCC2 increase the risk of breast cancer. *Am J Hum Genet*. 2012; 90:734–739. [PubMed: 22464251]
7. Ruark E, Snape K, Humburg P, Loveday C, Bajrami I, Brough R, et al. Mosaic PPM1D mutations are associated with predisposition to breast and ovarian cancer. *Nature*. 2013; 493:406–410. [PubMed: 23242139]
8. Snape K, Ruark E, Tarpey P, Renwick A, Turnbull C, Seal S, et al. Predisposition gene identification in common cancers by exome sequencing: insights from familial breast cancer. *Breast Cancer Res Treat*. 2012; 134:429–433. [PubMed: 22527104]
9. Feng BJ, Tavtigian SV, Southey MC, Goldgar DE. Design considerations for massively parallel sequencing studies of complex human disease. *PLoS One*. 2011; 6:e23221. [PubMed: 21850262]
10. Park DJ, Odefrey FA, Hammet F, Giles GG, Baglietto L, ABCFS MCCS, Hopper JL, et al. FAN1 variants identified in multiple-case early-onset breast cancer families via exome sequencing: no evidence for association with risk for breast cancer. *Breast Cancer Res Treat*. 2011; 130:1043–1049. [PubMed: 21858661]
11. Hilbers FS, Wijnen JT, Hoogerbrugge N, Oosterwijk JC, Collee MJ, Peterlongo P, et al. Rare variants in XRCC2 as breast cancer susceptibility alleles. *J Med Genet*. 2012; 49:618–620. [PubMed: 23054243]
12. van Asperen CJ, Brohet RM, Meijers-Heijboer EJ, Hoogerbrugge N, Verhoef S, Vasen HF, et al. Cancer risks in BRCA2 families: estimates for sites other than breast and ovary. *J Med Genet*. 2005; 42:711–719. [PubMed: 16141007]
13. Mucci E, Milne RL, Mendez-Villamil EY, Hopper JL, John EM, Andrulis IL, et al. Risk of pancreatic cancer in breast cancer families from the breast cancer family registry. *Cancer Epidemiol Biomarkers Prev*. 2013; 22:803–811. [PubMed: 23456555]
14. Win AK, Lindor NM, Winship I, Tucker KM, Buchanan DD, Young JP, et al. Risks of colorectal and other cancers after endometrial cancer for women with Lynch syndrome. *J Natl Cancer Inst*. 2013; 105:274–279. [PubMed: 23385444]

15. Dowty JG, Win AK, Buchanan DD, Lindor NM, Macrae FA, Clendenning M, et al. Cancer risks for MLH1 and MSH2 mutation carriers. *Hum Mutat.* 2013; 34:490–497. [PubMed: 23255516]
16. Win AK, Lindor NM, Young JP, Macrae FA, Young GP, Williamson E, et al. Risks of primary extracolonic cancers following colorectal cancer in lynch syndrome. *J Natl Cancer Inst.* 2012; 104:1363–1372. [PubMed: 22933731]
17. Bellizzi AM, Frankel WL. Colorectal cancer due to deficiency in DNA mismatch repair function: a review. *Adv Anat Pathol.* 2009; 16:405–417. [PubMed: 19851131]
18. Yandell M, Huff C, Hu H, Singleton M, Moore B, Xing J, et al. A probabilistic disease-gene finder for personal genomes. *Genome Res.* 2011; 21:1529–1542. [PubMed: 21700766]
19. Hu H, Huff CD, Moore B, Flygare S, Reese MG, Yandell M. VAAST 2.0: improved variant classification and disease-gene identification using a conservation-controlled amino acid substitution matrix. *Genet Epidemiol.* 2013; 37:622–634. [PubMed: 23836555]
20. Xiao J, Liu CC, Chen PL, Lee WH. RINT-1, a novel Rad50-interacting protein, participates in radiation-induced G(2)/M checkpoint control. *J Biol Chem.* 2001; 276:6105–6111. [PubMed: 11096100]
21. Lin X, Liu CC, Gao Q, Zhang X, Wu G, Lee WH. RINT-1 serves as a tumor suppressor and maintains Golgi dynamics and centrosome integrity for cell survival. *Mol Cell Biol.* 2007; 27:4905–4916. [PubMed: 17470549]
22. Sun Y, Shestakova A, Hunt L, Sehgal S, Lupashin V, Storrie B. Rab6 regulates both ZW10/RINT-1 and conserved oligomeric Golgi complex-dependent Golgi trafficking and homeostasis. *Mol Biol Cell.* 2007; 18:4129–4142. [PubMed: 17699596]
23. Wainman A, Giansanti MG, Goldberg ML, Gatti M. The Drosophila RZZ complex - roles in membrane trafficking and cytokinesis. *J Cell Sci.* 2012; 125:4014–4025. [PubMed: 22685323]
24. Schmitt HD. Dsl1p/Zw10: common mechanisms behind tethering vesicles and microtubules. *Trends Cell Biol.* 2010; 20:257–268. [PubMed: 20226673]
25. Civril F, Wehenkel A, Giorgi FM, Santaguida S, Di Fonzo A, Grigorean G, et al. Structural analysis of the RZZ complex reveals common ancestry with multisubunit vesicle tethering machinery. *Structure.* 2010; 18:616–626. [PubMed: 20462495]
26. John EM, Hopper JL, Beck JC, Knight JA, Neuhausen SL, Senie RT, et al. The Breast Cancer Family Registry: an infrastructure for cooperative multinational, interdisciplinary and translational studies of the genetic epidemiology of breast cancer. *Breast Cancer Res.* 2004; 6:R375–R389. [PubMed: 15217505]
27. Mann GJ, Thorne H, Balleine RL, Butow PN, Clarke CL, Edkins E, et al. Analysis of cancer risk and BRCA1 and BRCA2 mutation prevalence in the kConFab familial breast cancer resource. *Breast Cancer Res.* 2006; 8:R12. [PubMed: 16507150]
28. Smith P, McGuffog L, Easton DF, Mann GJ, Pupo GM, Newman B, et al. A genome wide linkage search for breast cancer susceptibility genes. *Genes Chromosomes Cancer.* 2006; 45:646–655. [PubMed: 16575876]
29. Tavtigian SV, Oefner PJ, Babikyan D, Hartmann A, Healey S, Le Calvez-Kelm F, et al. Rare, evolutionarily unlikely missense substitutions in ATM confer increased risk of breast cancer. *Am J Hum Genet.* 2009; 85:427–446. [PubMed: 19781682]
30. Le Calvez-Kelm F, Lesueur F, Damiola F, Vallee M, Voegelé C, Babikyan D, et al. Rare, evolutionarily unlikely missense substitutions in CHEK2 contribute to breast cancer susceptibility: results from a breast cancer family registry case-control mutation-screening study. *Breast Cancer Res.* 2011; 13:R6. [PubMed: 21244692]
31. Le Calvez-Kelm F, Oliver J, Damiola F, Forey N, Robinot N, Durand G, et al. RAD51 and Breast Cancer Susceptibility: No Evidence for Rare Variant Association in the Breast Cancer Family Registry Study. *PLoS One.* 2012; 7:e52374. [PubMed: 23300655]
32. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009; 25:1754–1760. [PubMed: 19451168]
33. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010; 26:841–842. [PubMed: 20110278]
34. Pope BJ, Nguyen-Dumont T, Odefrey F, Hammet F, Bell R, Tao K, et al. FAVR (Filtering and Annotation of Variants that are Rare): methods to facilitate the analysis of rare germline genetic

- variants from massively parallel sequencing datasets. *BMC Bioinformatics*. 2013; 14:65. [PubMed: 23441864]
35. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010; 38:e164. [PubMed: 20601685]
 36. Voegelé C, Tavtigian SV, de Silva D, Cuber S, Thomas A, Le Calvez-Kelm F. A Laboratory Information Management System (LIMS) for a high throughput genetic platform aimed at candidate gene mutation screening. *Bioinformatics*. 2007; 23:2504–2506. [PubMed: 17709339]
 37. Nguyen-Dumont T, Calvez-Kelm FL, Forey N, McKay-Chopin S, Garritano S, Gioia-Patricola L, et al. Description and validation of high-throughput simultaneous genotyping and mutation scanning by high-resolution melting curve analysis. *Hum Mutat*. 2009; 30:884–890. [PubMed: 19347964]
 38. Garritano S, Gemignani F, Voegelé C, Nguyen-Dumont T, Le Calvez-Kelm F, De Silva D, et al. Determining the effectiveness of High Resolution Melting analysis for SNP genotyping and mutation scanning at the TP53 locus. *BMC Genet*. 2009; 10:5. [PubMed: 19222838]
 39. Felsenstein J. PHYLIP - Phylogeny Inference Package (version 3.2). *Cladistics*. 1989; 5:164–166.
 40. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol*. 2004; 11:377–394. [PubMed: 15285897]
 41. Tavtigian SV, Byrnes GB, Goldgar DE, Thomas A. Classification of rare missense substitutions, using risk surfaces, with genetic- and molecular-epidemiology applications. *Hum Mutat*. 2008; 29:1342–1354. [PubMed: 18951461]
 42. Easton DF, Deffenbaugh AM, Pruss D, Frye C, Wenstrup RJ, Allen-Brady K, et al. A systematic genetic assessment of 1,433 sequence variants of unknown clinical significance in the BRCA1 and BRCA2 breast cancer-predisposition genes. *Am J Hum Genet*. 2007; 81:873–883. [PubMed: 17924331]
 43. Southey MC, Tesoriero A, Young MA, Holloway AJ, Jenkins MA, Whitty J, et al. A specific GFP expression assay, penetrance estimate, and histological assessment for a putative splice site mutation in BRCA1. *Hum Mutat*. 2003; 22:86–91. [PubMed: 12815598]
 44. Teo ZL, Park DJ, Provenzano E, Chatfield CA, Odefrey FA, Nguyen-Dumont T, et al. Prevalence of PALB2 mutations in Australasian multiple-case breast cancer families. *Breast Cancer Res*. 2013; 15:R17. [PubMed: 23448497]
 45. Goldgar DE, Healey S, Dowty JG, Da Silva L, Chen X, Spurdle AB, et al. Rare variants in the ATM gene and risk of breast cancer. *Breast Cancer Res*. 2011; 13:R73. [PubMed: 21787400]
 46. Southey MC, Teo ZL, Dowty JG, Odefrey FA, Park DJ, Tischkowitz M, et al. A PALB2 mutation associated with high risk of breast cancer. *Breast Cancer Res*. 2010; 12:R109. [PubMed: 21182766]
 47. Lange K, Weeks D, Boehnke M. Programs for Pedigree Analysis: MENDEL, FISHER, and dGENE. *Genet Epidemiol*. 1988; 5(6):471–472. [PubMed: 3061869]
 48. Rothman, KJ.; Boice, JD. *Epidemiologic Analysis With a Programmable Calculator*. Washington DC: US Government Printing Office; 1979. nih Publication 79-1649

Significance

The work described here adds RINT1 to the growing list of genes in which rare sequence variants are associated with intermediate levels of breast cancer risk. RINT1 also is strongly associated with a spectrum of cancers found in mismatch repair defects that has both clinical applications and raises interesting biological questions.

Figure 1. Pedigrees of families found to carry mutations in *RINT1* via WES

Mutation status is indicated for all family members for whom a DNA sample was available.

a) Pedigree with c.343C>T (p.Q115X) **b)** Pedigree with c.1207G>T (p.D403Y) **c)** Pedigree with c.1132_1134del (p.M378del);

Mutation carriers are indicated in black text, obligate carriers are indicated in grey italics text. Cancer diagnosis and age at onset is indicated for affected members. *, DNA underwent WES; BC, Breast Cancer (black filled symbols); uk, age at diagnosis unknown; PC, Pancreatic Cancer; CC, Colon Cancer; MM, Malignant Melanoma; UK, unknown age; BIC, Bladder Cancer; LC, Lung Cancer; LiC, Liver Cancer; KC, Kidney Cancer; BnC, Brain Cancer; CUP, Cancer of unknown primary; NMSC, non-melanoma skin cancer; UC, Uterine cancer (all grey filled symbols); (V), verified cancer (via cancer registry or pathology report); wt, wildtype; d., age at death (years). Some symbols represent more than one person as indicated by a numeral.

d) RINT1 protein multiple sequence alignment covering positions M378 and D403.

Positions L376 and K383 are the nearest invariant positions before and after M378; as the spacing between those positions is invariant, deletion of M378 would be expected to damage the protein. As position D403 is invariant, a non-conservative substitution at that position would also be expected to damage the protein.

Figure 2. Localization of *RINT1* rare variants

Upper panel. Domain organization of *RINT1* and distribution of rare variants in the investigated series: orange, multiple-cases families (exome) - red, BCFR cases - blue, BCFR controls - green, KConFab and ABCFS multiple-case families. Symbols: triangle, truncating mutation or mutation altering splicing - filled diamond, putative pathogenic missense substitution - empty diamond, likely benign missense substitution.

Lower panel. Grantham Variation (GV) conservation profile across *RINT1* based on multiple protein sequence alignment through to Platypus (*Oana* - *Ornithorhynchus Anatinus*). *RINT1* TIP20 domain: from amino acid 147 to amino acid 701. Note that $GV=0$ is indicative of evolutionary invariance and $0 < GV < 65$ is indicative of positions where cross-species variation is limited to conservative amino acid substitutions.

Table 1

Distribution of *RINT1* rare variants (with frequency <0.5%) identified in the BCPR cases and controls.

Variant type	Variant	Effect on protein	Con.	Ca.	Align-GVGD	PolyPhen2.1 (Q6NU/Q1, Hum Div)
In-frame del	c.1329delA; 1334-1_1335delG/TT	p.(Phe445_Ala446delinsSer)	0	1	N/A	N/A
In-frame del	c.2361G>A	p.W787*	1	1	N/A	N/A
Missense	c.281C>G	p.T94R	1	0	C65	Prob. Dam.
Missense	c.301A>G	p.K101E	1	0	C55	Poss. Dam.
Missense	c.319T>G	p.L107V	0	1	C25	Prob. Dam.
Missense	c.376C>T	p.H126Y	1	1	C0	Benign
Missense	c.388G>A	p.A130T	3	3	C0	Benign
Missense	c.413C>T	p.A138V	1	3	C65	Benign
Missense	c.501A>T	p.Q167H	0	1	C15	Benign
Missense	c.532T>C	p.Y178H	0	1*	C65	Benign
Missense	c.736C>A	p.P246T	1	1	C0	Benign
Missense	c.778G>T	p.A260S	0	1	C0	Benign
Missense	c.782C>T	p.P261L	1	1	C65	Benign
Missense	c.1025T>C	p.M342T	0	1	C45	Benign
Missense	c.1121G>A	p.R374Q	0	1	C35	Prob. Dam.
Missense	c.1256C>G	p.P419R	0	1	C0	Prob. Dam.
Missense	c.1270A>T	p.S424C	0	1	C15	Benign
Missense	c.1385C>T	p.S462L	0	1	C65	Prob. Dam.
Missense	c.1449G>T	p.M483I	1	0	C0	Benign
Missense	c.1465A>G	p.I489V	0	1	C0	Benign
Missense	c.1519G>A	p.E507K	0	2	C55	Poss. Dam.
Missense	c.1562C>T	p.T521I	0	1	C65	Prob. Dam.
Missense	c.1949C>T	p.P650L	0	1	C65	Prob. Dam.
Missense	c.1985T>C	p.L662S	0	1	C65	Prob. Dam.
Missense	c.2036T>C	p.V679A	0	1	C0	Benign
Missense	c.2090A>G	p.N697S	0	1	C45	Prob. Dam.
Missense	c.2128C>T	p.R710W	0	1	C65	Prob. Dam.
Missense	c.2159G>T	p.C720F	0	1	C65	Prob. Dam.

Variant type	Variant	Effect on protein	Con.	Ca.	Align-GVGD	PolyPhen2.1 (Q6NUQ1, Hum Div)
Missense	c.2176T>C	p.Y726H	0	1	C65	Poss. Dam.
Missense	c.2179T>C	p.F727L	1	0	C15	Prob. Dam.
Missense	c.2362C>T	P788S	0	1*	C65	Prob. Dam.
Total Rare Variants			12	31		
Total Likely Pathogenic			6	23		

* This case carries both p.Y178H and p.P788S. In the analyses, only p. P788S was considered in the analyses. Variants highlighted in bold are deemed likely pathogenic, defined as A-GVGD score > C0 or PolyPhen “probably damaging”.

Table 2

Risk estimates and LR test p-values from *RINTI* case-control mutation screening.

Analysis	Crude OR (95% CI)	Adj OR (95% CI) (ethnicity)	Adj OR (95% CI) (center)	Adj* OR (95% CI) (ethnicity and center)
All rare variants (incl. C0)	2.24 (1.14, 4.38) p=0.019	2.31 (1.17, 4.55) p=0.016	2.31 (1.16, 4.62) p=0.018	2.33 (1.16, 4.65) p=0.017
PolyPhen2 (prob damaging)	3.73 (1.06, 13.13) p=0.040	4.00 (1.13, 14.18) p=0.032	3.10 (0.85, 11.30) p=0.086	3.19 (0.88, 11.59) p=0.078
Align-GVGD (> C0) ^d	3.17 (1.28, 7.85) p=0.013	2.96 (2.23, 3.92) p=0.0094	3.00 (1.18, 7.63) p=0.021	3.08 (1.22, 7.81) p=0.018
PolyPhen2 or Align-GVGD ^b	3.32 (1.35, 8.18) p=0.0091	3.53 (1.42, 8.74) p=0.0065	3.16 (1.25, 7.98) p=0.015	3.24 (1.29, 8.17) p=0.013
PolyPhen2 and Align-GVGD ^c	3.44 (0.97, 12.23) p=0.056	3.66 (1.02, 13.12) p=0.046	2.80 (0.76, 10.36) p=0.12	2.89 (0.78, 10.64) p=0.11

* OR are adjusted for race or ethnicity (Caucasian, East Asian, African American or Latina) and study center.

Notes:

^a In the binary analysis, only carriers of a missense substitution with grade>C0 or of an in-frame deletion (IFR) were considered.

^b Carriers of an IFR or carriers of a missense substitution with grade >C0, or predicted as probably damaging by PolyPhen2.

^c Carriers of an IFR or carriers of a missense substitution with grade >C0 and predicted as probably damaging by PolyPhen2.

Table 3
Analysis of incidence of first primary cancers in *RINT1* mutation-carrying families

Site Group	Cancers Observed ¹		RINT1 vs. Wild Type			RINT1 vs. population rates		
	RINT1	WT	HR	95% CI	p-value	SIR (carriers)	95% CI	p-value
All non-breast	51	2117	1.67	1.2–2.3	0.0009	2.23	1.3–3.5	0.005
Gastrointestinal (C15–C26)	13	363	2.51	1.4–4.4	0.001	2.46	0.8–5.9	0.050
Female (C51–C59)	12	251	2.89	1.6–5.3	0.0007	7.90	2.5–19	0.0008
Lynch syndrome ¹	18	375	3.24	1.9–5.7	4 × 10 ^{−5}	3.66	1.4–7.7	0.009
Lynch syndrome dx <60	13	191	3.79	2.2–6.5	10 ^{−6}	12.4	3.7–30	0.012
Extended Lynch syndrome ¹	30	823	2.44	1.5–3.9	0.0003	3.35	1.7–6.0	0.005
Extended Lynch syndrome dx <60	22	444	2.71	1.8–4.0	4 × 10 ^{−7}	10.9	4.7–21	0.0003
All but Lynch syndrome	33	1742	1.33	0.8–1.8	0.12	1.52	0.8–2.6	0.09
All but Extended Lynch syndrome	21	1294	1.16	0.9–1.9	0.50	1.21	0.5–2.4	0.33

¹ Number of cancers of each type observed in 16038 person-years in family members of RINT1 probands and 914178 person-years of observations in family members of wild-type probands.

² See text for definition